

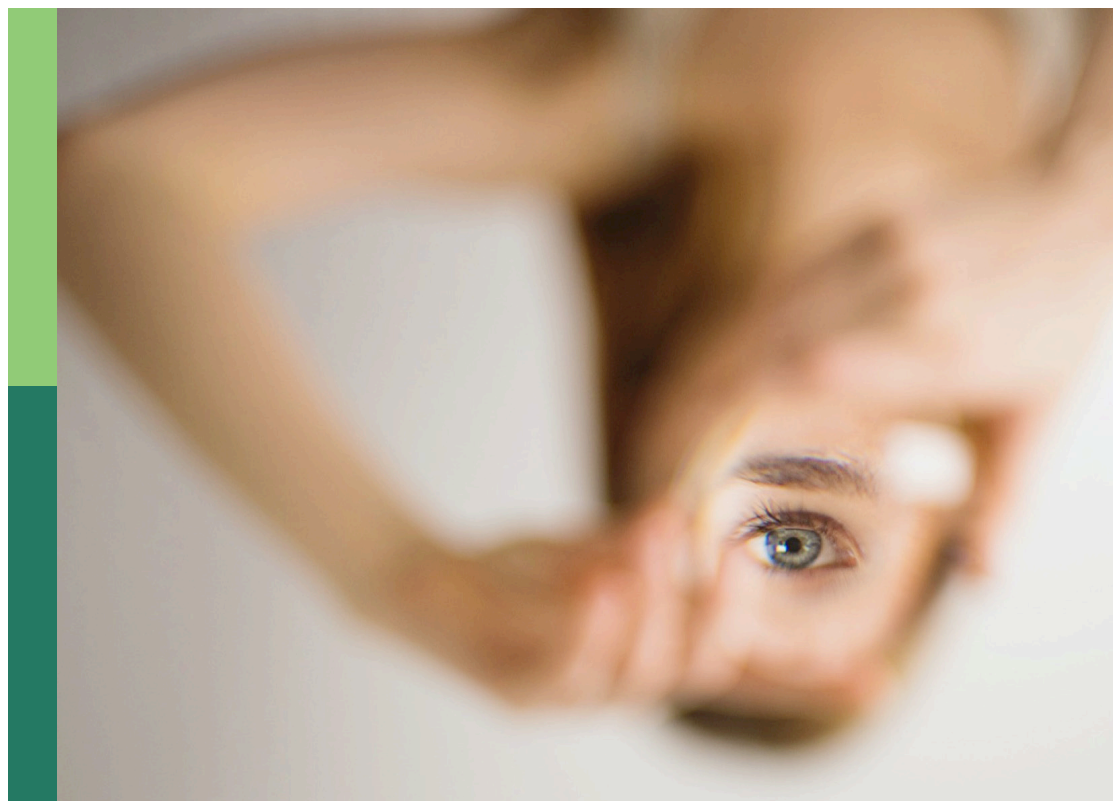
Reviews in psychology of language, volume II

Edited by

Thomas L. Spalding, Antonio Bova and
Antonio Benítez-Burraco

Published in

Frontiers in Psychology



FRONTIERS EBOOK COPYRIGHT STATEMENT

The copyright in the text of individual articles in this ebook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this ebook is the property of Frontiers.

Each article within this ebook, and the ebook itself, are published under the most recent version of the Creative Commons CC-BY licence. The version current at the date of publication of this ebook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or ebook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714
ISBN 978-2-8325-8374-6
DOI 10.3389/978-2-8325-8374-6

Generative AI statement

Any alternative text (Alt text) provided alongside figures in the articles in this ebook has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Reviews in psychology of language, volume II

Topic editors

Thomas L. Spalding — University of Alberta, Canada

Antonio Bova — Catholic University of the Sacred Heart, Italy

Antonio Benítez-Burraco — University of Seville, Spain

Citation

Spalding, T. L., Bova, A., Benítez-Burraco, A., eds. (2026). *Reviews in psychology of language, volume II*. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-8325-8374-6

Table of contents

04	Editorial: Reviews in psychology of language-volume II Antonio Benítez-Burraco, Antonio Bova and Thomas Spalding
06	Examining the relationship between foreign language anxiety and foreign language enjoyment from the perspective of positive psychology Qiangfu Yu
13	A cross-age odyssey of cognitive reading attributes: a scoping review Muhamad Firdaus Mohd Noh, Mohd Effendi Ewan Mohd Matore and Nur Ainil Sulaiman
32	Treatment of aphasia in linguistically diverse populations: current and future directions Gerald C. Imaezue, David Ajayi and Celine Davis
45	Language and cognitive function in children: a narrative review of neural, behavioral, and developmental evidence Ling Pu, Sergey Kiselev and Ningkun Xiao
58	The importance of intonation for children's understanding of verbal irony Jordanna Smith and Melanie Glenwright
68	Effect of the consonant context on the corner vowel produced by native and Chinese speakers: based on AESOP corpora Hongming Zhu
89	Parental emotional expression and children's oral narrative ability: a systematic review Zhi Hong Zhang, Suziyani Mohamed and Kamariah Abu Bakar
108	The influence of prime characteristics in semantic priming Ajay Mangat, Alexander Taikh, Christina Gagne and Thomas L Spalding
125	Chinese as a second language anxiety and intervention in a technology-assisted environment: a mini-review Bo Yu
139	Finger-tracking captures distinct cognitive signatures in L1 vs. L2 reading Eva Polo and Claudia Marzi



OPEN ACCESS

EDITED AND REVIEWED BY

Xiaolin Zhou,
Peking University, China

*CORRESPONDENCE

Antonio Benítez-Burraco
✉ abenitez8@us.es

RECEIVED 14 June 2026

ACCEPTED 22 June 2026

PUBLISHED 07 July 2026

CITATION

Benítez-Burraco A, Bova A and
Spalding T (2026) Editorial: Reviews in
psychology of language-volume II.
Front. Psychol. 17:1909048.
doi: 10.3389/fpsyg.2026.1909048

COPYRIGHT

© 2026 Benítez-Burraco, Bova and
Spalding. This is an open-access article
distributed under the terms of the
[Creative Commons Attribution License](#)
(CC BY). The use, distribution or
reproduction in other forums is
permitted, provided the original author(s)
and the copyright owner(s) are credited
and that the original publication in this
journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Editorial: Reviews in psychology of language-volume II

Antonio Benítez-Burraco ^{1*}, Antonio Bova² and
Thomas Spalding³

¹Department of Spanish, Linguistics, and Theory of Literature, Faculty of Philology, University of Seville, Seville, Spain, ²Department of Psychology, Faculty of Psychology, Università Cattolica del Sacro Cuore, Milan, Italy, ³Department of Psychology, Faculty of Arts, University of Alberta, Edmonton, AB, Canada

KEYWORDS

clinical linguistics, first language acquisition, language teaching, psycholinguistics, second language learning

Editorial on the Research Topic

Reviews in psychology of language-volume II

Over the last few decades, psycholinguistics has crucially contributed, together with neurolinguistics, to our understanding of how our brain processes language. Although languages are sensitive to their speakers' physical and social environment, the core features of language and how they are used depend primarily on how our brain is configured and works. Significant progress has been made in recent years in our understanding of language comprehension and production, language use in natural settings, language acquisition in children and language learning in adults, and language change throughout the lifespan. These are still key areas of interest for psycholinguistics, particularly now that they are being increasingly cross-fertilized by other related disciplines: neurolinguistics, as noted, but also biolinguistics, neuroscience, anthropology, and cognitive sciences, to name a few. At the same time, advances in many fields have opened up new avenues of research that practitioners of psycholinguistics and related fields should be aware of.

As with its first edition, the aim of this Research Topic is to gather comprehensive and up-to-date review articles on key aspects of psycholinguistics, with a focus on the topics and methodologies that are currently in the spotlight, such as the use of artificial intelligence for the improvement of our understanding of language processing by the brain or our teaching methods, or the role of linguistic and cognitive diversity in refining our understanding of language psychology. In this second volume of the Research Topic "*Reviews in Psychology of Language*," we have collected 10 contributions by 21 scholars.

Two articles address fundamental, traditional concerns of psycholinguistics. The contribution by [Pu et al.](#) re-examines the links between language and cognition. The authors support an active role for language in structuring our minds. They focus on the effect of language on executive function and social cognition. Adopting a multidisciplinary approach, they draw on evidence from behavioral, developmental, clinical, and neuroscience studies to propose a more integrative and dynamic view of the place of language within our cognitive architecture. The contribution by [Mangat et al.](#) focuses on what is perhaps the most familiar effect in psycholinguistics: semantic priming. The authors review the sources (automatic or strategic) of this effect and propose ways to improve current models of semantic priming.

The majority of articles in this Research Topic examine first language (L1) acquisition in children or second language (L2) learning in more or less formal contexts. Specifically, three articles focus on language acquisition by children. The contribution by [Zhang et al.](#) is a meta-analysis of approximately 100 studies examining whether the expression of emotional content by parents has a positive impact on the oral narrative abilities of their children. The authors conclude that the available evidence supports both direct and indirect effects. In the same vein, the article by [Smith and Glenwright](#) reviews the diverse research methods used for the determination of the impact of intonation on children's ability to understand verbal irony. The article also comments on the cues used by children to identify verbal irony, along with the cross-linguistic factors impacting this pragmatic use of intonation. Finally, the contribution by [Noh et al.](#) aims to characterize the different methodologies used for the evaluation of reading abilities in youth vs. adults. These authors review approximately 50 studies and conclude that adults are typically evaluated in terms of a broader range of skills and abilities and suggest some methodological strategies to overcome current limitations in analyzing reading abilities in children.

Two studies in this Research Topic compare aspects of language acquisition by L1 and L2 learners. The contribution by [Polo and Marzi](#) also focuses on reading abilities. More precisely, these authors introduce finger tracking as a new method of gaining insight into the predictive and integrative mechanisms involved in real-time language comprehension during reading. They apply this method to examine reading behavior by native speakers and learners of English. Similarly, the article by [Zhu](#) adopts a corpus approach to examine the differential sensitivity of native speakers and L2 learners of English to phonetic issues, particularly the consonant contexts in which vowels appear. Two other studies focus specifically on (methodological) aspects of second language learning. The contribution by [Yu](#) reviews the positive and negative impacts of technology on reducing the anxiety experienced by L2 learners of Chinese and includes some methodological recommendations for future research on this topic. In line with this contribution, the article by [Yu](#) reviews the potential of positive psychology for the improvement of L2 learning by reducing learning anxiety.

Finally, the contribution by [Imaezue et al.](#) is a clinical linguistics study that urges the adoption of more inclusive ways of treating aphasia, particularly those that are more sensitive to linguistic diversity.

Author contributions

AB-B: Conceptualization, Investigation, Writing – original draft. AB: Conceptualization, Investigation, Writing – original draft. TS: Conceptualization, Investigation, Writing – original draft.

Acknowledgments

The authors would like to thank all the contributors to this volume (both authors and reviewers).

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The authors AB-B, AB, and TS declared that they were an editorial board member of *Frontiers*, at the time of submission. This had no impact on the peer review process and the final decision.

Generative AI statement

The author(s) declared that Generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by *Frontiers* with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.



OPEN ACCESS

EDITED BY

Thomas L. Spalding,
University of Alberta, Canada

REVIEWED BY

Zdena Kralova,
Constantine the Philosopher University,
Slovakia

*CORRESPONDENCE

Qiangfu Yu
✉ yuqiangfu@xaut.edu.cn

RECEIVED 15 March 2025

ACCEPTED 03 April 2025

PUBLISHED 25 April 2025

CITATION

Yu Q (2025) Examining the relationship
between foreign language anxiety and foreign
language enjoyment from the perspective of
positive psychology.
Front. Psychol. 16:1594282.
doi: 10.3389/fpsyg.2025.1594282

COPYRIGHT

© 2025 Yu. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Examining the relationship between foreign language anxiety and foreign language enjoyment from the perspective of positive psychology

Qiangfu Yu*

Faculty of Humanities and Foreign Languages, Xi'an University of Technology, Xi'an, China

Negative emotions represented by foreign language anxiety (FLA) have long been an important topic in the study of second language acquisition (SLA), while the influence of positive emotions on foreign language learning (FLL) has long been neglected. Since Positive Psychology (PP) was introduced into the field of SLA, research on learners' positive and negative emotions in SLA based on the perspective of PP has begun to flourish. Particularly, the correlation between foreign language enjoyment (FLE) and FLA has become a hot topic. Based on the analysis of the literature on FLA and FLE, it was found that the research on FLA and FLE mainly focuses on the following categories: correlation between FLA and FLE, influencing factors of FLA and FLE, correlation of FLA and FLE with academic achievements, and dynamicity of FLA and FLE. This study summarized the major findings in the four research hotspots and concluded with some research lacunae and possible directions for future research on FLA and FLE from the perspective of PP.

KEYWORDS

foreign language anxiety, foreign language enjoyment, foreign language learning, foreign language learners, academic achievements, positive psychology

Introduction

Since the 1970s, emotions have been a hot topic in the field of SLA. However, a plethora of studies focus on the negative emotions, among which FLA is the most studied emotion in the past few decades (Scovel, 1978; Horwitz, 2001, 2010; MacIntyre, 2017). In recent years, with the introduction of PP into the field of SLA (Mercer and MacIntyre, 2014), positive emotions represented by FLE has gradually attracted the attention of researchers (Dewaele and MacIntyre, 2014, 2016, 2019; Dewaele et al., 2016; Dewaele et al., 2017; Mierzwa, 2020; Özer and Altay, 2021). PP favors a more holistic view of emotions of human beings, which means that instead of focusing solely on negative emotions, such as FLA, learners' positive emotions, such as FLE, are considered equally (Dewaele and MacIntyre, 2014; Dewaele et al., 2016; Saito et al., 2018). Therefore, research on the positive and negative emotions of learners in SLA from the perspective of PP is flourishing, among which the relationship between FLA and FLE and its related variables has become a hot research topic (Dewaele and MacIntyre, 2014, 2016; Dewaele and Dewaele, 2017, 2020; Dewaele et al., 2017, 2019; Jiang and Dewaele, 2019; Li, 2020a; Li et al., 2021; Li and Han, 2022). However, as a relatively new research topic in SLA, studies on FLA and FLE combined are relatively rare, compared with studies on the single emotion of FLA or FLE. The present study aims to investigate the research status of FLA and FLE in the past decade by analyzing the research hotspots, and to provide some constructive suggestions for future research, based on the research lacunae of the current research.

Literature review

FLA

Anxiety is a subjective feeling of tension and worry associated with the arousal of the autonomic nervous system (Spielberger et al., 1983), a fear without an identifiable threat (Ohman, 2008). Boudreau et al. (2018) pointed out that anxiety is essentially a diffused form of fear. FLA is considered to be anxiety experienced by individuals in the course of learning a foreign language (FL) in a classroom or in any situation in which a FL is used (Gardner and MacIntyre, 1993). Horwitz et al. (1986) defined FLA as a unique complex structure of self-cognition, belief, feeling and behavior related to classroom learning in the process of FLL, and designed the Foreign Language Classroom Anxiety Scale (FLCAS) to help language teachers and scholars better understand students' FLA. Gregersen and MacIntyre (2014) described FLA as a negative emotion of worry and anxiety caused by learning and using a second language in a classroom that requires self-expression. Overall, FLA is a kind of negative emotion in which learners feel frustrated, fearful and worried about FLL, including a series of anxiety related to specific language skills (Horwitz, 2016), and it is one of the strongest predictors of success or failure in FLL (Horwitz, 2010).

Studies on FLA mostly focus on the development and measurement of FLA scales (Horwitz et al., 1986; Wang, 2003; Dewaele and MacIntyre, 2014; Dewaele and Alfawzan, 2018; Dewaele et al., 2019a; Du, 2019; Jiang and Dewaele, 2019), the correlation between FLA and academic achievement (Gardner, 1985; Horwitz et al., 1986; Gardner and MacIntyre, 1993; Aida, 1994; Elkhafafi, 2005; Horwitz, 2010; Dewaele and MacIntyre, 2014; MacIntyre, 2017; Mierzwa, 2020; Dong, 2021) and influencing factors of FLA (Dewaele, 2013; Gregersen and MacIntyre, 2014; Oxford, 2017; MacIntyre, 2017; Dewaele et al., 2019a; Dewaele and MacIntyre, 2019; Jiang and Dewaele, 2019).

FLE

Enjoyment is regarded as an individual's sense of achievement and satisfaction from the process and result(s) of some activity (Csikszentmihalyi, 1990; Ainley and Hidi, 2013). Martin Seligman (2012), the founder of PP, proposed PERMA (positive emotion, engagement, relationships, meaning, and accomplishment) model from the perspective of PP. Enjoyment constitutes a positive psychological experience characterized by an elevated emotional state that emerges when individuals transcend homeostatic equilibrium. This adaptive process facilitates personal development and sustains well-being, distinguishing it from pleasure, a more immediate gratification mechanism that serves to fulfill biological drives such as hunger reduction or physical comfort (Seligman and Csikszentmihalyi, 2000). Dewaele and MacIntyre (2016) conceptualized enjoyment as a multifaceted emotional state arising from the dynamic interplay between task challenge and perceived competence, where individuals derive intrinsic satisfaction through accomplishing challengeable tasks. Succinctly put, rooted in PP, enjoyment serves as a foundational catalyst for personal growth, thereby providing theoretical grounding to integrate PP theories into FLL research.

FLE extends MacIntyre and Gregersen's (2012) application of PP principles to emotion regulation in second language acquisition. FLE, a relatively new concept, was proposed for the first time by Dewaele and MacIntyre (2014), who held that FLE can create a pleasant and reassuring psychological atmosphere for learners to explore unacquainted languages and cultures, thus promoting their FLL. Dewaele and MacIntyre (2016) further clarified that FLE is not only a positive emotion that promotes FLL, but also can alleviate the persistent effects of negative emotions by improving attention and awareness of FL input, and improve individuals' resilience in coping with FLL difficulties. Overall, FLE is a complex and relatively stable emotion (Boudreau et al., 2018), which enables learners to form good interpersonal relationships in FLL, fuel the FLL process (Dewaele, 2022), and make continuous progress toward their learning goals (Chen Wen et al., 2021).

Since Dewaele and MacIntyre (2014) first proposed the concept of FLE, many scholars have begun to shift their attention from the study of negative emotions represented by FLA to the study of positive emotions represented by FLE from the perspective of PP. Most studies on FLE focus on the development and measurement of FLE scales (Dewaele and MacIntyre, 2014, 2016; Dewaele and Dewaele, 2017; Dewaele and Alfawzan, 2018; Li et al., 2018; Jin and Zhang, 2018, 2019; Dewaele et al., 2019a, 2019b), the relationship between FLE and academic performance/achievements (Piechurska-Kuciel, 2017; Khajavy et al., 2017; Saito et al., 2018; Mierzwa, 2020; Dewaele and Alfawzan, 2018; Botes et al., 2021), influencing factors of FLE (Dewaele and MacIntyre, 2014, 2019; Dewaele and Dewaele, 2017; Dewaele et al., 2017, 2019; Dewaele and Alfawzan, 2018; Moskowitz and Dewaele, 2019; Jiang, 2020).

Methods

Firstly, previous studies on FLA and FLE published in international journals were searched in databases of Web of Science (WOS) in February, 2025. The author used the following searching parameters to conduct the search for previous studies: (((TS = (anxiety)) AND TS = (enjoyment)) AND TS = (foreign language)) AND PY = (2014–2024) and (((TS = (anxiety)) AND TS = (enjoyment)) AND TS = (second language)) AND PY = (2014–2024). Secondly, studies published in Chinese journals were searched in China National Knowledge Infrastructure (CNKI) database in February, 2025. The author used the Chinese counterparts of the searching parameters mentioned above and extracted the studies on FLA and FLE in Chinese Social Sciences Citation Index (CSSCI) source journals.

Studies were eligible for inclusion if they focused on FLA and FLE and if the relevant key words (e.g., anxiety, enjoyment, foreign language, second language) appeared in the titles or abstracts. We included all studies, both speculative or theoretical research design, empirical research design, qualitative and quantitative, in this review. Only CSSCI articles published in Chinese and WOS articles published in English were included. Then, the retrieved studies were evaluated for the eligibility by scanning and analyzing the full texts, and only studies focusing FLA and FLE were eligible for inclusion in

this review. Conference proceedings, editorials, book reviews and review articles were outside the scope of the study and therefore excluded from the data collection. The inclusion and exclusion criteria as well as the selection procedure are summarized in Figure 1.

Results and discussion

As a result, 31 papers were retrieved from Web of Science, and 15 from CSSCI. Based on iterative reading and analysis of the 46 retrieved

studies, the hotspots of research on FLA and FLE can be summarized under the following categories.

Correlation between FLA and FLE

From the perspective of PP, FLE aligns closely with the field's emphasis on strengths and well-being cultivating while FLA reflects challenges to psychological resilience, highlighting the complex interplay between FLA and FLE. Many studies retrieved focus on the correlation between the two significant affective factors. Some scholars

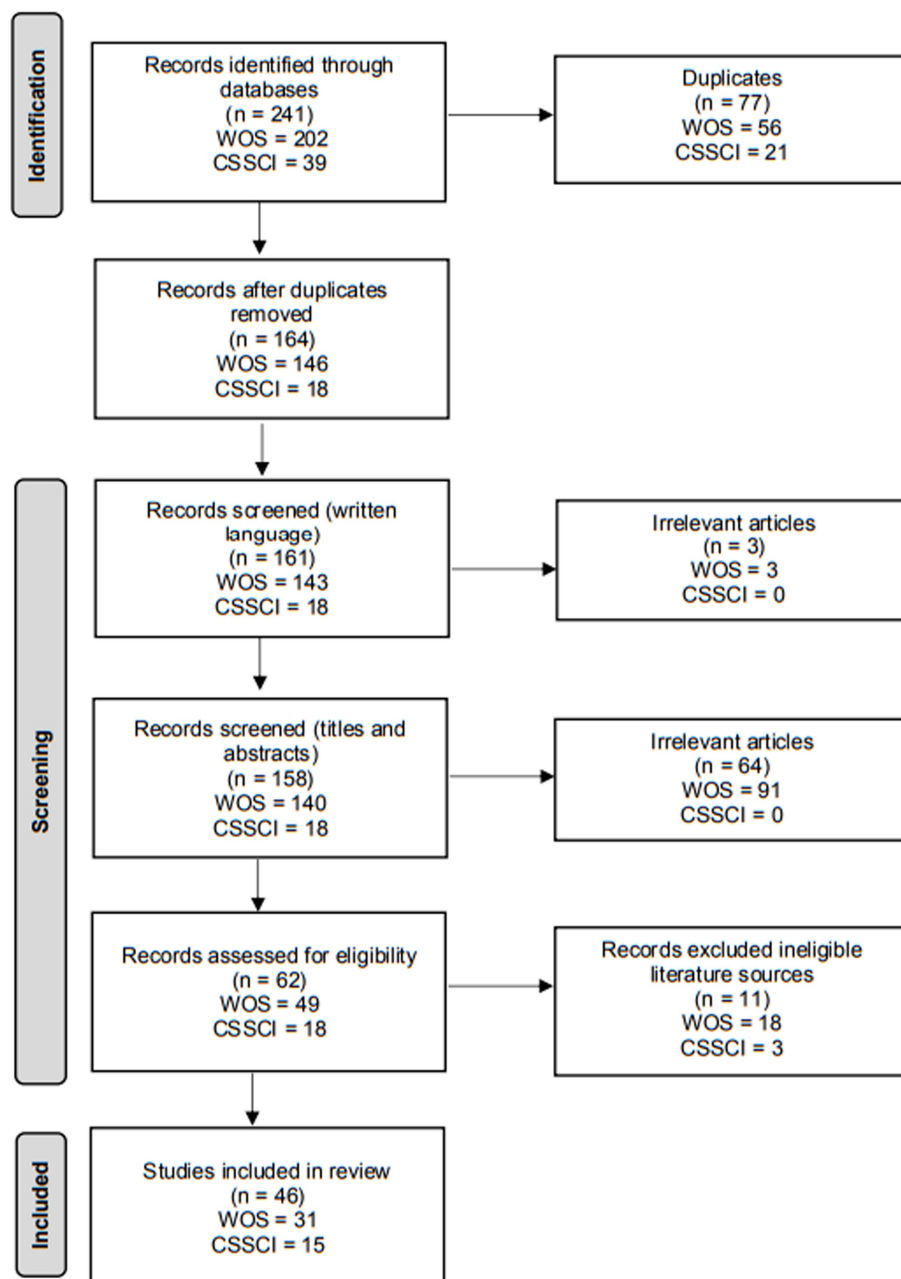


FIGURE 1
Flow diagram illustrating the process of article selection.

believed that there was a negative correlation between FLA and FLE (Maican and Cocorada, 2021). Dewaele and MacIntyre (2014) investigated the status quo of FLA and FLE among 1746 learners from different ages around the world and revealed the negative correlation between FLE and FLA, indicating that although FLE and FLA were related to a certain extent, they were independent in essence, which is of landmark significance to the development of PP. In a follow-up study, Dewaele and MacIntyre (2016) further explored the negative correlation between FLE and FLA, and verified the mutual independence of the two. Their findings resonate with PP's distinction between reducing pathology (e.g., anxiety) and cultivating well-being (e.g., enjoyment), suggesting dual pathways for optimizing FLL outcomes.

Notwithstanding many researchers championed the negative correlation between FLA and FLE, they were somewhat divided on the degree of negative correlation between FLA and FLE. Dewaele and Dewaele (2017) found that there was a weak negative correlation between FLA and FLE among 189 FL learners in two London secondary schools. Dewaele et al. (2017) also found a low negative correlation between FLA and FLE. However, Dewaele and MacIntyre (2019) found a moderate negative correlation between FLA and FLE among 750 FL learners. Similarly, Dewaele et al. (2019a) found a moderate negative correlation between FLA and FLE through a questionnaire survey among 210 Spanish learners of English as a FL. Notwithstanding, Shirvan and Taherian (2021) analyzed FLA and FLE among 371 undergraduates by using the potential growth curve model and triangular data collection method, and found that students' FLE level increased and FLA level decreased during the study period, and that FLA and FLE were significantly negatively correlated, which aligns with PP's upward spiral model, where momentary positive emotions accumulate to produce enduring resilience (Garland et al., 2010). Fang and Tang (2021) also found a significant negative correlation between FLA and FLE.

Contrary to dominant findings, some scholars also believed that there was a positive correlation between FLA and FLE. For example, Dewaele et al. (2019b) found a weak positive correlation between FLA and FLE among 592 undergraduates and middle school students in Kazakhstan who spoke Turkish as a FL, which reflects moderate anxiety might coexist with enjoyment during optimally challenging tasks, a phenomenon requiring nuanced PP frameworks to interpret.

Influencing factors of FLA and FLE

Dewaele and MacIntyre (2014) investigated FLA and FLE among 1746 learners of different ages around the globe and found the participants' age, gender, cultural background, FL level, education level and the number of FLs being learned were related to FLA and FLE. Based on the research above, Dewaele et al. (2016) specifically studied gender differences in FLA and FLE, and found that females experienced more emotions in FL classes, and their FLA and FLE were much higher than those of males. However, Wang (2023) found gender did not affect FLA and FLE of Chinese EFL learners. Dewaele et al. (2017) found that although the age of the participants had no correlation with FLA, it was positively correlated with FLE, and the girls had higher levels of FLA and

FLE. The study also revealed FLE was related to positive classroom relationships and FL, while lower FLA correlated with the attitudes toward FLL and peer support, which mirrors PP's broaden-and-build theory. Besides, Zhao and Zhang (2021) found no significant difference in FLA and FLE of English majors in different grades, and the students' FLA and FLE came from classroom activities, teachers' teaching methods, classmates' support and personal progress in English learning.

Correlation of FLA and FLE with academic achievements

Since FLA and FLE have an important effect on learning performance, they should be studied together from the perspective of PP to better explain learners' emotional experience (Özer and Altay, 2021). Dewey et al. (2018) investigated 36 overseas students and found a negative correlation between language proficiency and FLA and a positive correlation between language proficiency and FLE. Li et al. (2020) explored FLA/FLE effects among 1,307 Chinese students, finding FLA negatively predicted learning performance while FLE positively predicted it. Similarly, Mierzwa (2020) found that FLE was positively correlated with academic performance, while FLA was negatively correlated with academic performance.

Dewaele and Alfawzan (2018) used a combination of quantitative and qualitative methods to investigate FLA/FLE effects on FL achievement among FL students and found students with higher FLE had a lower level of FLA and a better academic performance. Moreover, it was emphasized that the positive correlation between FLE and academic performance was stronger than the negative correlation between FLA and academic performance. However, some scholars hold the opposite view, believing that FLA affected FL performance more than FLE. For example, Dewaele and Ergün (2020) investigated the correlation of FLA and FLE with academic performance of 110 Turkish middle school students learning English, and found that FLA was the strongest predictor of academic performance. Li (2020a) explored the relationship between students' trait emotional intelligence, emotions and academic performance among 1,307 grade-two high school students in China, and found that the students with high FLE felt less FLA and had better learning performance, and that the influence of FLA on learning performance was more significant than that of FLE.

The above studies only confirm the predictive effect of FLA and FLE on general FL achievement. Dewaele and Li (2018) explored the correlation of FLA and FLE with achievements in specific FL skills (listening, speaking, reading, writing, vocabulary and grammar), suggesting that future research should integrate positive emotion regulation strategies into skill-specific interventions from the perspective of PP.

Dynamicity of FLA and FLE

FLA and FLE are not completely opposed (Li and Han, 2022), mirroring PP's dual-factor model of mental health that emphasizes the co-existence of distress reduction and well-being enhancement,

and FLA and FLE are dynamic, so the relationship between them also changes dynamically (Dewaele and Dewaele, 2017), which deserves more dynamic research methods (Mierzwa, 2020). Boudreau et al. (2018) adopted dynamic systems theory to reveal the nonlinear interplay between FLA and FLE in second language communication, indicating that the dynamic relationship between FLA and FLE can be conceptualized and dynamically measured, which resonates with the emotion scaffolding concept in PP. By comparing the levels of FLA and FLE of the participants of different ages, Dewaele and Dewaele (2017) investigated how students' FLA and FLE as well as their independent variables evolved over time. It was found that there was little change in FLA while FLE increased slightly, and that the prediction of different variables on FLA and FLE also varied with time. Aubrey (2022) adopted the idiodynamic approach and found that the negative relationship between them fluctuated from strong to very weak. Of note, Pan and Zhang (2021) adopted the longitudinal study method to unpack the changes of FLA and FLE in FL classroom over time, and found that FLE was less stable than FLA as time went on, which further validated the findings of Dewaele and Dewaele (2017).

Conclusions and suggestions for future research

Although recent years witnessed more research on FLA and FLE since the introduction of PP into the field of language education, the research samples are mainly confined to college students and high school students. There is a scarcity of research focused on FLA and FLE of other student groups, such as FL beginner learners and FL learners with master degrees and PhD degrees. Meanwhile, FLA and FLE are different in different contexts. For example, Dewaele and MacIntyre (2014) found that the degree of FLA and FLE of Asian students was different from that of learners in other parts of the world. North American students experienced more FLE and less FLA, while Chinese students experienced a lower degree of FLE and a higher degree of FLA. This view is also supported by the findings of Jiang and Dewaele (2019) and Li (2020b). Besides, the levels of FLA and FLE of learners learning different FLs in the same country are also different. For example, in Belgium, English learners have significantly lower levels of FLA and higher levels of FLE than Dutch learners (De Smet et al., 2018). Therefore, in the light of PP, future studies on FLA and FLE should focus more attention on FL learners other than high school students and college students, FL learners in different contexts, and learners learning different FLs, especially non-common FLs.

With respect to research methods, most studies adopted quantitative research methods overwhelmingly represented by questionnaire survey. Future research on FLE and FLA should adopt both quantitative methods and qualitative methods to improve research reliability. Quantitative methods should employ context-based scales or questionnaires with good reliability and validity while qualitative methods can use classroom observation, video recording, student journals, field investigation, and particularly audio-recorded think-aloud protocols, instead of relying solely on semi-structured interviews. Kráľová et al. (2021) used qualitative research methods (questionnaire, interview and

observation) and quantitative research methods (scale) to discuss which learning activities would lead to positive or negative emotions among high school students learning English as a FL. This study combined the four research methods into the research design, offering certain reference value for future study. Meanwhile, some advanced techniques like Event-related Potentials, Positron Emission Tomography and functional Magnetic Resonance Imaging can be incorporated into future research to uncover neural mechanism of FLE and FLA by measuring the electromagnetic, blood flow and neuronal activities of research samples.

Finally, most of the studies were cross-sectional studies, taking FLA and FLE as relatively stable individual variation variables. However, from the perspective of PP, FLA and FLE of FL learners are dynamic, and the influencing factors are complex and diverse. Diachronic studies can more scientifically reveal the dynamic changes of FLA and FLE. In the future, longer diachronic studies can be conducted on the dynamicity of FLA and FLE of learners of a certain FL, on the comparison of FLA and FLE among learners learning different FLs, and on the correlation of FLA and FLE with other variables.

Author contributions

QY: Writing – review & editing, Writing – original draft.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported by grants from the Research Project of Education and Teaching Reform, Xi'an University of Technology (xjy2448) and the Research Project on Graduate Education and Teaching Reform, Xi'an University of Technology (252042342).

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Aida, Y. (1994). Examination of Horwitz, Horwitz, and Cope's construct of foreign language anxiety: the case of students of Japanese. *Mod. Lang. J.* 78, 155–168. doi: 10.1111/j.1540-4781.1994.tb02026.x
- Ainley, M., and Hidi, S. (2013). Interest and enjoyment. *Int. Handbook Emotions Educ.*, eds. R. Pekrun and L. Linnenbrink (New York: Routledge). 205–220. doi: 10.4324/9780203148211.ch11
- Aubrey, S. (2022). The relationship between anxiety, enjoyment, and breakdown fluency during second language speaking tasks: an idiodynamic investigation. *Front. Psychol.* 13:968946. doi: 10.3389/fpsyg.2022.968946
- Botes, E., Dewaele, J. M., and Greiff, S. (2021). Taking stock: an overview of the literature and a preliminary meta-analysis of foreign language enjoyment and other individual difference variables. *PsyArXiv* 1, 1–37. doi: 10.31234/osf.io/guaj5
- Boudreau, C., MacIntyre, P. D., and Dewaele, J. M. (2018). Enjoyment and anxiety in second language communication: an idiodynamic approach. *Stud. Second Lang. Learn. Teach.* 8, 149–170. doi: 10.14746/ssl.2018.8.1.7
- Csikszentmihalyi, M. (1990). Flow: the psychology of optimal experience. *J. Leis. Res.* 24, 93–94. doi: 10.1080/00222216.1992.11969876
- De Smet, A., Mettewie, L., Galand, B., Hiligsmann, P., and Van Mensel, L. (2018). Classroom anxiety and enjoyment in CLIL and non-CLIL: does the target language matter? *Stud. Second Lang. Learn. Teach.* 8, 47–71. doi: 10.14746/ssl.2018.8.1.3
- Dewaele, J. M. (2013). The link between foreign language classroom anxiety and psychoticism, extraversion and neuroticism among adult bi- and multilinguals. *Mod. Lang. J.* 97, 670–684. doi: 10.1111/j.1540-4781.2013.12036.x
- Dewaele, J. M. (2022). “Enjoyment” in The Routledge handbook of second language acquisition and individual difference. eds. S. Li, P. Hiver and M. Papi (New York: Routledge), 190–206.
- Dewaele, J. M., and Alfawzan, M. (2018). Does the effect of enjoyment outweigh that of anxiety in foreign language performance? *Stud. Second Lang. Learn. Teach.* 8, 21–45. doi: 10.14746/ssl.2018.8.1.2
- Dewaele, J. M., and Dewaele, L. (2017). The dynamic interactions in foreign language classroom anxiety and foreign language enjoyment of pupils aged 12 to 18. A Pseudo-longitudinal investigation. *J. Eur. Second Lang. Assoc.* 1, 12–22. doi: 10.22599/jesla.6
- Dewaele, J. M., and Dewaele, L. (2020). Are foreign language learners' enjoyment and anxiety specific to the teacher? An investigation into the dynamics of learners' classroom emotions. *Stud. Second Lang. Learn. Teach.* 10, 45–65. doi: 10.14746/ssl.2020.10.1.3
- Dewaele, J. M., and Ergün, A. L. P. (2020). How different are the relations between enjoyment, anxiety, attitudes/motivation and course marks in pupils' Italian and English as foreign languages? *J. Eur. Second Lang. Assoc.* 4, 45–57. doi: 10.22599/jesla.65
- Dewaele, J. M., and Li, C. C. (2018). Emotions in SLA. *Stud. Second Lang. Learn. Teach.* 8, 15–19. doi: 10.14746/ssl.2018.8.1.1
- Dewaele, J. M., and MacIntyre, P. D. (2014). The two faces of Janus? Anxiety and enjoyment in the foreign language classroom. *Stud. Second Lang. Learn. Teach.* 4, 237–274. doi: 10.14746/ssl.2014.4.2.5
- Dewaele, J. M., and MacIntyre, P. D. (2016). “Foreign language enjoyment and foreign language classroom anxiety. The right and left feet of FL learning?” in Positive psychology in SLA. eds. P. MacIntyre, T. Gregersen and S. Mercer (Bristol: Multilingual Matters), 215–236. doi: 10.21832/9781783095360-010
- Dewaele, J. M., and MacIntyre, P. D. (2019). “The predictive power of multicultural personality traits, learner and teacher variables on foreign language enjoyment and anxiety” in Evidence-based second language pedagogy: A collection of instructed second language acquisition studies. eds. M. Sato and S. Loewen (London: Routledge), 263–286.
- Dewaele, J. M., MacIntyre, P. D., Boudreau, C., and Dewaele, L. (2016). Do girls have all the fun? Anxiety and enjoyment in the foreign language classroom. *Theory Practice Second Language Acquisition* 1, 41–63.
- Dewaele, J. M., Magdalena, A. F., and Saito, K. (2019a). The effect of perception of teacher characteristics on Spanish EFL learners' anxiety and enjoyment. *Mod. Lang. J.* 103, 412–427. doi: 10.1111/modl.12555
- Dewaele, J. M., Özdemir, C., Karci, D., Uysal, S., Özdemir, E. D., and Balta, N. (2019b). How distinctive is the foreign language enjoyment and foreign language classroom anxiety of Kazakh learners of Turkish? *Applied Linguistics Review* 13, 243–265. doi: 10.1515/applirev-2019-0021
- Dewaele, J. M., Witney, J., Saito, K., and Dewaele, L. (2017). Foreign language enjoyment and anxiety: the effect of teacher and learner variables. *Lang. Teach. Res.* 22, 676–697. doi: 10.1177/1362168817692161
- Dewey, D. P., Belnap, R. K., and Steffen, P. (2018). Anxiety: stress, foreign language classroom anxiety, and enjoyment during study abroad in Amman, Jordan. *Annu. Rev. Appl. Linguist.* 38, 140–161. doi: 10.1017/S0267190518000107
- Dong, L. (2021). A meta-analysis on Chinese EFL learners' foreign language anxiety and foreign language learning outcomes. *Foreign Language World* 1, 54–61.
- Du, X. (2019). The reevaluation of the foreign language classroom anxiety scale (FLCAS). *Proceed. 2nd Int. Conf. Humanities Educ. Soc. Sci.* Dordrecht: Atlantis Press. 626–631. doi: 10.2991/ichess-19.2019.130
- Elkhafaifi, H. (2005). Listening comprehension and anxiety in the Arabic language classroom. *Mod. Lang. J.* 89, 206–220. doi: 10.1111/j.1540-4781.2005.00275.x
- Fang, F., and Tang, X. (2021). The relationship between Chinese English major students' learning anxiety and enjoyment in an English language classroom: a positive psychology perspective. *Front. Psychol.* 12:705244. doi: 10.3389/fpsyg.2021.705244
- Gardner, R. C. (1985). Social psychology and second language learning: The role of attitudes and motivation. London: Edward Arnold.
- Gardner, R. C., and MacIntyre, P. D. (1993). On the measurement of affective variables in second language learning. *Lang. Learn.* 43, 157–194. doi: 10.1111/j.1467-1770.1992.tb00714.x
- Garland, E. L., Fredrickson, B., Kring, A. M., Johnson, D. P., Meyer, P. S., and Penn, D. L. (2010). Upward spirals of positive emotions counter downward spirals of negativity: insights from the broaden-and-build theory and affective neuroscience on the treatment of emotion dysfunctions and deficits in psychopathology. *Clin. Psychol. Rev.* 30, 849–864. doi: 10.1016/j.cpr.2010.03.002
- Gregersen, T., and MacIntyre, P. D. (2014). Capitalizing on language learners' individuality: from premise to practice. Bristol: Multilingual Matters. doi: 10.21832/9781783091218
- Horwitz, E. K. (2001). Language anxiety and achievement. *Annu. Rev. Appl. Linguist.* 21, 112–126. doi: 10.1017/S0267190501000071
- Horwitz, E. K. (2010). Foreign and second language anxiety. *Lang. Teach.* 43, 154–167. doi: 10.1017/S026144480999036X
- Horwitz, E. K. (2016). Factor structure of the foreign language classroom anxiety scale: comment on park (2014). *Psychol. Rep.* 119, 71–76. doi: 10.1177/0033294116653368
- Horwitz, E. K., Horwitz, M. B., and Cope, J. (1986). Foreign language classroom anxiety. *Mod. Lang. J.* 70, 125–132. doi: 10.1111/j.1540-4781.1986.tb05256.x
- Jiang, Y. (2020). An investigation of the effect of teacher on Chinese university students' foreign language enjoyment. *Foreign Language World* 1, 60–68.
- Jiang, Y., and Dewaele, J. M. (2019). How unique is the foreign language classroom enjoyment and anxiety of Chinese EFL learners? *System* 82, 13–25. doi: 10.1016/j.system.2019.02.017
- Jin, Y., and Zhang, L. J. (2018). The dimensions of foreign language classroom enjoyment and their effect on foreign language achievement. *Int. J. Biling. Educ. Biling.* 24, 948–962. doi: 10.1080/13670050.2018.1526253
- Jin, Y., and Zhang, L. J. (2019). A comparative study of two scales for foreign language classroom enjoyment. *Percept. Mot. Skills* 126, 1024–1041. doi: 10.1177/0031512519864471
- Khajavy, G. H., MacIntyre, P. D., and Barabadi, E. (2017). Role of the emotions and classroom environment in willingness to communicate. *Stud. Second Lang. Acquis.* 40, 605–624. doi: 10.1017/S0272263117000304
- Kráľová, Z., Kovacikova, E., Repova, V., and Škorvavová, E. (2021). Activities in English classes inducing positive/negative emotions. *Educ. Sci. J.* 23, 136–155. doi: 10.17853/1994-5639-2021-1-136-155
- Li, C. (2020a). A positive psychology perspective on Chinese EFL students' trait emotional intelligence, foreign language enjoyment and EFL learning achievement. *J. Multiling. Multicult. Dev.* 41, 246–263. doi: 10.1080/01434632.2019.1614187
- Li, C. (2020b). Emotional intelligence and English achievement: the mediating effects of enjoyment, anxiety and burnout. *Foreign Language World* 1, 69–78.
- Li, C., Dewaele, J. M., and Jiang, G. (2020). The complex relationship between classroom emotions and EFL achievement in China. *Applied Linguistics Review* 11, 485–510. doi: 10.1515/applirev-2018-0043
- Li, C., and Han, Y. (2022). The predictive effects of foreign language enjoyment, anxiety, and boredom on learning outcomes in online English classrooms. *Modern Foreign Languages* 2, 207–219.
- Li, C., Huang, J., and Li, B. (2021). The predictive effects of classroom environment and trait emotional intelligence on foreign language enjoyment and anxiety. *System* 96:102393. doi: 10.1016/j.system.2020.102393
- Li, C., Jiang, G., and Dewaele, J. M. (2018). Understanding Chinese high school students' foreign language enjoyment: validation of the Chinese version of the foreign language enjoyment scale. *System* 76, 183–196. doi: 10.1016/j.system.2018.06.004
- MacIntyre, P. D. (2017). “An overview of language anxiety research and trends in its development” in New insights into language anxiety: Theory, research and educational implications. eds. C. Gkonou, M. Daubney and J. M. Dewaele (Bristol: Multilingual Matters), 11–30.
- MacIntyre, P. D., and Gregersen, T. (2012). Emotions that facilitate language learning: the positive broadening power of the imagination. *Stud. Second Lang. Learn. Teach.* 2, 193–213. doi: 10.14746/ssl.2012.2.2.4
- Maican, M. A., and Cocorada, E. (2021). Online foreign language learning in higher education and its correlates during the COVID-19 pandemic. *Sustain. For.* 13:781. doi: 10.3390/su13020781
- Mercer, S., and MacIntyre, P. D. (2014). Introducing positive psychology to SLA. *Stud. Second Lang. Learn. Teach.* 4, 153–172. doi: 10.14746/ssl.2014.4.2.2
- Mierzwa, E. (2020). Anxiety and enjoyment in the foreign language classroom. *East West Cultural Passage* 19, 107–121. doi: 10.2478/ewcp-2019-0007

- Moskowitz, S., and Dewaele, J. M. (2019). Is teacher happiness contagious? A study of the link between perceptions of language teacher happiness and student attitudes. *Innov. Lang. Learn. Teach.* 15, 117–130. doi: 10.1080/17501229.2019.1707205
- Ohman, A. (2008). “Fear and anxiety” in Handbook of emotions. eds. M. Lewis, J. M. Haviland-Jones and L. F. Barrett. 3rd ed (New York: Guilford), 709–729.
- Oxford, R. L. (2017, 2017). “Anxious language learners can change their minds: ideas and strategies from traditional psychology and positive psychology” in New insights into language anxiety: Theory, research and educational implications. eds. C. Gkonou, M. Daubney and J. M. Dewaele (Bristol: Multilingual Matters), 177–197.
- Özer, Z., and Altay, İ. F. (2021). Examining the level of enjoyment and anxiety among Turkish EFL students. *J. Lang. Linguistic Studies* 17, 663–671. doi: 10.17263/jlls.903532
- Pan, C., and Zhang, X. (2021). A longitudinal study of foreign language anxiety and enjoyment. *Lang. Teach. Res.* 27, 1552–1575. doi: 10.1177/1362168821993341
- Piechurska-Kuciel, E. (2017). “L2 or L3? Foreign language enjoyment and proficiency” in *Multiculturalism, multilingualism and the self: Studies in linguistics and language Learning*. eds. D. Gabrys-Barker and D. Gałajda (Spring: Cham), 97–111.
- Saito, K., Dewaele, J. M., and Abe, M. (2018). Motivation, emotion, learning experience and second language comprehensibility development in classroom settings: a cross-sectional and longitudinal study. *Lang. Learn.* 68, 709–743. doi: 10.1111/lang.12297
- Scovel, T. (1978). The effect of affect on foreign language learning: a review of the anxiety research. *Lang. Learn.* 28, 129–142. doi: 10.1111/j.1467-1770.1978.tb00309.x
- Seligman, M. (2012). *Flourish: A visionary new understanding of happiness and well-being*. New York: Atria Books.
- Seligman, M., and Csikszentmihalyi, M. (2000). Positive psychology: an introduction. *Am. Psychol.* 55, 5–14. doi: 10.1037/0003-066X.55.1.5
- Shirvan, M. E., and Taherian, T. (2021). Longitudinal examination of university students’ foreign language enjoyment and foreign language classroom anxiety in the course of general English: latent growth curve modeling. *Int. J. Biling. Educ. Biling.* 24, 31–49. doi: 10.1080/13670050.2018.1441804
- Spielberger, C. D., Gorsuch, R., Lushene, R. E., and Vagg, P. R. (1983). *Manual for the state-trait anxiety inventory (form Y1 – Y2)*. Palo Alto: Consulting Psychogysts Press.
- Wang, C. K. (2003). The adaptation and validation of the foreign language classroom anxiety scale when applied to Chinese college students. *Psychol. Sci.* 2, 281–284.
- Wang, Z. (2023). Psychological factors and production behaviors of Chinese undergraduate EFL learners. *PLoS One* 18:e0288014. doi: 10.1371/journal.pone.0288014
- Wen, C., Liu, G., and Li, J. (2021). An empirical study of Chinese L2 learners’ foreign language enjoyment and anxiety. *J. Yunnan Normal Univ.* 1, 44–53. doi: 10.16802/j.cnki.ynsddw.2021.01.006
- Zhao, W., and Zhang, X. (2021). A study of foreign language anxiety and enjoyment of English major college students. *Modern Commun.* 7, 169–173.



OPEN ACCESS

EDITED BY

Antonio Bova,
Catholic University of the Sacred Heart, Italy

REVIEWED BY

Ugur Özbilen,
Akdeniz University, Türkiye
Prajit Dhar,
University of Potsdam, Germany

*CORRESPONDENCE

Mohd Effendi Ewan Mohd Matore
✉ effendi@ukm.edu.my

RECEIVED 04 April 2025

ACCEPTED 28 May 2025

PUBLISHED 20 June 2025

CITATION

Mohd Noh MF, Mohd Matore MEE and
Sulaiman NA (2025) A cross-age odyssey of
cognitive reading attributes: a scoping review.
Front. Psychol. 16:1605898.
doi: 10.3389/fpsyg.2025.1605898

COPYRIGHT

© 2025 Mohd Noh, Mohd Matore and
Sulaiman. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

A cross-age odyssey of cognitive reading attributes: a scoping review

Muhamad Firdaus Mohd Noh¹,
Mohd Effendi Ewan Mohd Matore^{2*} and Nur Ainil Sulaiman³

¹Faculty of Education, Universiti Kebangsaan Malaysia, Bangi, Selangor, Malaysia, ²Research Centre of Education Leadership and Policy, Faculty of Education, Universiti Kebangsaan Malaysia, Bangi, Selangor, Malaysia, ³Centre of Teaching and Learning Innovation, Faculty of Education, Universiti Kebangsaan Malaysia, Bangi, Selangor, Malaysia

Reading is a fundamental cognitive-linguistic process that involves the dynamic interaction of multiple interrelated cognitive and perceptual mechanisms. Existing reading models are often limited in fully capturing the intricate relationships between reading attributes across different age groups. This review aims to compare the cognitive attributes utilized in reading assessments for young and adult readers. Using a scoping review methodology, the study analyzed 47 empirical studies selected through a systematic search of reputable academic databases, Scopus, Web of Science, and Google Scholar, from an initial pool of 331 publications. The selection followed a screening process based on three inclusion and exclusion criteria: types of publications, language, and skills assessed. A distinctive pattern emerges in the assessment of cognitive reading skills across age groups, with adult readers evaluated on a wider array of attributes that encompass both fundamental skills and higher order cognitive abilities. In contrast, young readers' assessments tend to center on a narrower spectrum of subskills, primarily emphasizing literal and interpretive comprehension. This developmental pattern calls for the refinement of existing assessment models to better capture the progressive nature of cognitive reading development. Tailoring assessment tools and instructional strategies to align with learners' cognitive demands is imperative for internal stakeholders, while external stakeholders are urged to develop age-appropriate assessments. Future research could address the study's limitations by exploring advanced technologies, such as eye tracking, conducting rigorous reviews, performing cross-linguistic comparative studies, and evaluating diverse assessment methods to enhance the accuracy, effectiveness, and generalizability of reading assessments across various contexts.

KEYWORDS

reading skills, cognitive reading attributes, language assessment, scoping review, cognitive diagnostic

1 Introduction

Reading is a fundamental cognitive-linguistic process that involves the dynamic interaction of multiple interrelated cognitive and perceptual mechanisms. As a core aspect of language comprehension, reading requires the coordination of lexical access, working memory, syntactic parsing, and semantic integration, enabling individuals to construct meaning from written texts (Hosseini Alast and Baleghizadeh, 2024). Unlike other polytomous language skills such as speaking and writing, which involve productive

language use and require rater involvement for assessment (Mohd Noh and Mohd Matore, 2022, 2020), reading is a visually mediated process that necessitates efficient orthographic, phonological, and semantic activation to support fluent comprehension (Teng and Mizumoto, 2024). The psychological mechanisms underlying reading are deeply interconnected with broader cognitive functions, including executive control, attentional regulation, and memory retrieval, which contribute to individual differences in reading proficiency. Due to its complexity and the diverse theoretical perspectives that frame it, including developmental, cognitive, pedagogical, and computational frameworks, reading requires a comprehensive assessment approach capable of capturing its full range of cognitive attributes and skills. Within the psychology of language, multiple theoretical frameworks have been proposed to explain how cognitive and linguistic processes converge during reading. Classic reading models, such as psycholinguistic reading model (Goodman, 1967) and information processing model (Gough, 1972) provide insights into how readers derive meaning from texts. More recent perspectives such as the interactive-compensatory model (Stanovich, 1980). More recent models, like the cognitive processing model (Khalifa and Weir, 2009) and the direct and indirect effects model of reading (Kim, 2020), emphasize the hierarchical and interactive nature of reading, where multiple cognitive and contextual skills directly and indirectly contribute to comprehension.

While these models offer valuable insights into reading mechanisms, they often overlook developmental variations in cognitive reading processes. The way young and adult readers engage with text differs not only in terms of linguistic complexity but also in the cognitive operations employed during comprehension. Research has shown that adult readers rely more on higher-order cognitive functions, such as inference-making and semantic integration, while young readers focus more on basic decoding and literal comprehension (Siew et al., 2022). However, existing reading models do not always explicitly address the gradual shifts in cognitive reading attributes throughout the lifespan. To address this gap, this review examines how cognitive reading attributes develop across different age groups by synthesizing findings from empirical studies. Another key limitation in traditional reading models is their reliance on broad theoretical constructs rather than fine-grained, diagnostic insights into individual cognitive mechanisms. Cognitive Diagnostic Assessment (CDA) emerges as an innovative framework capable of providing such granularity, offering a psychometric approach that systematically examines specific cognitive reading attributes. Distinct from conventional reading assessments, which yield general single proficiency scores, CDA enables a detailed diagnosis of strengths and weaknesses in core reading processes. This methodological advancement aligns with contemporary perspectives in reading cognition, developmental psycholinguistics, and the psychology of language by emphasizing a learner-specific approach to reading assessment. Despite its potential, research integrating CDA into developmental reading studies remains underexplored. While existing CDA studies have successfully retrofitted high-stakes assessments to diagnose reading attributes (Safari and Ahmadi, 2023), fewer studies have systematically applied CDA to trace the evolution of cognitive reading attributes

across different age groups. Thus, the current review aims to identify and compare the reading attributes used in previous research to measure the reading skills of individuals from different age groups.

2 Literature review

2.1 Reading models

The diversity of reading models reflects various theoretical perspectives on how individuals engage with and comprehend texts. Table 1 presents the different models of reading (Goodman, 1967). Early conceptualization of reading as a psycholinguistic guessing game suggests that readers rely on contextual cues, linguistic structure, and prior knowledge to make predictions during reading, emphasizing the role of higher-level cognitive processes in guiding comprehension. This model exemplifies a top-down approach to reading, where meaning is primarily constructed through the reader's background knowledge and cognitive expectations rather than solely relying on the text itself. In contrast, Gough's (1972) and LaBerge and Samuels's (1974) stage-based models also known as the information processing model describe reading as a sequential process where readers decode text progressively, from letters to words to complete comprehension, highlighting the role of automaticity in word recognition and information processing. These models represent a bottom-up approach to reading, where comprehension is built from the accurate and systematic decoding of smaller units of text, such as letters and words before higher-level meaning can be constructed.

Apart from these stage-based reading models, component-based models, such as the sociocultural reading model, describe how a reader's background knowledge, conceptual abilities, and process strategies work together to create comprehension (Coady, 1979). Similarly, the Simple View of Reading conceptualizes reading comprehension as the product of two core components: decoding and linguistic comprehension, emphasizing that both are necessary for proficient reading (Hoover and Gough, 1990). This model highlights that weaknesses in either decoding or comprehension can hinder overall reading ability, underscoring the importance of balanced instruction targeting both skills. Meanwhile, the interactive reading model integrates both top-down and bottom-up processes, emphasizing a compensatory mechanism where strengths in one cognitive process balance weaknesses in another (Stanovich, 1980; Rumelhart, 1977). Rumelhart's (1977) interactive model synthesizes these perspectives by proposing that reading involves a dynamic interaction of both text-driven (perceptual) and knowledge-driven (cognitive) processes occurring simultaneously. Whereas, Stanovich's (1980) interactive-compensatory model extends the idea by suggesting that strengths in one cognitive process, such as contextual knowledge, can offset weaknesses in another, such as phonological decoding.

Expanding on these foundational models, the cognitive processing model of reading (Khalifa and Weir, 2009) integrates cognitive psychology perspectives to describe reading as a dynamic process involving both bottomup and top-down processing. It

TABLE 1 Types of reading models and their characteristics.

Type of reading models	Characteristics
Psycholinguistic reading model (Goodman, 1967)	Reading is an active process where meaning is constructed from the overall context and prior knowledge, with the reader forming and testing hypotheses about the text.
Information processing model (Gough, 1972)	Reading involves sequential processing from recognizing individual letters to understanding the entire text, with all steps working together for comprehension.
Automatic processing model (LaBerge and Samuels, 1974)	Reading is an automatic process where word stimuli are transformed into meaning through several memory systems.
Three-component reading model or socio-cultural reading model (Coady, 1979)	Reading comprehension is influenced by cognitive skills and sociocultural factors, depending on conceptual ability, background knowledge, and processing strategies.
Two-component reading model or simple view of reading model (SVR; Hoover and Gough, 1990)	Reading comprehension results from the development of word identification and linguistic comprehension over time.
Interactive reading model (Rumelhart, 1977)	Reading is a dynamic interaction between prior knowledge and the text, with top-down and bottom-up processes occurring simultaneously and relying on mental schemas.
Interactive-compensatory reading model (Stanovich, 1980)	Reading involves compensating for weaknesses in one area with strengths in another, using both bottom-up and top-down processes to construct meaning.

emphasizes the importance of goal setting during reading and distinguishes between local comprehension (understanding specific details) and global comprehension (grasping overall meaning), illustrating the complex interplay of cognitive skills required for effective reading. Similarly, the integrated reading model, also referred to as the Direct and Indirect Effects Model of Reading (DIER; Kim, 2020) proposes reading comprehension as an outcome of hierarchical, dynamic, and interactive contributions from multiple language, cognitive, and contextual skills. This model explains how some skills, such as word recognition, directly impact comprehension, while others, such as working memory and inferencing, provide indirect support by facilitating the integration of information. These models collectively aim to explain the complex, multi-stage nature of reading, emphasizing the interaction of perceptual, cognitive, and contextual mechanisms rather than rigid classifications, while acknowledging the dynamic and adaptive nature of reading comprehension.

2.2 Reading taxonomies and cognitive structures

Reading taxonomies provide systematic classification framework that organizes and categorizes the various cognitive skills, processes, and components involved in reading

comprehension. These reading taxonomies underscore the varied approaches to categorizing and assessing the intricate skills involved in reading, distinguishing them from reading models, which focus on the theoretical frameworks of the reading process itself. Each taxonomy emphasizes different aspects of reading comprehension and cognitive processes. Reading taxonomy by Davis (1968) focuses on sub-skills like recalling word meanings, making inferences, and understanding the author’s technique, which was empirically tested on students. Munby’s (1978) taxonomy, often used in curriculum development, outlines 19 sub-skills, including identifying main ideas and interpreting text through various cohesion devices, though it has been critiqued for lacking empirical validation. Heaton’s (1991) taxonomy provides a detailed list of 14 sub-skills, such as deducing word meaning and understanding conceptual meaning, aimed at assessing specific reading skills. Meanwhile, Hughes and Hughes’s (2020) divide reading skills into expedient and careful reading, further categorizing them into sub-skills like identifying discourse markers and interpreting complex sentences. Whereas, Anderson and Krathwohl (2001) updated version of Bloom’s Taxonomy introduces a cognitive dimension to reading by classifying skills into six levels: remembering, understanding, applying, analyzing, evaluating, and creating. This hierarchical model is echoed in Luebkke and Lorié (2013) divides reading skills into four sub-skills: identification, understanding and analyzing, making inferences, and application. Each of these taxonomies provides a different lens through which reading comprehension can be assessed, from the recall of explicit information to higher-order cognitive skills such as critical thinking and application. Collectively, these taxonomies highlight the complexity of reading comprehension and offer valuable tools for educators and researchers to assess and enhance reading skills in a structured and systematic manner.

2.3 The exploration of reading attributes through cognitive diagnostic assessment

An attribute refers to a specific cognitive skill or ability that contributes to task performance, particularly in educational contexts (Wang et al., 2021). Cognitive attributes are the foundational components underlying a learner’s ability to process, interpret, and respond to tasks, making them essential units for understanding academic performance (Zhang et al., 2024). Meanwhile, cognitive reading attributes are the cognitive skills involved in the process of reading and comprehending text. These attributes encompass a range of cognitive processes, including word recognition, decoding, inferencing, summarizing, comprehension monitoring, and understanding text structures (Li et al., 2021). Together, they influence a reader’s ability to extract, interpret, and construct meaning from written material. Understanding these attributes is crucial for assessing reading proficiency and guiding instructional strategies tailored to individual learner needs. The underlying sub-skills of reading have been explored in previous research studies using various approaches, including factor analysis (Nightingale et al., 2023; Lin, 2024), structural equation modeling (SEM) (Lee and Lee, 2023; Peng et al., 2021), item response theory (IRT) (Geramipour et al., 2021; Polat et al., 2022), and cognitive

diagnostic modeling (CDM) (Askari and Karami, 2024; Chen et al., 2023), each providing unique insights into the cognitive processes and relationships among reading components.

Cognitive reading attributes have been investigated within CDA principles by examining how specific cognitive skills contribute to reading comprehension and how they can be accurately measured (Mei and Chen, 2022). CDA represents a modern advancement in educational measurement by integrating psychological and measurement theories that offers insights into students' cognitive processes and addresses the limitations of conventional assessment methods (Maas et al., 2024). CDA provides a holistic and contextual understanding of individual cognitive processes by diagnosing specific cognitive reading attributes, drawing on insights from psychological theories and advanced measurement techniques (Ketabi et al., 2021). Other approaches often provide a narrow view of students' cognitive capabilities, falling short to diagnostically capture the intricacies of reading comprehension and the diverse cognitive reading attributes involved (von Davier and Lee, 2019).

The evolution of CDA is deeply rooted in the historical development of educational psychology and measurement theories. Early educational psychology prioritized intelligence measurement through specific tests, focusing on intellectual and academic skills (van der Linden, 2016). However, as understanding of human cognition expanded, theories such as cognitive development theory (Piaget, 1950) and information processing theory (Miller, 1956) highlighted the importance of cognitive processes like information handling, storage, and retrieval—key elements in reading comprehension. Concurrent advancements in measurement theories supported a more holistic and contextual approach to assessing these cognitive processes (Miller and Lovler, 2020). This integration ultimately led to the development of CDA models, which enable the exploration of cognitive reading attributes across various contexts, including language skills (Sessoms and Henson, 2018).

One common method used within CDA studies for investigating reading attributes is the retrofitting of existing high-stakes assessments for diagnostic purposes. Retrofitting involves reanalyzing existing reading tests to ensure they provide relevant and useful diagnostic information (Sessoms and Henson, 2018). This approach has been applied to large-scale assessments such as the Program for International Student Assessment (PISA; Chen and Chen, 2016) and the Test of English as a Foreign Language (TOEFL; Safari and Ahmadi, 2023), among others. Retrofitting allows educators to gain diagnostic insights from assessments not originally designed for this purpose, although challenges remain, such as ensuring a sufficient number of items to measure specific cognitive reading attributes accurately (Gierl and Cui, 2008). Despite the efficiencies of retrofitting, some researchers advocate for the development of entirely new CDA items and assessments tailored specifically to measure reading attributes. Although more resource intensive, this approach allows for a more detailed and personalized analysis of learners' cognitive strengths and weaknesses in reading. For example, a study in Turkey developed a diagnostic reading test for university students (Toprak and Cakir, 2021). Developing new CDA items ensures that the assessments are specifically designed to measure key cognitive reading attributes, such as decoding, inferencing, and

comprehension, providing more precise diagnostic information and supporting targeted instructional interventions (Ketabi et al., 2021).

3 Methods

3.1 Design

The current study employed a scoping review method as it is appropriate for gaining an in-depth understanding of a broad issue and charting existing research to better inform future studies (Munn et al., 2022). The key difference between a scoping review and a systematic literature review is that while a systematic review focuses on synthesizing specific evidence to answer a well-defined research question (Sabtu and Mohd Matore, 2024), a scoping review aims to map the existing literature on a broader topic, identifying gaps and trends without necessarily assessing the quality of individual studies (Nadmilail et al., 2023). It is particularly useful for analyzing emerging evidence when it is uncertain if more specific questions can be effectively addressed by a precise systematic review (Pollock et al., 2021). Scoping reviews are ideal for evaluating and understanding the extent of knowledge in a developing field, as well as for identifying, mapping, reporting, or discussing the characteristics and concepts within that area (Peters et al., 2022). The scoping review was conducted based on the guidelines recommended by Levac et al. (2010) and Arksey and O'Malley (2005) which involves five stages: (i) identifying the research objectives, (ii) identifying relevant studies, (iii) selecting the studies, (iv) charting the data, and (v) collating, summarizing, and reporting the results.

3.2 Review protocol

The review was guided by the Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR). The review process included the identification of relevant studies, their selection based on predefined criteria, and the abstraction and synthesis of key data. All screening and selection processes were conducted independently by two reviewers, following the PRISMA-ScR guidelines to minimize bias and ensure the reliability of the findings.

This rigorous process ensures the inclusion of high-quality evidence to address the research questions posed in this study.

3.3 Research question formulation

The research questions were formulated based on the Population, Interest and Context (PICO) framework (Lockwood et al., 2015). The PICO framework aids researchers in crafting appropriate research questions for reviews by focusing on the core elements (Mohamed Shaffril et al., 2021). The current review has applied these components by examining, (i) Population: readers across different age groups, (ii) Interest: cognitive reading

attributes, and (iii) Context: cognitive diagnostic assessment studies. As a result, the review is guided by three research questions:

1. What are the publication trends in cognitive diagnostic assessment studies on reading, specifically in relation to test types, sample demographics and age distributions?
2. How do cognitive diagnostic assessment studies conceptualize and operationalize key reading attributes for young and adult readers?
3. How do the similarities and differences in cognitive reading attributes assessed in young and adult readers reflect their cognitive and linguistic development?

3.4 Database

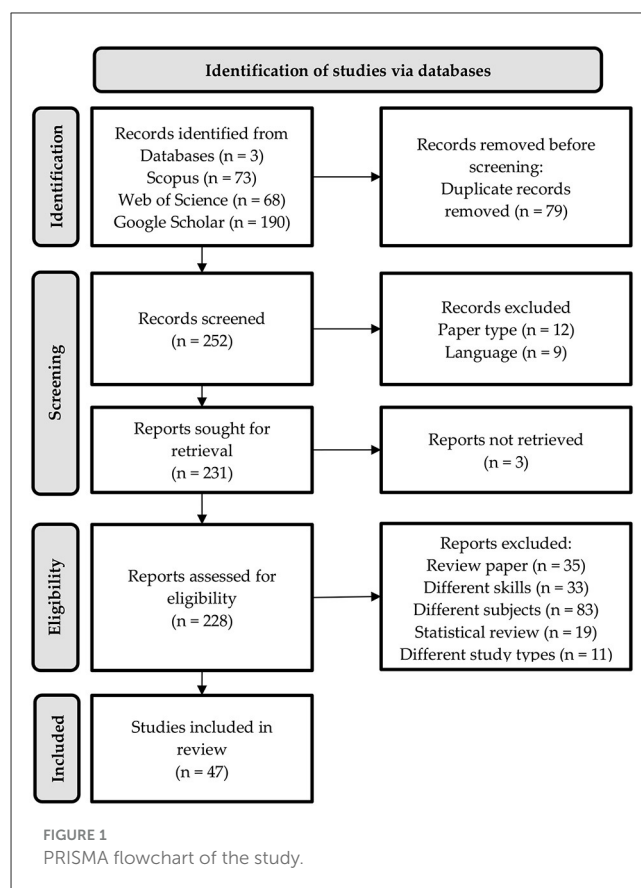
The review process began with the development of a systematic search strategy. This involved searching through three reputable academic databases, Scopus, Web of Science and Google Scholar. These databases have been used by previous researchers in reviewing the latest publication trends in various fields, providing wide coverage of peer-reviewed journals, conference papers, and other scholarly works (Sakaria et al., 2023). These databases were selected due to their extensive reach and credibility, ensuring that the review captures a broad spectrum of relevant and up-to-date research findings (Sabtu and Mohd Matore, 2024). Scopus, an esteemed database, aggregates content from over 30,000 journals representing a wide range of subject fields, including education, measurement, and language, enabling researchers to identify trends, gaps, and advancements in the field of study (Masdoki et al., 2021). This database, encompassing contributions from more than 11,000 publishers, undergoes a rigorous peer-review process. On the other hand, the Web of Science (WoS) is a platform providing access to a robust database across numerous academic disciplines (Singh et al., 2021). Originally established by the Institute for Scientific Information (ISI) and currently monitored by Clarivate Analytics, WoS covers over 30,000 journals spanning more than 250 disciplines (Birkle et al., 2020). Meanwhile, Google Scholar provides extensive access to a wide range of academic literature, including peer-reviewed articles, theses, books, and conference papers across various disciplines (Martín-Martín et al., 2021). Additionally, Google Scholar's extensive indexing of both traditional and gray literature ensures a thorough and inclusive search of relevant studies.

3.5 Systematic search strategies

This stage involves three steps; identification, screening, and eligibility, as illustrated in Figure 1.

3.5.1 Identification

The identification process begins with selecting appropriate keywords. To maximize the retrieval of relevant articles for the current review, the keywords are categorized into two groups: reading assessment and cognitive diagnostic assessment. Related terms for each keyword were identified and shown in Table 2. By



combining the keywords and related terms, the search process was conducted on the three databases with respective search strings as displayed in Table 3. The search on Scopus and Web of Science used a structured string incorporating Boolean operators, proximity operators, and wildcard symbols * to ensure relevant results. Boolean operators (OR) broadened the search by including synonyms like “reading assess*” OR “reading test*” OR “reading evaluat*”. Proximity operators (W/) refined the results by ensuring terms like “assess* W/reading” and “test* W/reading” appeared close together, maintaining relevance. Wildcards (*) captured word variations, such as “assess*” for “assess”, “assessment”, or “assessed”. Keywords like “cognitive diagnos*” and “classification AND diagnos*” ensured a focus on cognitive and diagnostic aspects. This strategy effectively targeted studies on reading assessment and cognitive diagnostics. Whereas, the search of the Google Scholar database was conducted using the Publish or Perish version 8.17 for Windows software, a tool designed to retrieve and analyze academic citations. This software allowed for efficient querying of Google Scholar using the same structured keywords and phrases as the Scopus and Web of Science searches. The search yielded a total of 331 publications, with 73 articles retrieved from Scopus, 68 from Web of Science, and 190 from Google Scholar.

3.5.2 Screening

Following the identification stage, 78 duplicate records were removed, leaving 253 records for initial screening. During the

screening stage, the records were assessed for relevance based on predefined inclusion and exclusion criteria as presented in Table 4. To ensure transparency and reduce the risk of selection bias, two independent reviewers conducted the title and abstract screening. Discrepancies between the reviewers were resolved through discussion until consensus was reached. If disagreements persisted, a third reviewer was consulted. The first criterion was the types of publications. Only publications on empirical studies were selected but not empirical studies using simulation data and other types of studies, including conceptual and review studies. The second criterion applied was language. Only publications in the English language were accepted. As a result, 21 records were excluded, with 12 records removed due to inappropriate paper types such as conceptual papers and book chapters, while nine records were excluded due to language limitations such as Chinese language and Persian language. This reduced the number of reports to 232, which were then sought for retrieval. However, three reports could not be retrieved, leaving a total of 229 reports for further evaluation.

TABLE 2 Keywords and related terms.

Keywords	Related terms
Reading assessment	Reading skill assessment, assessing reading, reading test, testing reading, evaluating reading
Cognitive diagnostic assessment	Cognitive diagnostic approach, cognitive diagnostic model, diagnostic classification model

TABLE 3 Search string.

Databases	Search strings
Scopus	((“reading assess*” OR “assess* W/reading” OR “reading test*” OR “test* W/reading” OR “reading evaluat*” OR “evaluat* W/reading” OR “reading comprehension” OR “skill-based assessment” OR “language assessment” OR “language testing”) AND (“cognitive diagnos*” OR “classification AND diagnos*”))
Web of Science	((“reading assess*” OR “assess* W/reading” OR “reading test*” OR “test* W/reading” OR “reading evaluat*” OR “evaluat* W/reading” OR “reading comprehension” OR “skill-based assessment” OR “language assessment” OR “language testing”) AND (“cognitive diagnos*” OR “classification AND diagnos*”))
Google Scholar	“cognitive diagnos*” AND “reading skill”

TABLE 4 Screening criteria.

No.	Criteria	Acceptance criteria	Rejection criteria
1.	Types of publication	Empirical studies with real data	Other than empirical studies with real data—empirical studies using simulation data, conceptual papers, review papers, method reviews
2.	Languages	English language	Other than English—publications reported in other than English language, studies on assessment of other than English language
3.	Focused skills	Reading skills	Other than reading skills—writing, speaking, listening skills

3.5.3 Eligibility

In the eligibility stage, the remaining reports were assessed in detail to determine their suitability for the review. Again, two reviewers independently assessed the full-text articles using the predefined inclusion and exclusion criteria. This dual-reviewer approach enhanced the methodological rigor of the review process. A total of 181 reports were excluded for various reasons: 35 were review papers, 33 focused on different skills, 83 addressed subjects outside the scope of the review such as Mathematics, diagnostic assessment and computerized adaptive testing (CAT), 19 were statistical reviews, and 11 followed different study types that did not meet the inclusion criteria, such as action research. Following this rigorous screening and eligibility process, 47 studies were deemed relevant and included in the final review. This multi-step process ensured the inclusion of only the most pertinent and high-quality studies, allowing for a robust and reliable scoping review of cognitive reading attributes across different age groups. The next step involved graphically presenting the selected articles. Summaries were created for each article, detailing the author, year, title of the study, methodologies, population, samples and test names. The final phase of this review framework included compiling, summarizing, and reporting the results. All data were stored and processed using Mendeley Desktop Version 1.19.8. A synthesis of the literature was then created by summarizing the key points and presenting them in text, tables, and figures. All 47 publications were thoroughly retrieved and reviewed in their entirety to ensure they met the inclusion criteria.

3.6 Systematic coding

Each publication was coded based on six variables: study design, test names, samples, sample types, sample ages, and cognitive attributes. Table 5 outlines the definitions and possible codes for each variable. The study design refers to the scientific framework of the study, categorized as either a retrofitting study or a true study. Test names indicate the reading assessments used, including wellknown standardized tests such as the Progress in International Reading Literacy Study (PIRLS) and the Program for International Student Assessment (PISA), as well as newly developed reading tests. The samples represent the test-takers, which include primary school students, high school students, undergraduate students, and postgraduate students. Sample types further classify test-takers based on their reading development stage, distinguishing between young and adult readers. Meanwhile, sample age specifies whether the exact age of the participants is mentioned or not. Lastly, cognitive attributes refer to the specific reading skills assessed in each study, such as finding explicit information (EXP),

TABLE 5 Article coding for the present review.

No.	Variable	Definition	Possible codes
1.	Study design	Scientific design of the study	Retrofitting study, true study
2.	Test names	The name of the test where the data was analyzed or the name of the newly developed test	Progress in International Reading Literacy Study (PIRLS), Program for International Student Assessment (PISA), Grade 6 provincial reading achievement assessment, newly developed reading test
3.	Samples	The test-takers	Primary school students, high school students, undergraduate students and postgraduate students
4.	Sample types	The classification of the samples based on the purpose of the test	Young reader, adult reader
5.	Sample age	The age of the test-takers	Exact age of the samples, age range, not mentioned
6.	Attributes	The sub-skills used or analyzed to assess reading skills of the study samples	Finding explicit information (EXP), making inferences (INF), generalizing or synthesizing main ideas (GEN), making interpretation of texts (INT), lexical or vocabulary knowledge (LEX), evaluating or analyzing the text (EVA)

TABLE 6 Reading tests and sample profiles in CDA studies involving young readers.

No.	Study	Reading tests	Sample types	Sample ages
1.	Yumsek (2023)	A large-scale K–12 ELP assessment	Secondary school students	17 years old
2.	Nallasamy and Khairani (2022)	A newly-developed cognitive diagnostic reading test	Primary school students	10 years old
3.	Toprak-yildiz (2021)	Progress in International Reading Literacy Study (PIRLS)	Primary school students	10 years old
4.	George and Robitzsch (2021)	Progress in International Reading Literacy Study (PIRLS)	Primary school students	10 years old
5.	Liu and Bian (2021)	A large-scale Grade 5 and 6 reading comprehension test	Primary school students	11–12 years old
6.	Fan and Yan (2020)	National Matriculation English Test(NMET)	Secondary school students	18 years old
7.	Thi and Loye (2019)	Progress in International Reading Literacy Study (PIRLS)	Primary school students	10 years old
8.	Chen and Chen (2016)	Program for International Student Assessment (PISA)	Secondary school students	15 years old
9.	Chen and Chen (2015)	Program for International Student Assessment (PISA)	Secondary school students	15 years old
10.	Jang et al. (2015)	Grade 6 provincial reading achievement assessment	Primary school students	12 years old
11.	Jang et al. (2013)	Grade 6 provincial reading achievement assessment	Primary school students	12 years old

making inferences (INF), generalizing or synthesizing main ideas (GEN), interpreting texts (INT), lexical or vocabulary knowledge (LEX), and evaluating or analyzing texts (EVA). By coding publications according to these variables, the review systematically organizes reading research, allowing for a structured analysis of trends, patterns, and gaps in the field. This approach enhances the understanding of how reading skills have been assessed across various studies and contexts. The preparation of data for presentation and synthesis, including handling missing summary statistics, data conversions, and visual display of individual study results, was conducted using Microsoft Excel 2021.

4 Results

4.1 RQ1: What are the publication trends in cognitive diagnostic assessment studies on reading, specifically in relation to test types, sample demographics and age distributions?

The analysis identified 11 articles focusing on young readers, with each study utilized different types of reading tests. Table 6

shows the reading tests and sample profiles used in the studies. Only one study developed new reading assessment items (Nallasamy and Khairani, 2022), while the other 10 studies retrofitted existing non-diagnostic reading assessment items to extract the diagnostic reports such as from a large-scale K–12 ELP assessment measuring the social and academic language (Yumsek, 2023), Progress in International Reading Literacy Study (PIRLS; Toprak-yildiz, 2021), Program for International Student Assessment (PISA) (Chen and Chen, 2016), Grade 6 provincial reading achievement assessment (Jang et al., 2013, 2015), Grade 5 and 6 reading comprehension test (Liu and Bian, 2021), and the National Matriculation English Test (NMET) (Fan and Yan, 2020). In regards to the types of the samples, seven of these studies focused on primary school students, while the remaining four targeted secondary school students, as shown in Figure 2. Among the primary school samples, four studies involved students aged ten (Nallasamy and Khairani, 2022; Thi and Loye, 2019), and three studies involved students aged 12 (Jang et al., 2013, 2015). Regarding the secondary school samples, two studies focused on students aged 15 (Chen and Chen, 2016, 2015), while one study each involved students aged 17 (Fan and Yan, 2020) and 18 (Yumsek, 2023), respectively.

A total of 36 articles sampling adult readers were further analyzed after the eligibility procedure. The reviewed studies

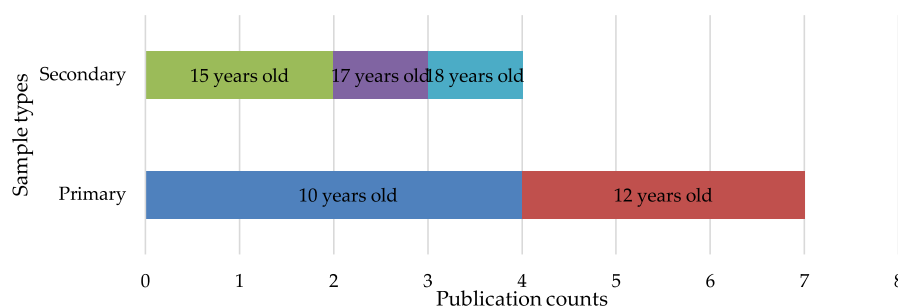


FIGURE 2
Publication counts based on sample types and ages.

employed a variety of reading tests and sample groups. Table 7 depicts the names of the reading tests, the types and the age of the samples. Findings reveal that only five studies developed new reading assessment items (Askari and Karami, 2024; Toprak and Cakir, 2021; Alavi and Ranjbaran, 2018; Liu and Read, 2023; Toprak-Yildiz, 2022). In contrast, the remaining 31 studies employed a retrofitting approach by adapting existing tests to analyze cognitive attributes of reading skills and to produce diagnostic profiling reports using different types of tests such as high-stakes language proficiency test (Boori et al., 2023), test for English Majors (Chen et al., 2023), International English Language Testing System (IELTS) (Jang et al., 2023) and high-stakes university entrance examination (Tonekaboni et al., 2021). Meanwhile, the study samples represent five distinct groups: undergraduate students, postgraduate students, undergraduate student candidates, postgraduate student candidates, and specific international examination test takers. While the majority of the studies did not report the ages of the samples, 12 studies did provide age ranges. None of these studies focused on a single age group; instead, the samples encompassed various age ranges. The youngest participants were 18 years old (Toprak-Yildiz, 2022) and the oldest was 59 years old (Jang et al., 2023).

Figure 3 displays the publication counts categorized by test types and sample types. Among the five studies that developed new assessment items, all utilized undergraduate students (Toprak and Cakir, 2021; Alavi and Ranjbaran, 2018), with one study also combining undergraduate and postgraduate students (Askari and Karami, 2024). This kind of study, also known as true CDA studies, undergoes a rigorous process of designing cognitive models, constructing Qmatrix, developing and validating new items and finally profiling test-takers mastery classes. For retrofitting studies that adapted existing tests, 15 studies involved undergraduate students (Chen et al., 2023), seven involved postgraduate students (Boori et al., 2024), two focused on undergraduate student candidates (Hemati and Baghaei, 2020) and, respectively, five studies targeted postgraduate student candidates (Javidanmehr et al., 2019) and specific examination test takers (Jang et al., 2023). Retrofitting is ubiquitous in CDA studies due to the need to adapt existing assessments to align with cognitive diagnostic models, allowing researchers to extract detailed information about the test-takers strengths and weaknesses without

developing entirely new tests (Mohd Noh and Mohd Matore, 2024).

4.2 RQ2: How do cognitive diagnostic assessment studies conceptualize and operationalize key reading attributes for young and adult readers?

The analysis of CDA studies reveals distinct approaches to defining and measuring reading attributes for both young and adult readers. Across the reviewed studies, reading attributes are conceptualized based on cognitive processing theories and language comprehension models. Based on the coding process, 13 codes of cognitive reading attributes emerged with at least one occurrence. Table 8 highlights the essential reading attributes employed in CDA studies to evaluate the reading skills of young readers. Topping the list are finding explicit information (EXP) and making inferences (INF). These are followed by generalizing or synthesizing main ideas (GEN), making interpretation of texts (INT), lexical or vocabulary knowledge (LEX), evaluating or analyzing the text (EVA), grammatical knowledge (GRAM), syntactic knowledge (SYN) and summarizing (SUM). Finally, others represent four other attributes with <3 occurrences, including sequencing, understanding charts, expressing in written forms and recognizing authors' styles.

Table 9 presents key reading attributes of adult readers assessed in CDA studies. Regarding the number of attributes used to assess reading skills among adult readers, they ranged between 3 (Jang et al., 2023) and 10 (Kim, 2015). The coding process has produced 21 codes with at least one occurrence of attributes used in the reviewed studies to measure the reading skills of adult readers. The key attributes are led by lexical knowledge (LEX), making inferences (INF) and finding explicit information (EXP). These are followed by syntactic knowledge (SYN), summarizing ideas (SUM), grammar knowledge (GRAM), generalizing or synthesizing (GEN), identifying main ideas (MI), making interpretation (INT), connecting ideas (CON), recognizing authors' styles (AUT), scanning or skimming (SCAN), understanding paragraphs (PARA), evaluating (EVA), building a mental model (BUI), identifying supporting details (DET). Others

TABLE 7 Reading tests and sample profiles in CDA studies involving adult readers.

No.	Study	Reading test	Sample types	Sample ages
1.	Askari and Karami (2024)	Newly developed	Bachelor's and master's degrees students	19–47 years old
2.	Boori et al. (2024)	Iranian high-stakes language proficiency test	PhD students	Not mentioned
3.	Boori et al. (2023)	Iranian high-stakes language proficiency test	PhD students	Not mentioned
4.	Chen et al. (2023)	Test for English Majors (TEM)	Final year bachelor's degree students	Not mentioned
5.	Jang et al. (2023)	International English Language Testing System (IELTS)	IELTS test-takers	14–59 years old
6.	Mohammed et al. (2023)	B1 Preliminary English Test (PET)	Tertiary education students	19 and 39 years old
7.	Shahmirzadi and Marashi (2023)	General English test	PhD students	Not mentioned
8.	Sun and Hwang (2023)	College English Test Band Four (CET-4)	Tertiary education students	Not mentioned
9.	Wang (2023)	Large-scale Spanish proficiency test	Bachelor's degree students	21–22 years old
10.	Cai and Chen (2022)	Test for English Majors Band 4 (TEM-4)	Bachelor's degree students	Not mentioned
11.	He et al. (2022)	Large scale in-house EFL exit test	Bachelor's degree students	19–21 years old
12.	Toprak-Yildiz (2022)	Newly developed	Bachelor's degree students	18–20 years old
13.	Du and Ma (2021)	College English Test Band 4 (CET-4)	First year bachelor's degree students	Not mentioned
14.	Tonekaboni et al. (2021)	High-stakes university entrance examination (UEE) Master of Arts (M.A.) exam	Master degree's students	Not mentioned
15.	Toprak and Cakir (2021)	A new diagnostic L2 reading comprehension test	Bachelor's degree students	18–23 years old
16.	Liu and Read (2021)	DELNA reading assessment	Bachelor's degree students	Not mentioned
17.	Hemati and Baghaei (2020)	Iranian University Entrance Examination (IUEE)	Bachelor's degree candidates	Not mentioned
18.	Mirzaei et al. (2020)	International English Language Testing System (IELTS)	IELTS test-takers	Not mentioned
19.	Tabatabaee-Yazdi (2020)	Iranian University Entrance Examination (IUEE)	Bachelor's degree candidates	20–44 years old
20.	Zhao et al. (2020)	University final English examination	First year bachelor's degree students	Not mentioned
21.	Javidanmehr et al. (2019)	University entrance exam for PhD programs	PhD candidates	25–50 years old
22.	Ravand (2019)	Iranian University Entrance Examination (IUEE)	Master's degree candidates	Not mentioned
23.	Alavi and Ranjbaran (2018)	Newly developed Iranian National University	Bachelor's degree students	Not mentioned
24.	Ravand and Robitzsch (2018)	Entrance Examination (INUEE) - master program candidate admission	Master's degree candidates	22–35 years old
25.	Li and Wang (2017)	Test of Practical Chinese	Tertiary education students	Not mentioned
26.	Ranjbaran and Alavi (2017)	Reading comprehension test for Bachelor's Degree General English course	Bachelor's degree students	Not mentioned
27.	Ravand (2016)	Iranian National University Entrance Examination (INUEE)—master program candidate admission	Master's degree candidates	22–25 years old
28.	Sook Yi (2016)	Examination for the Certificate of Proficiency in English (ECPE)	TOEFL test-takers	Not mentioned
29.	Baghaei and Ravand (2015)	Iranian National University Entrance Examination (INUEE)	Master degree candidates	Not mentioned
30.	Kim (2015)	Reading section of a placement test used in the adult ESL program	Tertiary education students	18 years old and above
31.	Li et al. (2015)	Michigan English Language Assessment Battery (MELAB)	Tertiary education students	Not mentioned
32.	Li and Suen (2013)	Michigan English Language Assessment Battery (MELAB)	Tertiary education students	Not mentioned
33.	Jang (2009a)	LanguEdge Courseware reading test	Bachelor's degree students	Not mentioned
34.	Jang (2009b)	LanguEdge Courseware reading test	Bachelor's degree students	Not mentioned
35.	Lee and Sawaki (2009)	Test of English as a Foreign Language (TOEFL iBT)	TOEFL test-takers	Not mentioned
36.	Sawaki et al. (2009)	Test of English as a Foreign Language (TOEFL iBT)	TOEFL test-takers	Not mentioned

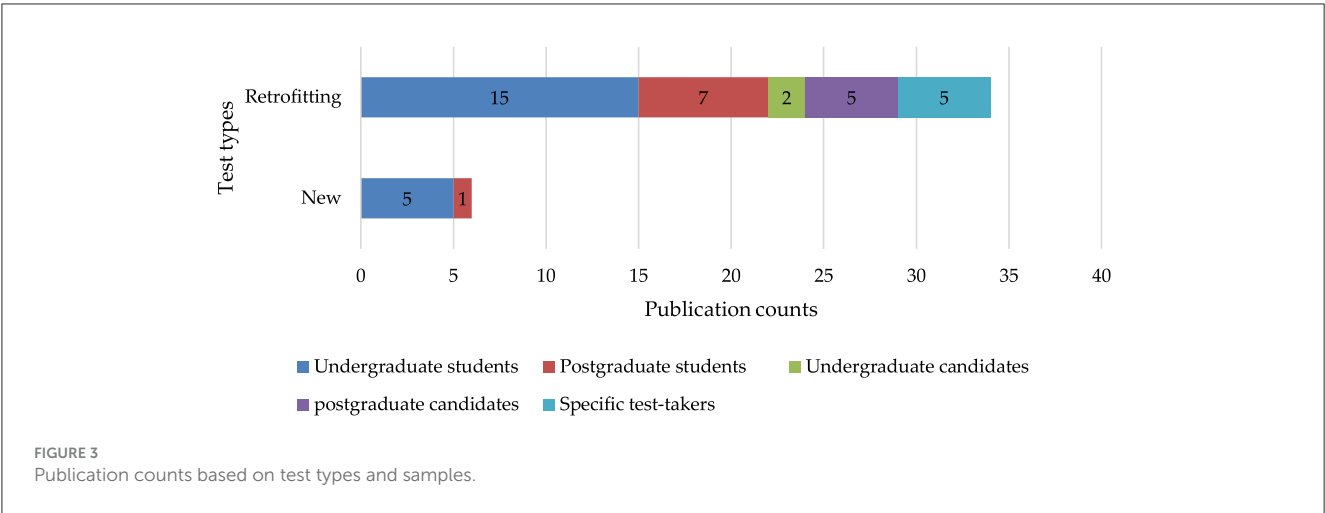


TABLE 8 Key reading attributes of young readers in CDA studies.

No.	Study	No. of Attributes	EXP	INF	GEN	INT	LEX	EVA	GRA	SYN	SUM	Others
1.	Yumsek (2023)	6	/	/			/		/		/	/
2.	Nallasamy and Khairani (2022)	5	/	/	/		/			/		
3.	Toprak-yildiz (2021)	4	/	/		/		/				
4.	George and Robitzsch (2021)	9	/	/	/			/				/
5.	Liu and Bian (2021)	4	/	/	/							
6.	Fan and Yan (2020)	7	/	/	/		/		/	/		/
7.	Thi and Loye (2019)	5	/	/		/		/				/
8.	Chen and Chen (2016)	6	/	/	/	/		/				
9.	Chen and Chen (2015)	6	/	/	/	/		/				
10.	Jang et al. (2015)	4	/	/		/	/		/		/	
11.	Jang et al. (2013)	3	/	/		/	/		/		/	
Total			11	11	6	6	5	5	4	2	3	4

EXP, Finding explicit information; INF, Making inferences; GEN, Generalizing or synthesizing ideas; INT, Making interpretation; LEX, Lexical knowledge; EVA, Evaluating text; GRAM, Grammar knowledge; SYN, Syntactic knowledge; SUM, Summarizing ideas.

refer to other attributes with <3 occurrences including negation, applying background knowledge, predicting, pragmatic knowledge and holding memory.

4.3 RQ3: How do the similarities and differences in cognitive reading attributes assessed in young and adult readers reflect their cognitive and linguistic development?

A comparison of the cognitive reading attributes assessed in young and adult readers highlights key similarities and differences, which reflect developmental changes in cognitive and linguistic processing. Table 10 presents the distribution of attributes across the two groups, showing that while both young and adult readers rely on fundamental comprehension skills such as finding explicit

information and making inferences, the complexity and scope of assessed attributes increase with age.

For young readers, the findings indicate that “finding explicit information” and “making inferences” are the most emphasized skills, with both being assessed 100% of the time. These skills are fundamental in evaluating reading comprehension, as they help determine a reader’s ability to understand and extract meaning from texts. Moderately assessed skills include “synthesizing” and “making interpretations” (55%), which suggest an effort to measure higher-order thinking, along with “lexical knowledge” and “evaluating content” (45%), highlighting the role of vocabulary and critical assessment in reading. Meanwhile, “grammar knowledge” (36%) and “summarizing ideas” (27%) are given less emphasis, suggesting that structural language knowledge and text summarization are secondary to comprehension skills. Less frequently assessed attributes include “understanding graphs”, “expressing in written forms”, and “syntactic knowledge” (18%),

TABLE 9 Key reading attributes of adult readers assessed in CDA studies.

Study	No of attributes	LEX	INF	EXP	SYN	SUM	GRAM	GEN	MI	INT	CON	AUTH	SCAN/SKM	PARA	EVA	BUILD	DET	Others
Askari and Karami (2024)	7	/	/		/	/	/							/				
Boori et al. (2023)	5	/	/	/			/		/									
Boori et al. (2024)	5	/	/	/			/		/									
Chen et al. (2023)	8	/	/	/					/	/		/			/			
Jang et al. (2023)	3		/	/		/												
Mohammed et al. (2023)	4		/				/		/								/	/
Shahmirzadi and Marashi (2023)	5	/	/	/	/						/							
Sun and Hwang (2023)	6	/		/	/	/		/										
Wang (2023)	5	/		/	/	/												
Cai and Chen (2022)	5		/	/		/				/					/			
He et al. (2022)	4	/	/	/	/			/										
Toprak-Yildiz (2022)	5	/	/	/			/				/							
Du and Ma (2021)	8	/	/	/	/	/								/				
Tonekaboni et al. (2021)	6	/	/		/			/	/							/		
Toprak and Cakir (2021)	5		/	/	/							/		/				
Liu and Read (2021)	6	/	/	/		/		/		/								
Hemati and Baghaei (2020)	5	/	/	/			/								/			
Mirzaei et al. (2020)	6	/	/				/	/					/					
Tabatabaee-Yazdi (2020)	5	/	/	/			/								/			
Zhao et al. (2020)	5		/	/				/		/	/							/
Javidanmehr et al. (2019)	5	/	/	/	/						/							
Ravand (2019)	5	/	/			/	/							/				

(Continued)

TABLE 9 (Continued)

Study	No of attributes	LEX	INF	EXP	SYN	SUM	GRAM	GEN	MI	INT	CON	AUTH	SCAN/SKM	PARA	EVA	BUILD	DET	Others
Alavi and Ranjbaran (2018)	9	/	/	/		/				/		/	/					/
Ravand and Robitzsch (2018)	5	/	/		/				/								/	/
Li and Wang (2017)	8	/	/	/					/				/					/
Ranjbaran and Alavi (2017)	9	/	/	/		/				/		/	/					/
Ravand (2016)	5	/	/		/				/								/	/
Sook Yi (2016)	4	/		/				/			/							
Baghaei and Ravand (2015)	5	/	/		/				/								/	
Kim (2015)	10	/	/	/	/	/	/						/	/				/
Li et al. (2015)	4	/	/	/	/													
Li and Suen (2013)	4	/		/	/													
Jang (2009a)	9	/	/	/	/	/				/						/		/
Jang (2009b)	9	/	/	/	/	/				/						/		/
Lee and Sawaki (2009)	4	/		/				/										
Sawaki et al. (2009)	4	/		/				/				/						
Total		32	30	28	17	13	10	9	9	8	5	5	5	5	4	3	4	10

LEX, Lexical knowledge; INF, Making inferences; EXPL, Finding explicit information; SYNT, Syntactic knowledge; SUM, Summarizing ideas; GRAM, Grammar knowledge; GEN, Generalizing or synthesizing ideas; MI, Identifying main idea; INT, Making interpretation; CON, Connecting ideas; AUTH, Recognizing authors' styles; SCAN/SKIM, Scanning or skimming; PARA, Understanding paragraphs; EVA, Evaluating text; BUILD, Building a mental model; DET, Identifying supporting details.

TABLE 10 Frequency and percentage of attributes used in assessing reading skills of adult and young readers.

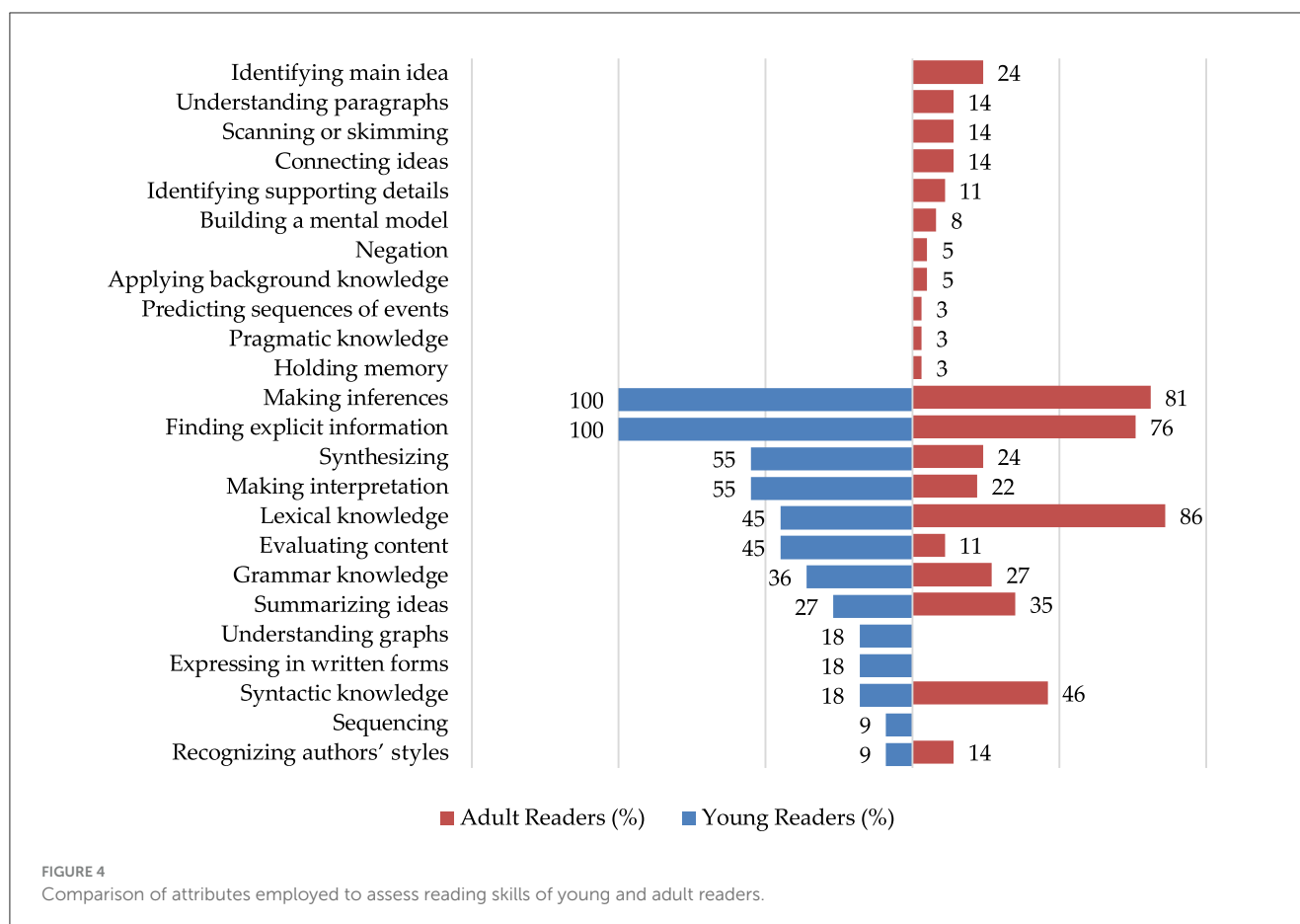
No.	Attributes	Adult readers		Young readers	
		Frequency	Percentage	Frequency	Percentage
1.	Applying background knowledge	2	5%	n/a	n/a
2.	Building a mental model	3	8%	n/a	n/a
3.	Connecting ideas	5	14%	n/a	n/a
4.	Evaluating text	4	11%	5	45%
5.	Expressing in written forms	n/a	n/a	2	18%
6.	Finding explicit information	29	78%	11	100%
7.	Grammar knowledge	10	27%	4	36%
8.	Holding memory	1	3%	n/a	n/a
9.	Identifying main idea	9	24%	n/a	n/a
10.	Identifying supporting details	4	11%	n/a	n/a
11.	Lexical knowledge	32	86%	5	45%
12.	Making inferences	31	84%	11	100%
13.	Making interpretation	8	22%	6	55%
14.	Negation	2	5%	n/a	n/a
15.	Pragmatic knowledge	1	3%	n/a	n/a
16.	Predicting	1	3%	n/a	n/a
17.	Recognizing authors' styles	5	14%	1	9%
18.	Scanning or skimming	5	14%	n/a	n/a
19.	Sequencing	n/a	n/a	1	9%
20.	Syntactic knowledge	17	46%	2	18%
21.	Summarizing ideas	14	38%	3	27%
22.	Synthesizing	9	24%	6	55%
23.	Understanding graphs	n/a	n/a	2	18%
24.	Understanding paragraphs	5	14	n/a	n/a

which may reflect their lesser role in early reading development. Finally, “sequencing” and “recognizing authors’ styles” are the least prioritized skills, assessed only 9% of the time. The findings suggest a strong focus on comprehension and meaning-making, with less emphasis on linguistic structure and stylistic analysis in assessing young readers’ abilities.

As for adult readers, the findings reveal that “lexical knowledge” (86%), “making inferences” (81%), and “finding explicit information” (76%) are the most frequently assessed attributes in evaluating adult readers’ skills. This suggests a strong emphasis on vocabulary knowledge, the ability to draw conclusions, and identifying direct information from texts. “Syntactic knowledge” (46%) and “summarizing ideas” (35%) are also significant, reflecting the importance of understanding sentence structures and condensing key points. “Grammar knowledge” (27%), along with “identifying the main idea” and “synthesizing” (24% each), highlights a moderate focus on language rules and higherorder thinking. “Making interpretation” (22%) indicates a role in deeper text analysis. Several other attributes, including “understanding paragraphs”, “scanning or

skimming”, “connecting ideas”, and “recognizing authors’ styles” (14% each), receive less emphasis, suggesting they are assessed selectively. “Identifying supporting details” and “evaluating content” (11% each) are considered minor components, while “building a mental model” (8%) is assessed infrequently. The least prioritized attributes include “negation” and “applying background knowledge” (5% each), as well as “predicting sequences of events”, “pragmatic knowledge”, and “holding memory” (3% each) receiving minimal attention. The findings highlight a strong focus on comprehension, vocabulary, and sentence structure in adult reading assessments, while aspects related to memory, background knowledge, and prediction are less emphasized.

Moreover, studies involving adult readers have assessed a greater number of attributes, totaling 21, compared to 13 attributes identified in studies on young readers, as illustrated in Figure 4. Notably, certain attributes were exclusively examined in CDA studies on adult readers, including understanding paragraphs, connecting ideas, applying background knowledge, and pragmatic knowledge. Conversely, some attributes were



uniquely assessed among young readers, such as understanding graphs and sequencing.

5 Discussion

5.1 Developmental progression of cognitive reading skills

The findings of this review indicate that reading frameworks applied across age groups are shaped by varying conceptualizations of cognitive reading processes. In young readers, the dominant attributes, finding explicit information and making inferences, suggest a framework that primarily focuses on basic comprehension and text-based understanding. These attributes align with early reading development models, which emphasize the acquisition of literal comprehension skills as the foundation for reading proficiency (Kim and Suk, 2020). The analysis reveals that young readers are predominantly assessed on literal and interpretive levels of comprehension (Nurjanah and Putri, 2020), reflecting a framework that prioritizes understanding surface-level information and basic meaning derivation. This observation also aligns with findings from neurocognitive studies, where children rely more on literal processing, supported by stronger inter-regional connectivity (Liu et al., 2018). However, the limited

inclusion of syntactic knowledge, along with the absence of critical and creative comprehension skills such as evaluating content or connecting ideas, suggests that early reading frameworks place less emphasis on complex higher-order cognitive processes, which are essential for developing more sophisticated reading abilities.

Conversely, the more diverse set of attributes in adult readers' assessments highlights a broader cognitive spectrum, where vocabulary mastery, sentence structure, and higher-order skills such as synthesizing ideas and evaluating content are crucial components of reading competence. This shift signals a more holistic reading framework in adult readers, consistent with interactive and strategic reading models that require integrating both text-based and knowledge-based processes (Khalifa and Weir, 2009). The broader cognitive spectrum observed in adult readers is further supported by neurocognitive evidence, which indicates that adults not only rely on semantic processing networks (Liu et al., 2024). As reading development progresses, semantic, pragmatic, and syntactic networks play a crucial role in predicting reading ability at earlier stages of life, while more complex integrative networks emerge as stronger predictors of reading proficiency later in life (Horowitz-Kraus et al., 2024). This developmental trajectory suggests that the future-reading network, which involves the integration of multiple cognitive processes, becomes more prominent as readers mature, allowing adult readers to engage with narratives in a more sophisticated manner.

5.2 Diagnostic precision in measuring higher-order reading skills

The diagnostic precision of cognitive diagnostic assessment models in capturing higher-order reading skills such as synthesizing, evaluating, and interpreting texts remains a contentious issue across different age groups. While these skills are more frequently assessed in adult readers, their limited use in young readers' assessments could stem from both genuine developmental constraints and methodological limitations. Higher-order cognitive processes often require abstract reasoning, metacognitive awareness, and the integration of prior knowledge, abilities that are typically more advanced in older readers. This aligns with the hierarchical structure of reading development, where early stages focus on literal comprehension and basic meaning retrieval, while more advanced stages involve critical appraisal and inferential thinking (Chen et al., 2023). Nevertheless, the limited inclusion of higher-order skills in young readers' assessments may also be attributed to the inherent complexity of these skills and the difficulty of designing reliable, scalable tasks that isolate and measure them independently (Molokopoeva and Simard, 2024). Unlike lower-level processes, which can be assessed using controlled measures with clear indicators of specific deficits, higher-order reading involves interconnected cognitive processes that operate simultaneously (Alderson et al., 2015).

Moreover, automaticity in lower-level skills such as retrieving explicit information and lexical knowledge is essential for freeing up cognitive resources needed for more complex comprehension tasks (Suzuki and Elgort, 2023). Therefore, CDA models may prioritize lower-level attributes to establish a solid foundation before progressing to higher-order skills. However, this approach risks oversimplifying the reading process and underestimating young readers' potential to engage in advanced cognitive tasks when provided with appropriate scaffolding and instructional support. Developing age-appropriate assessments that integrate both fundamental and higher-order skills is essential to providing a more detailed diagnostic profile of young readers' cognitive abilities. Additionally, CDA models often rely on retrofitting existing assessments, which are not originally designed to probe such complex cognitive processes (Ketabi et al., 2021). Consequently, the underrepresentation of higher-order skills might not solely reflect developmental readiness but rather the inadequacy of assessment tools to elicit and diagnose these abilities. Addressing this limitation requires the development of customized assessment instruments that integrate distinct item formats, interactive tasks, and multimodal stimuli to better capture the gradual emergence of higher-order thinking in young readers (Hemati and Baghaei, 2020).

6 Conclusion

This study has addressed all the three research questions by critically analyzing the cognitive attributes used in reading assessments across different age groups and providing valuable insights into how reading skills are conceptualized and measured from a cognitive perspective. By comparing the attributes used in assessments of young and adult readers, the study highlights

the distinct cognitive demands that readers encounter at different stages of development. The findings underscore the importance of age-specific approaches to reading instruction and assessment, which are essential for fostering reading proficiency across the lifespan. Ultimately, this research contributes to the broader discourse on educational assessment by advocating for more differentiated and contextually relevant assessment practices. By aligning reading assessments with the cognitive capabilities of learners, educators, policymakers, and assessment developers can better support the development of literacy skills to ensure that all individuals are equipped with the reading abilities necessary for their academic and professional success.

6.1 Theoretical and practical implications

This study contributes to the theoretical understanding of reading skill development across different age groups by highlighting the distinct cognitive attributes prioritized in assessing young and adult readers. The findings support and extend existing CDA frameworks by offering a comparative analysis of how reading skills are conceptualized and measured at various stages of life. By distinguishing between foundational skills emphasized in young learners and more complex, abstract skills in adults, the study provides a comprehensive perspective on the cognitive processes underlying reading proficiency. This reinforces the need for age-specific theoretical models that account for the evolving nature of reading comprehension and the adaptive strategies employed by readers as they mature.

The review also suggested that internal and external stakeholders should adjust their approaches based on the findings. Concerning internal stakeholders, especially educators, who play a key role in decision-making that directly impacts the institution's operations and pedagogical direction (Chan and Chou, 2020), the findings have direct implications for the design and implementation of reading instructional strategies and assessments. The identification of key reading attributes for different age groups enables educators to tailor their teaching methods and assessment tools to better align with the cognitive abilities of their students. For young learners, the emphasis on attributes like making inferences and finding explicit information suggests that teaching strategies should focus on enhancing these critical interpretive skills. In contrast, for adult learners, greater attention should be paid to vocabulary development, syntactic knowledge, and summarization techniques. Although readers of all ages must develop a range of reading skills, young learners acquire language implicitly through exposure, building interpretive abilities like analysis and inference that are essential for academic success. In contrast, adult learners benefit from explicit instruction in vocabulary, syntax, and summarization, which are vital for comprehending more complex texts and engaging in professional or academic reading. Thus, these distinctions highlight the need for a differentiated instructional approach that accommodates the developmental and cognitive differences between young and adult learners.

Regarding external stakeholders, such as policymakers, curriculum developers, educational authorities, and assessment

developers, who influence decisions, shape policies, and establish frameworks impacting the institution's operations within the broader educational ecosystem (Agnew, 2021), the study highlights the need for reading assessments that accurately reflect the cognitive demands placed on readers of different ages. The findings advocate for a more differentiated approach to assessment design, one that recognizes the varying complexities of reading tasks faced by young and adult readers. This has implications for standardized testing and literacy programs, suggesting that a one-size-fits-all model is inadequate. External stakeholders are encouraged to consider these insights when formulating educational policies, designing literacy interventions, and developing new assessment tools that are both equitable and effective across diverse populations.

6.2 Limitations and recommendations

Despite offering valuable insights, this study acknowledges several limitations that highlight opportunities for further research. The study focuses on reading assessment research through the cognitive diagnostic approach, which provides a fine-grained understanding of reading attributes but may not fully capture perspectives from other emerging assessment frameworks. The use of the scoping review methodology in identifying, screening, and analyzing publications facilitates a comprehensive mapping of the literature, yet it does not offer the quantitative synthesis that could be achieved through meta-analyses. Additionally, the study is confined to English language resources and reading assessments in the context of English as a first, second, or foreign language, which may restrict the generalizability of the findings across diverse linguistic and cultural settings. The selection of studies from three reputable academic databases with a focus on empirical studies ensures the inclusion of high-quality publications but may exclude valuable insights from other sources such as conference proceedings, book chapters, and gray literature.

To address these limitations, future research could explore several key areas. Future reviews on advanced technologies in reading assessment, such as eye tracking and electroencephalography (EEG), are needed to provide objective, real-time insights into readers' cognitive processes and offer a more holistic understanding of reading behaviors. Additionally, future studies could undertake more rigorous review methodologies, such as meta-analyses and bibliometric studies, to provide a more extensive understanding of prevailing trends and gaps in the literature on cognitive reading attributes. Cross-linguistic comparative research would also be valuable in examining the consistency of cognitive reading attributes across different languages and cultural contexts, thereby enhancing the generalizability of the findings. Furthermore, comparative investigations between conventional and technology-assisted assessments could yield critical insights into the effectiveness, adaptability, and efficiency of different assessment approaches. Finally, evaluating the strengths and limitations of various reading assessment methods would help identify the most reliable and valid techniques for measuring reading proficiency across diverse age groups and educational settings. Addressing these avenues

would significantly enrich the understanding of cognitive reading attributes and contribute to the advancement of reading assessment practices in various linguistic and educational contexts.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

MMN: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Visualization, Writing – original draft, Writing – review & editing. MMM: Funding acquisition, Resources, Supervision, Validation, Visualization, Writing – review & editing, Conceptualization. NS: Supervision, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. The study was funded by Ganjaran Penerbitan GP-KO21854 and TAP-KO21854 Universiti Kebangsaan Malaysia.

Acknowledgments

The authors would like to extend gratitude to Kumpulan Penyelidikan Universiti (KPU) Penilaian Pendidikan, Faculty of Education, UKM Bangi for continuous support.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Agnew, J. (2021). Internal and external stakeholder impact on curriculum design. *Acad. Lett.* 1–4. doi: 10.20935/AL2028
- Alavi, M., and Ranjbaran, F. (2018). Constructing and validating a Q-Matrix for cognitive diagnostic analysis of a reading comprehension test battery. *J. Engl. Lang. Teach. Learn.* 21, 1–15. Available online at: https://elt.tabrizu.ac.ir/article_17207_d9a9150ba8996ecfe90af3e07b905503.pdf
- Alderson, J. C., Brunfaut, T., and Harding, L. (2015). Towards a theory of diagnosis in second and foreign language assessment: insights from professional practice across diverse fields. *Appl. Linguist.* 36, 236–260. doi: 10.1093/applin/amt046
- Anderson, L. W., and Krathwohl, D. (2001). *A Taxonomy for Learning, Teaching and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. New York, NY: Longman.
- Arksey, H., and O'Malley, L. (2005). Scoping studies: towards a methodological framework. *Int. J. Soc. Res. Methodol. Theory Pract.* 8, 19–32. doi: 10.1080/1364557032000119616
- Askari, M., and Karami, H. (2024). On the relationship between sensory learning styles and reading subskill profiles: an application of Fusion Model. *Lang. Relat. Res.* 15, 245–274. doi: 10.2139/ssrn.4873700
- Baghaei, P., and Ravand, H. (2015). A cognitive processing model of reading comprehension in English as a foreign language using the linear logistic test model. *Learn. Individ. Differ.* 43, 100–105. doi: 10.1016/j.lindif.2015.09.001
- Birkle, C., Pendlebury, D. A., Schnell, J., and Adams, J. (2020). Web of science as a data source for research on scientific and scholarly activity. *Quant. Sci. Stud.* 1, 363–76. doi: 10.1162/qss_a_00018
- Boori, A. A., Ghazanfari, M., Ghonsooly, B., and Baghaei, P. (2023). The construction and validation of a Qmatrix for cognitive diagnostic analysis: the case of the reading comprehension section of the IAUPT. *Int. J. Lang. Test.* 13, 31–53. doi: 10.22034/IJLT.2023.383112.1227
- Boori, A. A., Ghazanfari, M., Ghonsooly, B., and Baghaei, P. (2024). A cognitive diagnostic modeling analysis of the reading comprehension section of an Iranian high-stakes language proficiency test. *Int. J. Lang. Test.* 14, 17–33. doi: 10.22034/ijlt.2023.399561.1256
- Cai, Y., and Chen, H. (2022). The fluctuating effect of thinking on language performance: new evidence for the Island Ridge Curve. *Lang. Assess. Q.* 19, 465–479. doi: 10.1080/15434303.2022.2080553
- Chan, S. J., and Chou, C. (2020). Who influences higher education decision-making in Taiwan? An analysis of internal stakeholders. *Stud. High Educ.* 45, 2101–2109. doi: 10.1080/03075079.2020.1823646
- Chen, H., Cai, Y., and de la Torre, J. (2023). Investigating second language (L2) reading subskill associations: a cognitive diagnosis approach. *Lang. Assess. Q.* 20, 166–89. doi: 10.1080/15434303.2022.2140050
- Chen, H., and Chen, J. (2015). Exploring reading comprehension skill relationships through the G-DINA model. *Educ. Psychol.* 36, 1049–1064. doi: 10.1080/01443410.2015.1076764
- Chen, H., and Chen, J. (2016). Retrofitting non-cognitive-diagnostic reading assessment under the Generalized DINA Model framework. *Lang. Assess. Q.* 13, 218–230. doi: 10.1080/15434303.2016.1210610
- Coady, J. (1979). "A psycholinguistic model of the ESL reader," in *Reading in A Second Language*, eds R. Mackay, B. Barkman, and R. R. Jordan (Rowley, MA: Newbury House).
- Davis, F. B. (1968). Research in comprehension in reading. *Read. Res. Q.* 3:499. doi: 10.2307/747153
- Du, W., and Ma, X. (2021). Probing what's behind the test score: application of multi-CDM to diagnose EFL learners' reading performance. *Read Writ.* 34, 1441–1466. doi: 10.1007/s11145-021-10124-x
- Fan, T., and Yan, X. (2020). Diagnosing English reading ability in Chinese senior high schools. *Stud. Educ. Eval.* 67, 1–12. doi: 10.1016/j.stueduc.2020.100931
- George, A. C., and Robertzsch, A. (2021). Validating theoretical assumptions about reading with cognitive diagnosis models. *Int. J. Test.* 21, 105–29. doi: 10.1080/15305058.2021.1931238
- Geramipour, M., Talebzadeh, H., and Mahdi, S. (2021). The optimal cognitive diagnostic model (CDM) for the grammar section of MA entrance examination of state universities for EFL candidates. *Lang. Relat. Res.* 12, 187–218. doi: 10.29252/LRR.12.1.6
- Gierl, M. J., and Cui, Y. (2008). Defining characteristics of diagnostic classification models and the problem of retrofitting in cognitive diagnostic assessment. *Meas. Interdiscip. Res. Perspect.* 6, 263–268. doi: 10.1080/15366360802497762
- Goodman, K. S. (1967). Reading: a psycholinguistic guessing game. *J. Read Spec.* 6, 126–135. doi: 10.1080/19388076709556976
- Gough, P. B. (1972). "One second of reading," in *Language by Ear and by Eye*, eds J. F. Kavanagh, and I. G. Mattingly (Cambridge: The MIT Press), 331–358.
- He, L., Xiong, L., and Min, S. (2022). Diagnosing listening and reading skills in the Chinese EFL context: performance stability and variability across modalities and performance levels. *System.* 106, 1–19. doi: 10.1016/j.system.2022.102787
- Heaton, J. (1991). *Writing English Language Test*. New York, NY: Longman.
- Hemati, S. J., and Baghaei, P. (2020). A cognitive diagnostic modeling analysis of the English reading comprehension section of the Iranian National University Entrance Examination. *Int. J. Lang. Test.* 10, 11–32. Available online at: https://www.ijlt.ir/article_114278_022ab6b64d4bc8acba9c561846828270.pdf
- Hoover, W. A., and Gough, P. B. (1990). The simple view of reading. *Read Writ. Interdiscip. J.* 2, 127–160. doi: 10.1007/BF00401799
- Horowitz-Kraus, T., Meri, R., Holland, S. K., Farah, R., Rohana, T., Haj, N., et al. (2024). Language first, cognition later: different trajectories of subcomponents of the future-reading network in processing narratives from kindergarten to adolescence. *Brain Connect.* 14, 60–69. doi: 10.1089/brain.2023.0012
- Hosseini Alast, S. F., and Baleghizadeh, S. (2024). The interplay of glossing with text difficulty and comprehension levels. *Lang. Teach. Res.* 28, 1201–1230. doi: 10.1177/13621688211020423
- Hughes, A., and Hughes, J. (2020). *Testing for Language Teachers. 3rd Edn.* Cambridge: Cambridge University Press.
- Jang, E. E. (2009a). Cognitive diagnostic assessment of L2 reading comprehension ability: Validity arguments for Fusion Model application to LanguEdge assessment. *Lang. Test.* 26, 31–73. doi: 10.1177/0265532208097336
- Jang, E. E. (2009b). Demystifying a Q-Matrix for making diagnostic inferences about L2 reading skills. *Lang. Assess. Q.* 6, 210–238. doi: 10.1080/15434300903071817
- Jang, E. E., Barron, C., Kim, H., and Russell, B. (2023). Mixed methods investigation into test score users' perspectives about IELTS reading skill profiles. *Lang. Teach. Res.* 37, 231–247. doi: 10.32038/ltrq.2023.37.13
- Jang, E. E., Dunlop, M., Park, G., and Van Der Boom, E. H. (2015). How do young students with different profiles of reading skill mastery, perceived ability, and goal orientation respond to holistic diagnostic feedback? *Lang. Test.* 32, 1–25. doi: 10.1177/0265532215570924
- Jang, E. E., Dunlop, M., Wagner, M., and Kim, Y., Gu, Z. (2013). Elementary school ELLs' reading skill profiles using Cognitive Diagnosis Modeling: roles of length of residence and home language environment. *Lang. Learn.* 63, 400–436. doi: 10.1111/lang.12016
- Javidanmehr, Z., Reza, M., and Sarab, A. (2019). Retrofitting non-diagnostic reading comprehension assessment: application of the G-DINA Model to a high-stakes reading comprehension test. *Lang. Assess. Q.* 16, 1–18. doi: 10.1080/15434303.2019.1654479
- Ketabi, S., Alavi, S. M., and Ravand, H. (2021). Diagnostic test construction: insights from cognitive diagnostic modeling. *Int. J. Lang. Test.* 11, 22–35. Available online at: https://www.ijlt.ir/article_128357_68e74c779145d7be54c24f610a549879.pdf
- Khalifa, H., and Weir, C. J. (2009). *Examining Reading: Research and Practice in Assessing Second Language Reading*. Cambridge: Cambridge University Press. Available online at: <http://hdl.handle.net/10547/224632> (accessed February 10, 2025).
- Kim, A. Y. (2015). Exploring ways to provide diagnostic feedback with an ESL placement test: cognitive diagnostic assessment of L2 reading ability. *Lang. Test.* 32, 227–258. doi: 10.1177/0265532214558457
- Kim, Y., and Suk, G. (2020). Hierarchical and dynamic relations of language and cognitive skills to reading comprehension: testing the Direct and Indirect Effects Model of Reading (DIER). *J. Educ. Psychol.* 94:659. doi: 10.1037/edu0000407
- Kim, Y. S. G. (2020). Toward integrative reading science: the direct and indirect effects model of reading. *J. Learn. Disabil.* 53, 469–491. doi: 10.1177/0022219420908239
- LaBerge, D., and Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cogn. Psychol.* 6, 293–323. doi: 10.1016/0010-0285(74)90015-2
- Lee, H., and Lee, J. H. (2023). Extending the simple view of reading in second and foreign language learning: a meta-analytic structural equation modeling approach. *Rev. Educ. Res.* 94, 467–500. doi: 10.3102/00346543231186605
- Lee, Y. W., and Sawaki, Y. (2009). Application of three cognitive diagnosis models to ESL reading and listening assessments. *Lang. Assess. Q.* 6, 239–263. doi: 10.1080/15434300903079562
- Levac, D., Colquhoun, H., and O'Brien, K. K. (2010). Scoping studies: advancing the methodology. *Represent. Interv.* 5, 1–18. doi: 10.1186/1748-5908-5-69
- Li, H., Hunter, C. V., and Lei, P. W. (2015). The selection of cognitive diagnostic models for a reading comprehension test. *Lang. Test.* 33, 391–409. doi: 10.1177/0265532215590848
- Li, H., and Suen, H. K. (2013). Constructing and validating a Q-matrix for cognitive diagnostic analyses of a reading test. *Educ. Assess.* 18, 1–25. doi: 10.1080/10627197.2013.761522

- Li, S., and Wang, J. (2017). "Diagnostic assessment of L2 Chinese learners' reading comprehension ability," in *Chinese as A Second Language Assessment*, eds D. Zhang, and C. Lin (Singapore: Springer), 183–202.
- Li, Y., Zhen, M., and Liu, J. (2021). Validating a reading assessment within the cognitive diagnostic assessment framework: Q-matrix construction and model comparisons for different primary grades. *Front. Psychol.* 12:e786612. doi: 10.3389/fpsyg.2021.786612
- Lin, J. (2024). Developing a reading proficiency scale for Chinese as a second language: a confirmatory factor analysis approach. *Lang. Aware.* 33, 597–624. doi: 10.1080/09658416.2023.2291521
- Liu, H., and Bian, Y. (2021). Model selection for cognitive diagnostic analysis of the reading comprehension test. *Front. Psychol.* 12:e644764. doi: 10.3389/fpsyg.2021.644764
- Liu, X., Gao, Y., Di, Q., Hu, J., Lu, C., Nan, Y., et al. (2018). Differences between child and adult large-scale functional brain networks for reading tasks. *Hum. Brain Mapp.* 39, 662–679. doi: 10.1002/hbm.23871
- Liu, X., and Read, J. (2021). Investigating the skills involved in reading test tasks through expert judgement and verbal protocol analysis: convergence and divergence between the two methods. *Lang. Assess. Q.* 18, 357–381. doi: 10.1080/15434303.2021.1881964
- Liu, X., and Read, J. (2023). Designing a new diagnostic reading assessment for a local post-admission assessment program: a needs-driven approach. *Educ. Linguist.* 61, 183–102. doi: 10.1007/978-3-031-33541-9_5
- Liu, X., Zhang, L., Yu, S., Bai, Z., Qi, T., Mao, H., et al. (2024). The effects of age and reading experience on the lifespan neurodevelopment for reading comprehension. *J. Cogn. Neurosci.* 36, 239–260. doi: 10.1162/jocn_a_02086
- Lockwood, C., Munn, Z., and Porritt, K. (2015). Qualitative research synthesis: methodological guidance for systematic reviewers utilizing meta-aggregation. *Int. J. Evid. Based Healthc.* 13, 179–187. doi: 10.1097/XEB.0000000000000062
- Luecke, S., and Lorie, J. (2013). Use of Bloom's Taxonomy in developing reading comprehension specifications. *J. Appl. Test Technol.* 14, 1–27. Available online at: <https://jattjournal.net/index.php/atp/article/view/45250/36643>
- Maas, L., Madison, M. J., and Brinkhuis, M. J. S. (2024). Properties and performance of the one-parameter loglinear cognitive diagnosis model. *Front. Educ.* 9:e1287279. doi: 10.3389/educ.2024.1287279
- Martín-Martín, A., Thelwall, M., Orduna-Malea, E., and Delgado López-Cózar, E. (2021). Google Scholar, Microsoft Academic, Scopus, Dimensions, Web of Science, and OpenCitations' COCI: a multidisciplinary comparison of coverage via citations. *Scientometrics.* 126, 871–906. doi: 10.1007/s11192-020-03690-4
- Masdoki, M., Din, R., and Effendi Ewan Mohd Matore, M. (2021). Teaching 4.0 competency in higher learning institutions: a systematic mapping review. *Int. J. Learn. Teach. Educ. Res.* 20, 217–31. doi: 10.26803/ijlter.20.10.12
- Mei, H., and Chen, H. (2022). Cognitive diagnosis in language assessment: a thematic review. *RELC J.* 55, 1–9. doi: 10.1177/00336882211072898
- Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychol. Rev.* 63, 81–97. doi: 10.1037/h0043158
- Miller, L. A., and Lovler, R. L. (2020). *Foundations of Psychological Testing: A Practical Approach*. 6th Edn. Los Angeles, CA: SAGE.
- Mirzaei, A., Vincheh, M. H., and Hashemian, M. (2020). Retrofitting the IELTS reading section with a general cognitive diagnostic model in an Iranian EAP context. *Stud. Educ. Eval.* 64, 1–10. doi: 10.1016/j.stueduc.2019.100817
- Mohamed Shaffril, H. A., Samsuddin, S. F., and Abu Samah, A. (2021). The ABC of systematic literature review: the basic methodological guidance for beginners. *Qual. Quant.* 55, 1319–1346. doi: 10.1007/s11135-020-01059-6
- Mohammed, A., Kareem, A., Dawood, S., Alghazali, T., Khlaif, Q., Sabti, A. A., et al. (2023). A cognitive diagnostic assessment study of the Reading Comprehension Section of the Preliminary English Test (PET). *Int. J. Lang. Test.* 13, 1–20. doi: 10.22034/ijlt.2022.362849.1195
- Mohd Noh, M. F., and Mohd Matore, M. E. E. (2020). Rating performance among raters of different experience through multi-facet Rasch measurement (MFRM) model. *J. Meas. Eval. Educ. Psychol.* 11, 147–162. doi: 10.21031/epod.662964
- Mohd Noh, M. F., and Mohd Matore, M. E. E. (2022). Rater severity differences in English language speaking assessment based on rating experience, training experience and teaching experience through many-faceted Rasch measurement analysis. *Front. Psychol.* 13:e941084. doi: 10.3389/fpsyg.2022.941084
- Mohd Noh, M. F., and Mohd Matore, M. E. E. (2024). Conundrum and considerations in Cognitive Diagnostic Assessment for language proficiency evaluation. *Inf. Manag. Bus. Rev.* 16, 63–72. doi: 10.22610/imbr.v16i2(I).3690
- Molokopceva, T., and Simard, D. (2024). Interaction between levels of text representation and working memory during L2 reading comprehension: what about it? *Int. J. Appl. Linguist.* 34, 568–585. doi: 10.1111/ijal.12516
- Munby, J. (1978). *Communicative Syllabus Design*. Cambridge: Cambridge University Press.
- Munn, Z., Pollock, D., Khalil, H., Alexander, L., McLnerney, P., Godfrey, C. M., et al. (2022). What are scoping reviews? Providing a formal definition of scoping reviews as a type of evidence synthesis. *JBIM Evid. Synth.* 20, 950–952. doi: 10.111124/JBIES-21-00483
- Nadmilail, A. I., Mohd Matore, M. E. E., Maat, S. M., and Sheridan, L. (2023). Broad vs. narrow traits: A scoping review of measuring personality traits in teacher selection using the situational judgment test. *Front. Psychol.* 14:e1217321. doi: 10.3389/fpsyg.2023.1217321
- Nallasamy, R., and Khairani, A. Z. (2022). Validating reading comprehension assessment under the GDINA Model. *Malaysian J. Soc. Sci. Humanit.* 7, 1–12. doi: 10.47405/mjssh.v7i11.1877
- Nightingale, E., Branum-Martin, L., and Greenberg, D. (2023). Modeling the relations between reading skills and assessment methods in adults who struggle with foundational reading skills. *Read Writ.* 36, 2175–2198. doi: 10.1007/s11145-022-10384-1
- Nurjanah, R. L., and Putri, S. R. (2020). The effect of literal comprehension on the higher levels of comprehension in reading skill: a longitudinal case study. *Engl. Lang. Lit. Int. Conf.* 5, 471–476. Available online at: <https://jurnal.unimus.ac.id/index.php/ELLIC/article/view/9692/6267>
- Peng, P., Lee, K., Luo, J., Li, S., Joshi, R. M., Tao, S., et al. (2021). Simple view of reading in Chinese: a one-stage meta-analytic structural equation modeling. *Rev. Educ. Res.* 91, 3–33. doi: 10.3102/0034654320964198
- Peters, M. D. J., Godfrey, C., McInerney, P., Khalil, H., Larsen, P., Marnie, C., et al. (2022). Best practice guidance and reporting items for the development of scoping review protocols. *JBIM Evid. Synth.* 20, 953–968. doi: 10.111124/JBIES-21-00242
- Piaget, J. (1950). *The Psychology of Intelligence*. New York, NY: Harcourt Brace.
- Polat, M., Toraman, Ç., and Turhan, N. S. (2022). Reliability analysis of PISA 2018 reading literacy student questionnaire based on item response theory (IRT): Turkey sample. *Int. J. Curric. Instr.* 14, 1004–1028. Available online at: <https://ijci.net/index.php/IJCI/article/view/906/469>
- Pollock, D., Davies, E. L., Peters, M. D. J., Tricco, A. C., Alexander, L., McInerney, P., et al. (2021). Undertaking a scoping review: a practical guide for nursing and midwifery students, clinicians, researchers, and academics. *J. Adv. Nurs.* 77, 2102–2113. doi: 10.1111/jan.14743
- Ranjbaran, F., and Alavi, S. M. (2017). Developing a reading comprehension test for cognitive diagnostic assessment: a RUM analysis. *Stud. Educ. Eval.* 55, 167–179. doi: 10.1016/j.stueduc.2017.10.007
- Ravand, H. (2016). Application of a Cognitive Diagnostic Model to a high-stakes reading comprehension test. *J. Psychoeduc. Assess.* 34, 782–799. doi: 10.1177/0734282915623053
- Ravand, H. (2019). "Application of a hierarchical diagnostic classification model in assessing reading comprehension," in *Quantitative Data Analysis for Language Assessment Volume II*, eds V. Aryadoust, and M. Raquel (London; New York, NY: Routledge), 79–98.
- Ravand, H., and Robitzsch, A. (2018). Cognitive diagnostic model of best choice: a study of reading comprehension. *Educ. Psychol.* 38, 1255–1277. doi: 10.1080/01443410.2018.1489524
- Rumelhart, D. E. (1977). "Toward an interactive model of reading," in *Attention and Performance* ed. S. Dornic (Hillsdale, NJ: Erlbaum).
- Sabtu, S. H., and Mohd Matore, M. E. E. (2024). Systematic literature review on the implementation of the Six Sigma approach in education. *Int. J. Eval. Res. Educ.* 13, 262–270. doi: 10.11591/ijere.v13i1.26196
- Safari, F., and Ahmadi, A. (2023). Developing and evaluating an empirically-based diagnostic checklist for assessing second language integrated writing. *J. Second Lang. Writ.* 60, 1–15. doi: 10.1016/j.jslw.2023.101007
- Sakaria, D., Maat, S. M., and Mohd Matore, M. E. E. (2023). Factors influencing mathematics teachers' pedagogical content knowledge (PCK): a systematic review. *Pegem. Egit. Ogr. Derg.* 13, 1–14. doi: 10.47750/pegegog.13.02.01
- Sawaki, Y., Kim, H., and Gentile, C. (2009). Q-Matrix construction: Defining the link between constructs and test items in large-scale reading and listening comprehension assessments. *Lang. Assess. Q.* 6, 190–209. doi: 10.1080/15434300902801917
- Sessoms, J., and Henson, R. A. (2018). Applications of diagnostic classification models: a literature review and critical commentary. *Meas. Interdiscip. Res. Perspect.* 16, 1–17. doi: 10.1080/15366367.2018.1435104
- Shahmirzadi, N., and Marashi, H. (2023). Cognitive diagnostic assessment of reading comprehension for high-stakes tests: Using GDINA model. *Lang. Test Focus Int. J.* 8, 1–16. doi: 10.32038/ltf.2023.08.01
- Sieo, W. M., Amini, M., Lee, K. F., Lin, S. E., Tee, J. N., Siau, C. S., et al. (2022). Reading strategies in teaching and learning English as a foreign language: a mixed-method study. *Pertanika J. Soc. Sci./Humanit.* 30, 475–499. doi: 10.47836/pjssh.30.2.04
- Singh, V. K., Singh, P., Karmakar, M., Leta, J., and Mayr, P. (2021). The journal coverage of Web of Science, Scopus and Dimensions: a comparative analysis. *Scientometrics.* 126, 5113–5142. doi: 10.1007/s11192-021-03948-5
- Sook Yi, Y. (2016). Probing the relative importance of different attributes in L2 reading and listening comprehension items: an application of

- cognitive diagnostic models. *Lang. Test.* 34, 1–9. doi: 10.1177/0265532216646141
- Stanovich, K. E. (1980). Toward an interactive-compensatory model of individual differences in the development of reading fluency. *Read. Res. Q.* 16, 32–71. doi: 10.2307/747348
- Sun, H., and Hwang, Y. (2023). Enhancing EFL learners' English reading performance through the diagnostic feedback of CDA. *Korean J. Engl. Lang. Linguist.* 23, 535–553. doi: 10.15738/kjell.23..202307.535
- Suzuki, Y., and Elgort, I. (2023). "Measuring automaticity in a second language: a methodological synthesis of experimental tasks over three decades (1990–2021)," in *Practice and Automatization in Second Language Research: Perspectives from Skill Acquisition Theory and Cognitive Psychology*, ed. Y. Suzuki (New York, NY: Routledge), 206–34.
- Tabatabaee-Yazdi, M. (2020). Hierarchical diagnostic classification modeling of reading comprehension. *SAGE Open.* 10, 1–13. doi: 10.1177/2158244020931068
- Teng, M. F., and Mizumoto, A. (2024). Validation of metacognitive knowledge in vocabulary learning and its predictive effects on incidental vocabulary learning from reading. *Int. Rev. Appl. Linguist Lang. Teach.* 62, 1–32. doi: 10.1515/iral-2023-0294
- Thi, D. T. D., and Loye, N. (2019). Cognitive diagnostic analyses of the Progress in International Reading Literacy Study (PIRLS) 2011 results. *Mes Éval. Éduc.* 42, 127–166. doi: 10.7202/1084131ar
- Tonekaboni, F. R., Ravand, H., and Rezvani, R. (2021). The construction and validation of a Q-matrix for a high-stakes reading comprehension test: a G-DINA study. *Int. J. Lang. Test.* 11, 58–87. Available online at: https://www.ijlt.ir/article_128361.html
- Toprak, T. E., and Cakir, A. (2021). Examining the L2 reading comprehension ability of adult ELLs: developing a diagnostic test within the cognitive diagnostic assessment framework. *Lang. Test.* 38, 106–131. doi: 10.1177/0265532220941470
- Toprak-yildiz, T. E. (2021). An international comparison using cognitive diagnostic assessment: fourth graders' diagnostic profile of reading skills on PIRLS 2016. *Stud. Educ. Eval.* 70, 1–10. doi: 10.1016/j.stueduc.2021.101057
- Toprak-Yildiz, T. E. (2022). Developing a cognitive model of L2 reading comprehension for evaluative purposes in an EMI context. *Stud. Educ. Eval.* 75, 1–11. doi: 10.1016/j.stueduc.2022.101200
- van der Linden, W. J. (2016). *Handbook of Item Response Theory Volume 1: Model*. London; New York, NY: CRC Press.
- von Davier, M., and Lee, Y. S. (2019). *Handbook of Diagnostic Classification Models*. Cham: Springer.
- Wang, D., Cai, Y., and Tu, D. (2021). Q-matrix estimation methods for Cognitive Diagnosis Models: Based on partial known Q-Matrix. *Multivariate Behav. Res.* 56, 514–26. doi: 10.1080/00273171.2020.1746901
- Wang, M. (2023). Bridging assessment and learning: a cognitive diagnostic analysis of a large-scale Spanish proficiency test. *Porta Lingarum.* 40, 9–24. doi: 10.30827/portalin.vi40.15930
- Yumsek, M. (2023). Educational L2 constructs and diagnostic measurement. *Lang. Test Asia.* 13, 1–23. doi: 10.1186/s40468-022-00214-0
- Zhang, H., Wu, X., and Ju, M. (2024). Developing a cognitive model of solid geometry based on Interpretive Structural Modeling method. *Heliyon.* 10, 1–12. doi: 10.1016/j.heliyon.2024.e27063
- Zhao, H., Wang, W., and Huang, Y. (2020). A cognitive analysis of an English reading test through the GDINA model A cognitive analysis of an English reading test through the G-DINA model. *J. Phys. Conf. Series.* 1629, 1–7. doi: 10.1088/1742-6596/1629/1/012037



OPEN ACCESS

EDITED BY

Antonio Benítez-Burraco,
University of Seville, Spain

REVIEWED BY

Matthew Walenski,
East Carolina University, United States
Mile Vuković,
University of Belgrade, Serbia
Xiaoyun Zhong,
Qingdao University of Technology, China

*CORRESPONDENCE

Gerald C. Imaezue
✉ gimaezue@usf.edu

RECEIVED 16 April 2025

ACCEPTED 24 July 2025

PUBLISHED 14 August 2025

CITATION

Imaezue GC, Ajayi D and Davis C (2025)
Treatment of aphasia in linguistically diverse
populations: current and future directions.
Front. Psychol. 16:1612413.
doi: 10.3389/fpsyg.2025.1612413

COPYRIGHT

© 2025 Imaezue, Ajayi and Davis. This is an
open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](#). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or reproduction
is permitted which does not comply with
these terms.

Treatment of aphasia in linguistically diverse populations: current and future directions

Gerald C. Imaezue*, David Ajayi and Celine Davis

Department of Communication Sciences and Disorders, University of South Florida, Tampa, FL, United States

Aphasia is a multimodal language disorder that affects individuals across all language cultures, disrupting speaking, listening, reading, writing, and gestural communication. Although aphasia is challenging to manage in monolingual individuals, it becomes even more complex in linguistically diverse populations due to factors such as differences in language-specific features, limited linguistically customizable behavioral therapies and clinicians' cross-linguistic competence. This critical review examines current and emerging treatment approaches for linguistically diverse populations, highlighting the progression from traditional behavioral interventions to innovative modalities, including state-of-the-art AI-driven and culturally sensitive interventions designed to overcome cultural and linguistic barriers and enhance therapy outcomes. The review emphasizes the growing need for aphasia care specific to linguistically diverse populations, with a focus on personalized treatment strategies and innovations in digital therapeutics that account for cultural and linguistic nuances. Specifically, we advocate for: (1) personalizing restorative aphasia therapies to users' preferred languages; (2) restorative therapies that leverage universal nonverbal systems and neurobiological modulations as primary treatment modalities; and (3) digital innovations such as multilingual artificial intelligence systems for restorative aphasia therapy, particularly delivered through smartphones. Additionally, ethical considerations—including cultural responsiveness, clinician preparedness, and patient data protection—are discussed to inform future directions in equitable and effective aphasia care. Overall, this study provides insights to guide the development of inclusive and innovative aphasia interventions for linguistically diverse populations.

KEYWORDS

aphasia treatment, multilingual aphasia therapy, restorative aphasia therapy, nonlinguistic aphasia therapy, AI-driven aphasia therapy

1 Introduction

1.1 Aphasia, classification and prevalence

Aphasia is a multimodal language disorder that impairs expressive and receptive language skills and other communication modalities, such as speaking, auditory comprehension, writing, reading and gestural communication (Fridriksson and Hillis, 2021). Its severity, presentation and prognosis for recovery vary and are influenced by the interaction of complex intrinsic factors such as the site of lesion, extent of brain damage, time post onset, and extrinsic factors like treatment use, relevance and accessibility (Fridriksson and Hillis, 2021; Plowman et al., 2012). Stroke remains the leading cause of aphasia, as it accounts for at least a quarter of aphasia cases (Grönberg et al., 2022). However, other conditions such as traumatic brain injury, brain tumor and neurodegenerative diseases could lead to aphasia.

The National Aphasia Association (NAA) estimates that aphasia affects over two million individuals in the United States (Breining and Sebastian, 2020). Aphasia is a significant public health concern due to its adverse impact on functional communication, mobility, and overall quality of life (Bueno-Guerra et al., 2024).

Our study focuses on stroke-induced aphasia. Aphasia is broadly categorized into fluent and nonfluent subtypes. Fluent aphasia, like Wernicke's aphasia or conduction aphasia, is characterized by relatively intact speech production with impaired language comprehension and paragrammatic speech: frequent word substitution errors or word misuse during spontaneous speech (Cordella et al., 2024). Nonfluent aphasia, such as Broca's aphasia or transcortical motor aphasia, involves effortful, halting, fragmented, and agrammatic speech—omission of grammatical elements, with relatively preserved comprehension. Some persons with aphasia (PWA) exhibit severe production and comprehension deficits concurrently, a condition clinically defined as global aphasia (Sheppard and Sebastian, 2021).

1.2 Linguistic diversity in aphasia

Aphasia is a heterogeneous disorder, presenting diverse patholinguistic profiles even among monolingual speakers. These complexities are further amplified in multilingual individuals with aphasia and linguistically diverse populations. Multilingual aphasia specifically refers to cases where individuals acquired and use two or more languages before the onset of aphasia. The assessment and treatment of this condition requires careful consideration of factors such as age of language acquisition, pre- and post-stroke as well as patterns of language use (Ansaldi et al., 2014; Goral and Lerman, 2020). In contrast, aphasia in linguistically diverse populations refers more broadly to individuals who may or may not be multilingual but who reside in multilingual societies or belong to linguistic minority groups (Arslan and Penaloza, 2025). For these individuals, extrinsic factors such as lack of linguistically appropriate treatments and access to multilingual clinicians can pose significant barriers to aphasia care (Goral and Lerman, 2020).

Overall, treatment of aphasia in linguistically diverse populations is limited by differences in language-specific features across cultures which then lead to lack of culturally and linguistically appropriate assessment tools, limited linguistically customizable behavioral therapies and shortage of clinicians with cross-linguistic competence. For instance, most standardized tests are designed for monolingual speakers, making them less accurate for individuals from minority language backgrounds (Noroozi et al., 2021). Access to clinicians who speak the patient's language is also limited, which can affect communication, engagement, and treatment outcomes. In multilingual aphasia, factors like language dominance, proficiency differences, and cross-linguistic interference further complicate diagnosis and therapy planning, which must account for syntactic, phonological, and semantic variation across languages. As global migration increases, there is a growing need for aphasia treatments that are linguistically and culturally responsive (Ansaldi et al., 2014; Centeno and Harris, 2021; Centeno and Laures-Gore, 2024). Meeting this challenge will require the development of language-agnostic and culturally tailored interventions delivered using ubiquitous technologies such as smartphones, and AI-driven multilingual

technologies designed to support equitable and effective aphasia rehabilitation worldwide.

1.3 Aims of the study

This critical narrative review aims to evaluate the current landscape of aphasia treatments in linguistically diverse populations, explore innovative approaches such as personalized and technology-assisted therapies and their applications in multilingual contexts, and discuss strategies for developing linguistically inclusive and accessible treatments for multilingual patients with aphasia. To achieve these objectives, we first examine contemporary treatment approaches for aphasia in monolingual populations and consider their implications for diverse linguistic groups. We then review non-linguistic cognitive strategies that may help overcome language barriers in cross-linguistic settings, and finally, we highlight recent advances in artificial intelligence (AI) and multilingual speech and language technologies with the potential of driving personalized therapies for linguistically diverse populations.

2 Current restorative treatments for aphasia

2.1 Restorative behavioral aphasia therapies

Contemporary approaches to aphasia therapy primarily involve behavioral interventions aimed at restoration of impaired language functions or enabling compensatory strategies for functional communication, with both approaches having implications for improving quality of life (Sheppard and Sebastian, 2021). Restorative aphasia therapy is a key behavioral intervention that plays a crucial role in language recovery by using linguistic tasks to actively target and engage specific impaired linguistic systems, leveraging principles of use-dependent learning and neuroplasticity as well as psycholinguistic and neurolinguistic theories of language processing. Restorative aphasia therapies are recognized as the gold standard for aphasia treatment, employing evidence-based techniques to minimize aphasic speech and language errors and ultimately facilitating PWA's functional communication abilities. These interventions can be broadly categorized based on linguistic domain (e.g., production or comprehension), linguistic level (e.g., word, sentence or discourse), and modality (e.g., spoken, written, auditory or reading).

To address difficulty with spoken language production at the word level, therapies such as Semantic Feature Analysis (SFA) and Phonological Component Analysis (PCA) have been widely used. SFA enhances lexical retrieval by training persons with aphasia to generate and reflect on semantic attributes related to target words, thereby strengthening the semantic network and aiding access to word meaning (Bihovsky et al., 2024; Efstratiadou et al., 2018; Boyle and Coelho, 1995). In contrast, PCA emphasizes phonological characteristics such as syllable structure and initial sounds, enhancing phonological encoding during word retrieval (Leonard et al., 2008; Meteyard and Bose, 2018; Python et al., 2025). These therapies are particularly beneficial for individuals with anomia or other naming deficits, helping to reduce word-finding difficulties. At the sentence level of spoken language production, approaches like

Melodic Intonation Therapy (MIT), Constraint-Induced Language Therapy (CILT) and script training have been used to improve language production in PWA. MIT uses two distinct pitches (high and low) to create melodic speech engaging the relatively preserved right hemisphere to help compensate for impaired left-hemisphere networks (Albert et al., 1973; Conklyn et al., 2012; Norton et al., 2009; Haro-Martínez et al., 2019). CILT, on the other hand, promotes use-dependent learning by restricting the use of compensatory non-verbal modalities (e.g., gestures or writing), thereby forcing reliance on spoken language and encouraging functional sentence production through intensive practice (Meinzer et al., 2005, 2007; Pulvermüller et al., 2001; Raymer and Roitschb, 2023). Script training emphasizes the automatic production of script-based sentences to facilitate functional communication (Cherney et al., 2008; Cherney and Van Vuuren, 2022; Quique et al., 2022; Youmans et al., 2005).

Although language production-focused therapies are well represented in the literature, comprehension-based restorative therapies are equally essential. These interventions often aim to improve auditory or written language comprehension through activities such as sentence-picture matching, syntactic mapping, or treatment of underlying forms, depending on the individual's comprehension profile (Fleming et al., 2021; Thompson, 1996; Thompson and Shapiro, 2005; Swiderski et al., 2021; Wallace et al., 2022; Webster et al., 2013). These techniques help improve communication by engaging language-dominant and compensatory neural pathways to induce language recovery in PWA.

2.2 Restorative neuromodulation and pharmacological therapies

Neuromodulation in aphasia treatment consistently involves the use of noninvasive methods aimed at exciting or inhibiting neural activity in target regions of interest to enhance aphasia recovery. Among noninvasive brain stimulation (NIBS) techniques, repetitive transcranial magnetic stimulation (rTMS) and transcranial direct current stimulation (tDCS) are the most widely studied (Breining and Sebastian, 2020; Gao et al., 2019; Williams et al., 2024). NIBS techniques are primarily used as adjuvants to behavioral aphasia therapy, targeting ipsilesional, perilesional or contralesional language brain regions as well as nonconventional regions like the cerebellum or primary motor cortex to induce language recovery in aphasia (Baker et al., 2010; Chrysikou and Hamilton, 2011; Crosson et al., 2019; Darkow et al., 2017). Studies have shown consistently that pairing restorative language therapy with rTMS or tDCS leads to better language outcomes in patients with nonfluent aphasia than using NIBS alone (Allendorfer et al., 2021a; Breining and Sebastian, 2020; Fridriksson et al., 2018; Raymer and Johnson, 2024). Although the optimal parameters and mechanism of action underpinning the effect of NIBS-induced aphasia recovery remains unclear, studies have consistently shown that inhibition of the contralesional right hemisphere and excitation of ipsilesional or perilesional language regions are optimal for driving language recovery (Allendorfer et al., 2021b; Marangolo et al., 2016; Williams et al., 2024) but see Turkeltaub (2015) for discussions on contradictory evidence regarding the role of contralesional regions in aphasia recovery.

Pharmacological interventions are used to modulate impaired neurochemical pathways to drive language recovery in aphasia. Like

studies on neuromodulation in aphasia care, no studies have conclusively shown that pharmacological interventions alone led to better language recovery in PWA than when combined with restorative behavioral aphasia therapy (Saxena and Hillis, 2017). For instance, evidence from randomized controlled trials suggests that pharmacological agents such as donepezil and memantine may enhance naming, oral expression, auditory comprehension, and repetition in post-stroke aphasia, particularly when combined with restorative behavioral treatments (Cichon et al., 2021; Fridriksson and Hillis, 2021).

While these findings demonstrate the effects of neuromodulation and pharmacological interventions as adjuvants to restorative behavioral aphasia therapy, further studies are needed to establish their efficacy and clinical applicability, especially in diverse cross-linguistic populations.

2.3 Linguistic barriers to restorative aphasia therapies

Restorative therapies for aphasia including behavioral, neuromodulation, and pharmacological interventions have been predominantly studied in speakers of Indo-European languages, particularly English. This linguistic narrowness presents a critical limitation in global aphasia care, as aphasic symptoms do not manifest uniformly across languages. Structural differences among languages can shape both the nature of aphasia and its response to treatment. For instance, languages with rich inflectional morphology, such as Italian or German, often elicit different error patterns (e.g., more substitutions) than less morphologically complex languages like English, which tend to show more omissions (Kiran and Roberts, 2010). These differences necessitate language-specific therapeutic approaches. Clinicians must therefore tailor assessments and interventions to the grammatical and lexical features of each language, accounting for factors such as word order, subject pronoun usage, and verb inflection, rather than applying English-based protocols uncritically.

Restorative aphasia therapies rely heavily on dyadic verbal interactions, which can be compromised when speech-language pathologists (SLPs) do not speak the patient's primary language (Larkman et al., 2024). This linguistic mismatch risks compromising the validity of assessments, reducing patient engagement, and limiting therapeutic outcomes. Moreover, most evidence-based treatment protocols have been designed and validated in English, offering limited guidance for their adaptation to non-English-speaking or culturally diverse populations. Given that over 7,000 languages are spoken globally (Karim and Sima, 2015), and that stroke-induced aphasia occurs across all linguistic and cultural groups, the lack of cross-linguistic generalizability poses a significant barrier to equitable and effective aphasia care.

Restorative therapy accessibility is particularly uneven across language groups, presenting a profound obstacle to universal aphasia care. For instance, English-speaking individuals benefit from a broad array of standardized aphasia tests, treatment resources, and trained clinicians. In contrast, speakers of minority, indigenous and other languages often lack equivalent linguistically-sensitive therapies, normative data-driven assessments, access to cross-linguistically competent clinicians and professional support (Ansaldi et al., 2008;

Ansaldo et al., 2014; Goral and Lerman, 2020; Ivanova and Hallowell, 2013). Multilingual patients with aphasia face unique challenges, including variable proficiency across their spoken languages—a factor often overlooked in conventional therapy approaches (Goral et al., 2023). This disparity means that many individuals must undergo assessment and treatment in a second or majority language, increasing the risk of misdiagnosis and culturally incongruent aphasia care (Centeno and Ansaldo, 2016). Speakers of under-resourced languages are especially vulnerable, as linguistic and cultural barriers have been shown to impede access to comprehensive aphasia services (Centeno, 2009). Although the mechanism of action of some behavioral therapies may have cross-linguistic applicability, e.g., semantic feature analysis (Bihovsky et al., 2024; Kiran and Roberts, 2010), there is no study that have established universal applicability of restorative aphasia treatment. Hence, the global shortage of linguistically appropriate assessments and culturally trained professionals thus limits the scalability and inclusivity of restorative therapies (Arslan and Penaloza, 2025; Centeno and Harris, 2021; Gobbo and Marini, 2024).

Addressing these linguistic and cultural gaps is essential for achieving equitable outcomes in aphasia rehabilitation. Recent scholarship has emphasized the urgent need to develop aphasia assessments in a wider range of languages and to validate existing protocols cross-linguistically (Ivanova and Hallowell, 2013). In parallel, training programs must prepare SLPs to deliver therapy that is not only linguistically but also culturally responsive. Without these efforts, the reach and effectiveness of aphasia therapies will remain restricted to a subset of the global population. Expanding resources beyond hegemonic languages and using language-agnostic tools are therefore prerequisites for ensuring that restorative aphasia therapies are accessible, relevant, and effective for all individuals affected by this disabling condition.

To address these critical needs, we further advocate for: (1) customizing restorative behavioral aphasia therapies to users' preferred languages; (2) restorative therapies that leverage universal nonverbal systems and neurobiological modulations as primary treatment modalities; and (3) digital innovations such as multilingual artificial intelligence systems for restorative aphasia therapy delivered through smartphones.

3 Customizing aphasia therapy to users' preferred languages

3.1 Importance of language preference

Language preference is crucial in aphasia therapy for multilingual patients. Research shows that strategies like targeting dominant languages, alternating treatment languages, focusing on shared linguistic structures, using patients' preferred languages, and incorporating culturally relevant materials improve engagement, neuroplasticity, and recovery in both trained and untrained languages (Ansaldo et al., 2008; Edmonds and Kiran, 2006; Goral et al., 2023; Kiran and Iakupova, 2011; Kiran and Thompson, 2019; Kuzmina et al., 2019; Mooijman et al., 2024). Multilingual patients with aphasia often experience greater preservation and better performance in their first language (L1) compared to later learned languages. Correspondingly, delivering therapy in the preferred language tends to yield greater

recovery gains. For example, Scimeca et al. (2024) found that multilingual patients who received naming therapy in their L1 showed significantly larger improvements in their first language compared to those treated in their second language. Allowing patients to use the language they are most comfortable with also enhances engagement in therapy. Studies have demonstrated that when multilingual patients with aphasia were free to switch between languages at will (reflecting their natural communication preference), they achieved faster and more accurate naming performance in single language conditions (Mooijman et al., 2024). This suggests that respecting a multilingual person's language choice and permitting code-switching can reduce frustration, enhance participation, and leverage all available communicative resources for rehabilitation. Furthermore, tailoring aphasia therapy to a patient's linguistic background and preference has been shown to improve rehabilitation outcomes, particularly, for instance, by addressing syntactic and phonological differences between languages thereby making treatment more linguistically relevant for patients (Edmonds and Kiran, 2006).

Rehabilitation in a patient's preferred or most dominant language pre-stroke may optimally engage intact neural networks and facilitate reorganization. In multilinguals, the language that recovers better is often the most used and the one with stronger pre-stroke connections to cognitive control networks, underscoring the benefit of focusing therapy on the more proficient language (Kuzmina et al., 2019). Additionally, engaging PWA in their preferred language may enhance patient engagement, support recovery, and potentially improve generalization of therapeutic gains across different languages, particularly in cases where the later-learned language (L2) has become the primary language used for everyday conversation (Goral et al., 2023).

In summary, recent evidence from the past decade strongly supports tailoring aphasia therapy to the multilingual person's language preference, as this approach maximizes therapy engagement, encourages adaptive neuroplasticity, and ultimately improves communication outcomes.

3.2 Use of interpreters and cultural mediators

In clinical practice and community-based settings, language interpreters and cultural mediators are vital for delivering effective cross-linguistic aphasia therapy to multilingual patients with aphasia especially when their languages differ from those spoken by clinicians. Professional interpreters facilitate accurate two-way communication, ensuring that clinicians can understand patients' responses and patients can fully comprehend therapy instructions, preventing misunderstandings that could hinder therapy (Larkman et al., 2024). This accuracy is crucial in aphasia, where nuances of language ability need to be precisely gauged; studies have shown that miscommunications (e.g., omissions or alterations of patient responses) can impede diagnosis, goal-setting and cause emotional detachment from the therapeutic process underscoring the need for interpreter training in aphasia care (Babbitt et al., 2022; Larkman et al., 2023).

Equally important, cultural mediators (sometimes called cultural or language brokers) help tailor interventions to the patient's cultural context, bridging gaps in understanding and building trust between

patients and providers. Clinicians are advised to collaborate with both interpreters and cultural mediators to maximize therapy engagement and relevance for multilingual patients. Cultural mediators often take on multiple functions in these settings, acting as patient advocates and navigators who connect individuals and families with resources in a culturally sensitive manner (Sharma et al., 2023). For instance, mediating cultural differences by explaining rehabilitation practices in the light of patient's cultural differences and adjusting activities to be culturally meaningful could make interventions culturally relevant, ultimately enhancing patient's motivation and adherence to therapy activities. Such culturally responsive approaches have been shown to improve therapy engagement and satisfaction, as patients feel understood and respected in their values and communication styles (Sharma et al., 2023). Conversely, the absence of such support is associated with disparities; for example, multilingual patients with aphasia who required interpreters have been found less likely to receive certain evidence-based language interventions and had longer rehabilitation stays, indicating the risk of poorer outcomes without language assistance (Mellahn et al., 2025). This evidence highlights that interpreters and cultural mediators are not just add-ons but essential contributors to equitable and effective clinical aphasia care.

Overall, the combined efforts of interpreters and cultural mediators help ensure that patients with aphasia in cross-linguistic settings receive support that is linguistically accurate and culturally appropriate, leading to significantly long-term recovery and improved quality of life.

3.3 Aphasia assessment in multicultural settings

Although this paper focuses on restorative aphasia treatments, accurate assessment of language proficiency and dominance pre-and poststroke is critical for developing customized treatment plans for patients with aphasia in linguistically diverse settings. Tools such as the Bilingual Aphasia Test (BAT) (Paradis and Libben, 2014), the Language Experience and Proficiency Questionnaire (LEAP-Q) (Marian et al., 2007) and the Bilingual Switching Questionnaire (BSWQ) (Rodriguez-Fornells et al., 2012) provide structured ways to evaluate a patient's language history, dominance, and proficiency across languages. Dynamic formal and informal assessment methods, which assess language skills especially in real-time contexts, provide clinicians with a comprehensive understanding of a patient's communicative abilities (Doedens and Meteyard, 2020; Salter et al., 2006).

Most reported standardized aphasia assessments are designed for monolingual English-speaking populations, limiting their utility in multilingual or culturally diverse contexts. Adapting these tools involves translating and culturally modifying assessment items to ensure relevance and fairness. For instance, the Boston Diagnostic Aphasia Examination (Goodglass et al., 2001); the Western Aphasia Battery-Revised (Kertesz, 2007) and the Bilingual Aphasia Test (Paradis and Libben, 2014) have been adapted for multiple languages, but these versions must consider cultural factors, such as the cultural validity and familiarity of test stimuli (Centeno and Laures-Gore, 2024). Clinicians are encouraged to employ dynamic assessments using culturally appropriate stimuli alongside standardized tools, if

available, to capture nuanced linguistic abilities in underrepresented languages.

4 Toward universal restorative aphasia therapy

4.1 Recruiting nonlinguistic cognitive processes for language recovery

The rationale behind recruiting universal cognitive systems in aphasia rehabilitation is particularly significant when considering patients from diverse linguistic backgrounds. The fundamental cognitive processes that underpin language processing, such as attention, executive function and working memory, are believed to be universal across cultures. A growing body of research consistently supports the complex relationship between language, language recovery and nonlinguistic cognitive processes in adults with and without aphasia (Albert et al., 1973; Diedrichs et al., 2022; Dignam et al., 2017; Gilmore et al., 2019; Lambon Ralph et al., 2010; Seniów et al., 2009). Meta-analyses of neural activity in patients with aphasia performing language tasks reveal that their brains recruit regions beyond the traditional language networks (LaCroix et al., 2021). These include areas involved in domain-general cognitive processes across both cerebral hemispheres, with activation patterns similar to those of healthy controls (LaCroix et al., 2021). These insights suggest that therapies targeting these relatively preserved cognitive systems could enhance language recovery by leveraging relatively intact cognitive processes in aphasia patients, thus effectively addressing linguistic barriers in aphasia care.

A growing body of empirical evidence has explored the effectiveness of nonlinguistic cognitive training, often administered in conjunction with traditional aphasia therapy, in facilitating aphasia recovery particularly in stroke survivors. Studies, including a network meta-analysis evaluating various cognitive training interventions, found that combining computer-assisted cognitive training, conventional cognitive training, virtual reality-based cognitive training, telerehabilitation computer-assisted cognitive training, cognitive stimulation training, working memory training, or attention training with traditional aphasia therapy improved language outcomes in patients with aphasia than traditional aphasia therapy alone (Kong et al., 2024). Studies have further demonstrated that the incorporation of gradual attention training and programs targeting executive functioning alongside conventional aphasia therapy led to improvements in a range of language abilities, including reading comprehension, auditory comprehension, verbal sentence production and comprehension, naming, speech repetition, functional communication, and verbal fluency (Culicetto et al., 2025).

The studies discussed combined cognitive training with conventional aphasia therapy, which may be limited by linguistic diversity. Another research, while not focused on cognitive training, explores the use of intrinsic self-feedback mechanisms—specifically recursive self-feedback (RSF) for aphasia recovery. RSF is a self-directed automated procedure that enables individuals to iteratively monitor, evaluate, and adjust their own spoken output—typically by listening to recordings of their own speech and providing self-correction or elaboration at each cycle—without relying on external cues or clinician guidance (Imaezue et al., 2023). RSF has shown

promise in improving language outcomes, including sentence production, spontaneous speech and target language production in both monolingual and multilingual PWA (Imaezue, 2024a; Imaezue et al., 2024b; Imaezue et al., 2025). Intrinsic self-feedback mechanisms, such as error monitoring and correction, may function independently of specific linguistic features, potentially making RSF a language-independent approach to aphasia care (Imaezue, 2024a, Imaezue et al., 2024b). RSF is a promising and useful procedure in the toolbox of clinicians to meet the language needs of PWA in diverse linguistic populations. However, the evidence-based for this technique is based on few PWA which suggests that more research is needed to test the hypothesis of universal applicability of this procedure for aphasia care. By targeting fundamental intrinsic-cognitive feedback mechanisms, RSF may offer a more universally applicable approach to rehabilitation than cognitive training combined with conventional therapy.

4.2 Multisensory stimulation for language recovery

Language barriers can significantly hinder the effectiveness of behavioral aphasia treatment, particularly when therapy relies heavily on verbal instructions. Multisensory stimulation offers a viable solution by engaging multiple nonverbal sensory modalities and neural pathways to support language recovery. Incorporating visual, auditory, tactile, and proprioceptive cues reduces reliance on verbal communication, enhancing accessibility and reinforcing learning (Diedrichs et al., 2022; Schuell et al., 1955).

Multisensory techniques transcend linguistic and cultural differences because they primarily rely on fundamental sensory and motor experiences rather than language-specific structures. Visual cues, gestures, facial expressions, and tactile feedback are universally understood and may not require proficiency in a particular language. For example, gesture-based cueing, such as pointing or hand movements that mimic actions or iconic gestural description of objects conveys meaning across cultures and can aid patients with fluent and nonfluent aphasia across cultures in understanding and expressing themselves (Attard et al., 2013; Rose et al., 2016; Wadams et al., 2022). Although multisensory techniques are often considered adaptable across linguistic and cultural contexts, particularly those involving gestures, gesture-based cues such as pointing and facial expressions may not always be interpreted uniformly across cultures and should therefore be applied with cultural sensitivity and individualized consideration (Efron, 1941; Noroozi et al., 2021).

A longstanding debate in aphasia rehabilitation concerns whether nonverbal and multisensory treatments can generalize to improvements in language such as spontaneous speech, which is a key marker of functional communication in real-world contexts. Some studies report little to no direct generalization to spontaneous speech following gesture-based interventions (Marshall et al., 2012; Marshall et al., 2013; Clough and Duff, 2020). In contrast, other studies suggest that nonverbal treatments can support language recovery indirectly, by improving naming or functioning as compensatory strategies, especially when used within a scaffolding framework, where gestures are gradually withdrawn as verbal abilities return (Rose et al., 2002; Rose et al., 2013; Clough and Duff, 2020). Additionally, multisensory approaches that pair speech with visual or motor cues, such as

speech-associated gestures, have been shown through neuroimaging to reduce cognitive effort and increase neural activations during language production, likely by engaging shared motor-language networks (Skipper et al., 2007). Collectively, these findings suggest that while nonverbal treatments may not always lead to direct gains in language tasks like spontaneous speech, their integration into multisensory, individualized therapies could enhance functional communication and support language recovery post-stroke.

Similarly, rhythmic and melodic elements in therapy, such as those used in MIT, tap into music's cross-cultural potential to bypass linguistic barriers while facilitating speech production (Albert et al., 1973; Sparks and Holland, 1976; van der Meulen et al., 2014; Van der Meulen et al., 2016; Haro-Martínez et al., 2019; Gu et al., 2024). The adaptability of MIT has been demonstrated in tonal languages such as Mandarin (Gu et al., 2024) and syllable-timed languages like Spanish (Haro-Martínez et al., 2017), highlighting its flexibility across diverse prosodic systems. However, while music engages broadly shared cognitive and emotional systems, it is important to acknowledge that musical structures, such as scales, rhythm, and harmonic conventions, vary significantly across cultures and may affect therapeutic resonance and familiarity (Shapiro, 2005).

Likewise, visual supports such as culturally appropriate visual scene cues and pictorial representations can offer nonverbal context without requiring verbal explanation, supporting communication across language boundaries (Cohn, 2020). Yet, pictorial systems are not inherently universal; visual conventions, symbolism, and narrative structure are culturally mediated and may influence how such stimuli are interpreted (Čeněk and Čeněk, 2015; Cho and Ishida, 2011). Taken together, these findings underscore the need to select and adapt musical and visual materials with cultural sensitivity. When tailored to the individual's cultural and aesthetic background, these modalities can serve as effective tools for enhancing engagement, scaffolding verbal expression, and supporting therapy in cross-linguistic and cross-cultural settings.

Furthermore, combining auditory repetition with visual semantic cues for aphasia therapy has been shown to significantly improve naming abilities in aphasic patients (Rose et al., 2016). The use of visual motor imagery and action observation, based on mirror-neurons and frontoparietal mechanisms, have been shown to facilitate language outcomes in patients with aphasia by engaging multiple sensory modalities simultaneously, thereby promoting aphasia recovery (Lee et al., 2010; Rizzolatti and Craighero, 2004).

By leveraging these universally accessible sensory modalities, multisensory and nonverbal-based techniques ensure that treatment remains effective and adaptable for patients with aphasia from diverse cultural and linguistic backgrounds. Further research is suggested to explore this idea.

4.3 Neurobiological approaches as primary treatment modalities

Behavioral aphasia therapy has firmly established its role as an effective intervention for aphasia recovery (Breitenstein et al., 2017; Robey, 1998). However, the emerging evidence surrounding NIBS and pharmacological interventions presents a compelling case for their potential to evolve beyond their current status as adjunctive therapies in aphasia rehabilitation (Raymer and Johnson, 2024). The

mechanisms of action of NIBS in promoting neuroplasticity through direct neural modulation and the capacity of pharmacological agents to modulate neurochemical pathways suggest a future where these neurobiological approaches could take on new roles as primary treatment modalities with greater potentials for worldwide applications than conventional behavioral therapies which are constrained by linguistic barriers. To attain this shift, however, several critical advancements are necessary. Continued rigorous research is essential to optimize NIBS protocols, including refining and personalizing stimulation parameters, developing more precise targeting strategies, and determining the optimal timing for intervention relative to the onset of aphasia. For instance, integrating neuroimaging and genomics into aphasia rehabilitation allows identifying neural and genetic markers associated with language recovery, which may help tailor pharmacological, NIBS and behavioral interventions to enhance individual treatment outcomes (Fridriksson et al., 2018; Kristinsson and Fridriksson, 2022; Shah-Basak et al., 2020; Wang et al., 2014).

Extensive and well-controlled clinical trials are needed to definitively establish the efficacy and safety of promising and novel pharmacological agents, as well as to determine the most effective ways to combine them with behavioral therapies for synergistic benefits. The development of personalized treatment strategies, guided by individual patient characteristics, detailed neuroimaging data, and the specific nature of their language impairments, will be crucial for maximizing the effectiveness of both NIBS and pharmacological interventions (Gkintoni and Michou, 2024; Saxena and Hillis, 2017). Technological advancements, particularly in the realm of advanced neuroimaging techniques and the development of artificial intelligence (AI)-driven personalized neurobiological treatment protocols, will play a significant role in facilitating this transition. Furthermore, advancements in stem cell therapy may have implications for neural repair poststroke and aphasia recovery in cross-linguistic populations (Baker et al., 2019). Addressing these current limitations and knowledge gaps through focused research and continued innovation is essential to fully unlock the potential of neurobiological interventions and transform the future of universal aphasia care.

In this evolving landscape, behavioral therapies could transition to a complementary role, focusing on functional communication and the generalization of gains achieved through neurobiological modalities. The realization of this future holds the promise of a paradigm shift in aphasia care, offering the potential for more effective and comprehensive rehabilitation strategies that address the underlying neurological basis of this debilitating condition while also leveraging the crucial contributions of behavioral therapies.

5 Digital innovations for accessible aphasia care

5.1 Smartphone-based aphasia therapy for accessible care

In response to global challenges like the COVID-19 pandemic, speech-language therapy shifted from in-person services to telehealth due to restrictions on in-person healthcare delivery. Telehealth emerged as an effective alternative, enabling patients with aphasia to continue receiving timely and consistent therapy from clinicians

primarily through video conferencing platforms and other remote communication technologies (Dekhtyar et al., 2020; Hill et al., 2009; Taiebne and Keegan, 2024). This shift has also expanded access to care for multilingual and culturally diverse populations by facilitating connections with bilingual clinicians, integrating real-time translation and captioning tools, and allowing therapy to be tailored to individual linguistic needs, albeit across limited languages (Penaloza et al., 2021).

Telehealth-based aphasia therapy commonly relies on desktop or laptop computers, which may pose accessibility challenges (Brandenburg et al., 2013; Raymond et al., 2024). Many patients with aphasia, particularly those with co-occurring motor impairments, may find these devices difficult to use due to their size and the need for fine motor control. Additionally, desktops and laptops are less portable and may not support the flexible, frequent practice sessions that are essential for aphasia recovery. This highlights the need for mobile health (mHealth) solutions, particularly smartphone-based therapy, which offers a more accessible, portable, and user-friendly alternative with broader worldwide reach. Smartphones are widely available, affordable, and can be equipped with speech-language therapy apps, voice-activated controls, and adaptive accessibility features that can facilitate self-directed practice and clinician-guided interventions. The convenience afforded by the portability of smartphones allows patients with aphasia to engage in therapy from virtually any location at a time that suits their schedule, potentially leading to improved language outcomes, accessibility, adherence and a reduced burden on both patients and caregivers (Braley et al., 2021; Brandenburg et al., 2013; Imaezue and Goral, 2025; Jiang et al., 2024). By leveraging mHealth technologies, smartphone-based aphasia therapy presents a compelling practical avenue for addressing the limitations of traditional telehealth and enhancing inclusivity worldwide and responsiveness to the diverse needs of patients with aphasia (Jiang et al., 2024). However, smartphone-based interventions are not without limitations. The small screen size, variable hardware quality and device accessibility, poor user interface and experience design, word retrieval focused tasks, and the need for app-specific familiarity may pose accessibility, usability and engagement challenges for some users (Brandenburg et al., 2017; Greig et al., 2008; Nichol et al., 2021; Pontus et al., 2025; Vaezipour et al., 2020). Moreover, reliable internet access and device compatibility can vary across regions, potentially limiting scalability in low-resource contexts (Li and Wilson, 2013; McCool et al., 2022).

A growing number of specialized and evidence-based aphasia therapy mobile apps (e.g., Constant Therapy, Lingraphica and Tactus Therapy) are now available for smartphones, providing a rich array of tools for addressing various aphasia deficits. These apps often incorporate therapeutic techniques and exercises designed by SLPs, offering structured and engaging opportunities for patients with aphasia to practice their communication skills outside of formal therapy sessions (Nikolaev and Nikolaev, 2022; Vaismoradi et al., 2024). Notably, many of these platforms support multiple languages: Constant Therapy is available in English (US, India) and Spanish; Lingraphica's Talk App supports dual language mode (English and Spanish); and Tactus Therapy offers full or partial translations in up to ten languages, including French, German, Dutch, and Zulu. This multilingual accessibility significantly enhances global reach and ensures that individuals with aphasia from linguistically diverse backgrounds can engage with therapy materials in their preferred or native language.

Studies have demonstrated the benefits of smartphone-delivered RSF for both scripted-sentence production and spontaneous speech in linguistically diverse individuals with aphasia (Imaezue, 2024a; Imaezue, 2024b; Imaezue et al., 2023; Imaezue et al., 2024b). RSF involves repeated playback and rehearsal of the patient's own speech, enabling users to refine their output using intrinsic self-monitoring and correction mechanisms rather than external cueing. This approach is particularly advantageous for linguistically diverse populations because it uses the patient's own speech as the therapeutic stimulus. As a result, RSF inherently preserves the speaker's native accent, dialect, and phonological patterns. This is especially relevant given that accent, dialectal variation, and prosodic features can significantly influence speech intelligibility and therapeutic engagement—factors that are often underrecognized in conventional aphasia treatments (Goral and Hejazi, 2021; Mooijman et al., 2024). By bypassing the need for externally modeled stimuli, RSF ensures that feedback is linguistically and culturally congruent, which may facilitate more efficient reactivation of residual language networks. Furthermore, RSF is compatible with basic smartphone functionality (e.g., audio/video recording), making it highly accessible and cost-effective, even in under-resourced or remote contexts. These attributes make RSF an ideal procedure to bridge gaps in aphasia care for minority-language speakers or those without access to language-matched clinicians, as it allows them to rehearse speech using materials in their preferred language and phonetic system.

5.2 Personalized aphasia therapy using artificial intelligence

Artificial Intelligence (AI) is revolutionizing aphasia assessment and therapy by enabling more language-specific, culturally adaptive, and personalized interventions (Azevedo et al., 2024). Machine learning models, natural language processing, and automated speech analysis enhance diagnostic accuracy, predict recovery trajectories, and customize treatment strategies. Studies have shown that AI-powered applications can make aphasia treatment more precise, data-driven and evidence-based (Azevedo et al., 2024; Medenica et al., 2024).

Computerized interventions, such as virtual reality (VR), create immersive, gamified environments that engage patients and simulate real-world communication challenges, enhancing motivation and language practice (Cao et al., 2021). Beyond language outcomes, these non-linguistic approaches help address cultural, logistical, and linguistic barriers, as well as the shortage of speech-language therapists, making aphasia rehabilitation more accessible (Jacobs et al., 2021). An example is EVA Park, a multi-user virtual environment designed to enable individuals with aphasia to engage with their speech pathologists and connect with other PWA (Galliers et al., 2017). AI-driven speech technologies are increasingly being integrated into aphasia rehabilitation, enhancing accessibility and allowing patients to receive therapy remotely through tablet and computer-based software (Jiang et al., 2024). Innovative technology-based interventions, such as those examined by Repetto et al. (2021) provide opportunities for more intensive practice, increasing treatment intensity and leading to improved language recovery outcomes. Additionally, these interventions offer greater flexibility, enabling patients to select tasks and practice at their own pace, thereby fostering motivation and engagement in therapy (Repetto et al., 2021). In

addition to enhancing accessibility and treatment intensity, speech recognition and synthesis technologies such as text-to-speech converters and voice banking systems help PWA improve their expressive language skills (Latif et al., 2021).

While promising, AI is not without limitations. AI systems are prone to generating plausible yet inaccurate outputs, often described as “hallucinations” or, more appropriately, “AI misinformation, which may mislead patients or clinicians if used without oversight (Hatem et al., 2023; Ray and Majumder, 2023). Ensuring clinical accuracy, safeguarding patient data, and incorporating human review mechanisms will be essential to responsibly integrating AI into aphasia care. To realize the full potential of AI and digital tools in aphasia care, it is essential to expand linguistic coverage and incorporate cross-linguistic adaptability into design, including the ability to accommodate code-switching, mixed-language use, and regionally distinct language norms. Addressing these gaps is essential for AI-based interventions to deliver equitable, scalable aphasia therapy for linguistically diverse populations (Privitera et al., 2024).

5.3 Multilingual artificial intelligence for universal aphasia therapy

The integration of large language models (LLMs) into aphasia therapy presents promising opportunities for the development of multilingual AI agents capable of delivering individualized interventions across diverse linguistic contexts. LLMs are advanced artificial intelligence systems that leverage deep learning—particularly transformer-based architectures—to model, predict, and generate human language in both monolingual and multilingual formats (Annepaka and Pakray, 2025; Vaswani et al., 2023). Contemporary LLMs are not only multilingual but also multimodal, allowing them to process and generate language across various sensory modalities (e.g., text, speech, image, video). These capabilities make LLMs well-suited and highly promising for supporting aphasia therapy in diverse populations by enabling tailored interventions that span languages and communication modalities.

LLMs have been integrated into the management of neurological conditions such as dementia and psychosis, particularly in diagnosing and predicting language disorders (de Arriba-Pérez and García-Méndez, 2024). Studies have indicated that it could also be integrated into aphasia care by improving language assessment, enabling personalized therapy, and facilitating real-time feedback, ultimately improving communication outcomes for PWA (Themistocleous, 2024; Zhong, 2024). For instance, LLMs have shown potential to detect paraphasic errors, identify agrammatic structures, and refine automatic speech recognition (Cong et al., 2024). Specifically, Cong et al. (2024) conducted a large-scale NLP study examining pre-trained LLMs across English and Mandarin speech samples and found that while English models achieved moderate accuracy in aphasia detection, Chinese models performed poorly, highlighting significant discrepancies in performance across languages. Similarly, a study by van V et al. (2024) reported that applying LLM-based procedures to patients with Broca's aphasia improved speech corrections and syntactic organization, addressing their characteristic halted and fragmented speech patterns. A recent pilot study described the development and initial testing of Aphasia-GPT, a mobile web app designed to support people with aphasia by providing suggested

utterances based on users' spoken input. Testing with three individuals with moderate aphasia indicated the system currently functions best as a sentence production aid for non-fluent aphasia with relatively preserved articulation, though further improvements are needed to support a wider range of speech profiles and communicative needs (Bailey et al., 2024). These findings highlight the potential of LLMs and AI, particularly when paired with expert-in-the-loop, in revolutionizing aphasia therapy by integrating AI-driven language processing and human expert guidance into clinical practice.

While these studies are encouraging, they also reinforce concerns about equity and accuracy in underrepresented languages. Without sufficient, ethnolinguistically representative training data, even state-of-the-art models may struggle to reliably detect or support therapy in heritage, minority, or low-resource languages. LLMs can misinterpret culturally specific syntax, discourse patterns, or error types—potentially undermining assessment and intervention outcomes (Yari and Koto, 2025; Zhou et al., 2025). Moreover, models may produce plausible but incorrect or misleading outputs if used without clinician validation.

Multilingual conversational AI agents, built using multilingual LLMs, have the potential to adapt to more diverse aphasic errors than previously studied which make them potentially useful for clinical applications. For instance, our research employed a novel approach—*Agent-Based Conversational Dialogue* (ABCD)—to simulate context embedded therapeutic speech interactions between an AI-based therapist and an AI-modeled aphasic patient. The patient agent was designed to vocally imitate characteristic word- and discourse-level errors observed in English-speaking individuals with aphasia, enabling ecologically valid modeling of therapeutic dialogue. The study used GPT4o as the core LLM for both AI-agents and their behaviors were modulated using well-crafted prompts rather than traditional finetuning of pretrained models using large dataset of aphasic speech data. Our findings showed that the AI-therapist's conversational performance, in terms of context and logically-relevant utterances, were highly accurate despite aphasic speech inputs such as language mixing errors in English and Spanish, tangential utterances, and semantic, phonemic, perseveration and syntactic word errors (Imaezue and Marampelly, 2025). We further replicated the performance of the AI-therapist in an ABCD simulated setting across other languages such as Spanish and Hindi (Imaezue et al., 2024a).

While these preclinical findings are highly promising, further research is needed to test the responsiveness and accuracy of AI-driven therapy and AI-driven therapy plus expert-in-the-loop with cross-linguistic patients with aphasia across diverse quantitative and qualitative metrics. Furthermore, more studies are required to demonstrate the clinical efficacy of these systems, especially administered via smartphones, on aphasia recovery in humans. Notwithstanding, multilingual LLMs and related technologies are still limited to only a few languages. This suggests that its potential clinical applicability is still constrained by linguistic factors.

5.4 Ethical and practical considerations

As AI-driven tools and telehealth platforms are increasingly integrated into aphasia therapy, safeguarding patient data has become an essential ethical concern (ElHennawy, 2024). Abujaber and Nashwan (2024) argued that ethical frameworks for AI in healthcare must account

for cultural and linguistic diversity at all stages, from data acquisition to algorithmic deployment. By foregrounding these issues, we propose that data privacy and security are not peripheral concerns but essential components of linguistically responsive digital therapy. The storage and transmission of sensitive health information must adhere to stringent data protection regulations to prevent breaches and unauthorized access (Abujaber and Nashwan, 2024). Clinicians and developers must implement encryption, secure authentication protocols, and consent-based data-sharing policies to maintain patient trust and confidentiality. Moreover, the use of AI models trained on patient speech data necessitates strict measures to ensure de-identification and protect individuals' privacy while enhancing therapy effectiveness (Medenica et al., 2024). Failure to address these privacy concerns may undermine the widespread adoption of these technologies, particularly in culturally sensitive contexts where trust is paramount (Marks and Haupt, 2023).

5.5 Training and resource allocation

As multilingual and AI-assisted therapies continue to gain traction, there is a growing need for clinician training in both technological and linguistic competencies. SLPs must be equipped with the skills necessary to utilize AI-driven diagnostic tools, telehealth platforms, and multilingual therapy strategies effectively (Azevedo et al., 2024). Collaboration with interdisciplinary teams, including neurologists, psychologists, computer scientists and software developers, is essential in refining these interventions and ensuring their effective integration into clinical practice (Hinckley and Jayes, 2023). Furthermore, resource allocation is vital to support the development of culturally inclusive therapy programs, which can help bridge existing treatment gaps in multilingual aphasia care (Jacobs and Ellis, 2025). By investing in these areas, the healthcare system can ensure that new technologies for aphasia and related disorders are not only innovative but also accessible and relevant to diverse cross-linguistic patient populations.

6 Conclusion

In conclusion, this critical review underscores the evolving landscape of aphasia research in culturally and linguistically diverse populations, highlighting both traditional and emerging restorative treatment approaches. Although progress has been made in addressing multilingual contexts, further research is essential to advance neurobiologically grounded restorative therapies as primary treatment modalities for diverse language speakers. There is also a growing need to explore mobile health solutions and develop multilingual, multimodal AI-based therapies tailored to multicultural communities. Expanding these areas will support the creation of more inclusive, effective, and culturally responsive aphasia interventions—ultimately paving the way towards universal aphasia care.

Author contributions

GI: Writing – review & editing, Supervision, Writing – original draft, Methodology, Conceptualization, Investigation, Validation. DA: Writing – original draft, Writing – review & editing, Investigation. CD: Writing – original draft, Investigation, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This Financial support was received from the University of South Florida for the publication of this article.

Conflict of interest

Author GI is the founder of Converstation AI LLC.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Abujaber, A. A., and Nashwan, A. J. (2024). Ethical framework for artificial intelligence in healthcare research: a path to integrity. *World J. Methodol.* 14:94071. doi: 10.5662/wjm.v14.i3.94071
- Albert, M. L., Sparks, R. W., and Helm, N. A. (1973). Melodic intonation therapy for aphasia. *Arch. Neurol.* 29, 130–131. doi: 10.1001/archneur.1973.00490260074018
- Allendorfer, J. B., Nenert, R., Nair, S., Vannest, J., and Szaflarski, J. P. (2021a). Functional magnetic resonance imaging of language following constraint-induced aphasia therapy primed with intermittent theta burst stimulation in 13 patients with post-stroke aphasia. *Med. Sci. Monit.* 27:e930100. doi: 10.12659/MSM.930100
- Allendorfer, J. B., Nenert, R., Vannest, J., and Szaflarski, J. P. (2021b). A pilot randomized controlled trial of intermittent theta burst stimulation as stand-alone treatment for post-stroke aphasia: effects on language and verbal functional magnetic resonance imaging (fMRI). *Med. Sci. Monit.* 27:e934818. doi: 10.12659/MSM.934818
- Annepaka, Y., and Pakray, P. (2025). Large language models: a survey of their development, capabilities, and applications. *Knowl. Inf. Syst.* 67, 2967–3022. doi: 10.1007/s10115-024-02310-4
- Ansaldi, A. I., Marcotte, K., Scherer, L., and Raboyeau, G. (2008). Language therapy and bilingual aphasia: clinical implications of psycholinguistic and neuroimaging research. *J. Neurolinguistics* 21, 539–557. doi: 10.1016/j.jneuroling.2008.02.001
- Ansaldi, A. I., Saidi, L. G., and Abutalebi, J. (2014). Aphasia therapy in the age of globalization: cross-linguistic therapy effects in bilingual aphasia. *Behav. Neurol.* 2014:603085. doi: 10.1155/2014/603085
- Arslan, S., and Penalzoza, C. (2025). Across countries and cultures: the assessment of aphasia in linguistically diverse clinical populations. *Aphasiology* 1–6, 1–6. doi: 10.1080/02687038.2025.2468546
- Attard, M. C., Rose, M. L., and Lanyon, L. (2013). The comparative effects of multi-modality aphasia therapy and constraint-induced aphasia therapy-plus for severe chronic Broca's aphasia: an in-depth pilot study. *Aphasiology* 27, 80–111. doi: 10.1080/02687038.2012.725242
- Azevedo, N., Kehayia, E., Jarema, G., Le Dorze, G., Beaujard, C., and Yvon, M. (2024). How artificial intelligence (AI) is used in aphasia rehabilitation: a scoping review. *Aphasiology* 38, 305–336. doi: 10.1080/02687038.2023.2189513
- Babbitt, E. M., Ginsberg-Jaeckle, M., Larkin, E., Escarcega, S., and Cherney, L. R. (2022). Classifying interpreter behaviors during aphasia assessments: survey results and checklist development. *Am. J. Speech Lang. Pathol.* 31, 2329–2347. doi: 10.1044/2022_AJSLP-21-00306
- Bailey, D. J., Herget, F., Hansen, D., Burton, F., Pitt, G., Harmon, T., et al. (2024). Generative AI applied to AAC for aphasia: a pilot study of aphasia-GPT. *Aphasiology* 1–16, 1–16. doi: 10.1080/02687038.2024.2445663
- Baker, E. W., Kinder, H. A., and West, F. D. (2019). Neural stem cell therapy for stroke: a multimethod approach to restoring neurological function. *Brain Behav.* 9:e01214. doi: 10.1002/brb3.1214
- Baker, J. M., Rorden, C., and Fridriksson, J. (2010). Using transcranial direct-current stimulation to treat stroke patients with aphasia. *Stroke* 41, 1229–1236. doi: 10.1161/STROKEAHA.109.576785
- Bihovsky, A., Ben-Shachar, M., and Meir, N. (2024). Semantic feature analysis (SFA) treatment in L1 and L2 in bilingual aphasia: effects of cognitive and language factors. *Aphasiology* 38, 683–711. doi: 10.1080/02687038.2023.2226359
- Boyle, M., and Coelho, C. A. (1995). Application of semantic feature analysis as a treatment for aphasic Dysnomia. *Am. J. Speech Lang. Pathol.* 4, 94–98. doi: 10.1044/1058-0360.0404.94
- Braley, M., Pierce, J. S., Saxena, S., De Oliveira, E., Taraboanta, L., Anantha, V., et al. (2021). A virtual, randomized, control trial of a digital therapeutic for speech, language, and cognitive intervention in post-stroke persons with aphasia. *Front. Neurol.* 12:626780. doi: 10.3389/fneur.2021.626780
- Brandenburg, C., Worrall, L., Copland, D., and Rodriguez, A. D. (2017). Barriers and facilitators to using the CommFit™ smart phone app to measure talk time for people with aphasia. *Aphasiology* 31, 901–927. doi: 10.1080/02687038.2016.1219016
- Brandenburg, C., Worrall, L., Rodriguez, A. D., and Copland, D. (2013). Mobile computing technology and aphasia: an integrated review of accessibility and potential uses. *Aphasiology* 27, 444–461. doi: 10.1080/02687038.2013.772293
- Breining, B. L., and Sebastian, R. (2020). Neuromodulation in post-stroke aphasia treatment. *Curr. Phys. Med. Rehabil. Rep.* 8, 44–56. doi: 10.1007/s40141-020-00257-5
- Breitenstein, C. D., Grewe, T. P., Flöel, A. P., Ziegler, W. P., Springer, L. P., Martus, P. P., et al. (2017). Intensive speech and language therapy in patients with chronic aphasia after stroke: a randomised, open-label, blinded-endpoint, controlled trial in a health-care setting. *Lancet* 389, 1528–1538. doi: 10.1016/S0140-6736(17)30067-3
- Bueno-Guerra, N., Provencio, M., Tarifa-Rodríguez, A., Navarro, A., Sempere-Iborra, C., Jordi, P., et al. (2024). Impact of post-stroke aphasia on functional communication, quality of life, perception of health and depression: a case-control study. *Eur. J. Neurol.* 31:e16184. doi: 10.1111/ene.16184
- Cao, Y., Huang, X., Zhang, B., Kranz, G. S., Zhang, D., Li, X., et al. (2021). Effects of virtual reality in post-stroke aphasia: a systematic review and meta-analysis. *Neurol. Sci.* 42, 5249–5259. doi: 10.1007/s10072-021-05202-5
- Čeněk, J., and Čeněk, Š. (2015). Cross-cultural differences in visual perception. *J. Educ. Cult. Soc.* 6, 187–206. doi: 10.15503/jecs20151.187.206
- Centeno, J. G. (2009). Issues and principles in service delivery to communicatively impaired minority bilingual adults in neurorehabilitation. *Semin. Speech Lang.* 30, 139–152. doi: 10.1055/s-0029-1225951
- Centeno, J. G., and Ansaldi, A. I. (2016). Customizing treatment for bilingual speakers with aphasia: bilingual clients with aphasia have unique first-versus second-language experiences before their stroke. That's why clinicians need to incorporate individually tailored bilingualism principles into rehabilitation. *ASHA Leader* 21, 52–56. doi: 10.1044/leader.FTR2.21102016.52
- Centeno, J. G., and Harris, J. L. (2021). Implications of United States service evidence for growing multiethnic adult neurorehabilitation caseloads worldwide. *Can. J. Speech Lang. Pathol. Audiol.* 45:77.
- Centeno, J. G., and Laures-Gore, J. S. (2024). Intersectionality in aphasia services for ethnically diverse adult populations. *Semin. Speech Lang.* 45, 1–4. doi: 10.1055/s-0043-1778016
- Cherney, L. R., Halper, A. S., Holland, A. L., and Cole, R. (2008). Computerized script training for aphasia: preliminary results. *Am. J. Speech Lang. Pathol.* 23, SupplS343–S360. doi: 10.1016/j.jcomdis.2011.04.002
- Cherney, L. R., and Van Vuuren, S. (2022). Complexity and feedback during script training in aphasia: a feasibility study. *Arch. Phys. Med. Rehabil.* 103, S205–S214. doi: 10.1016/j.apmr.2022.03.002
- Cho, H., and Ishida, T. (2011). “Exploring cultural differences in pictogram interpretations” in *The language grid*, Cognitive Technologies. ed. T. Ishida (Berlin, Heidelberg: Springer).
- Chrysikou, E. G., and Hamilton, R. H. (2011). Noninvasive brain stimulation in the treatment of aphasia: exploring interhemispheric relationships and their implications for neurorehabilitation. *Restor. Neurol. Neurosci.* 29, 375–394. doi: 10.3233/RNN-2011-0610
- Cichon, N., Włodarczyk, L., Saluk-Bijak, J., Bijak, M., Redlicka, J., Gorniak, L., et al. (2021). Novel advances to post-stroke aphasia pharmacology and rehabilitation. *J. Clin. Med.* 10:3778. doi: 10.3390/jcm10173778

Generative AI statement

The authors declare that no Gen AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Clough, S., and Duff, M. C. (2020). The role of gesture in communication and cognition: implications for understanding and treating neurogenic communication disorders. *Front. Hum. Neurosci.* 14:323. doi: 10.3389/fnhum.2020.00323
- Cohn, N. (2020). Visual narrative comprehension: universal or not? *Psychon. Bull. Rev.* 27, 266–285. doi: 10.3758/s13423-019-01670-1
- Cong, Y., LaCroix, A. N., and Lee, J. (2024). Clinical efficacy of pre-trained large language models through the lens of aphasia. *Scientific Reports*, 14, 15573–15516. doi: 10.1038/s41598-024-66576-y
- Conklyn, D., Novak, E., Boissy, A., Bethoux, F., and Chemali, K. (2012). The effects of modified melodic intonation therapy on nonfluent aphasia: a pilot study. *J. Speech Lang. Hear. Res.* 55, 1463–1471. doi: 10.1044/1092-4388(2012/11-0105)
- Cordella, C., Di Filippo, L., Kolachalama, V. B., and Kiran, S. (2024). Connected speech fluency in poststroke and progressive aphasia: a scoping review of quantitative approaches and features. *Am. J. Speech Lang. Pathol.* 33, 2083–2120. doi: 10.1044/2024_AJSLP-23-00208
- Crosson, B., Rodriguez, A. D., Copland, D., Fridriksson, J., Krishnamurthy, L. C., Meinzer, M., et al. (2019). Neuroplasticity and aphasia treatments: new approaches for an old problem. *J. Neurol. Neurosurg. Psychiatry* 90, 1147–1155. doi: 10.1136/jnnp-2018-319649
- Culicetto, L., Latella, D., Lo Buono, V., Orecchio, F., Murdaca, A. M., Quartarone, A., et al. (2025). Executive functions training improves language abilities in aphasia rehabilitation: a systematic review. *J. Pers. Med.* 15:92. doi: 10.3390/jpm15030092
- Darkow, R., Martin, A., Würzt, A., Flöel, A., and Meinzer, M. (2017). Transcranial direct current stimulation effects on neural processing in post-stroke aphasia. *Hum. Brain Mapp.* 38, 1518–1531. doi: 10.1002/hbm.23469
- de Arriba-Pérez, F., and García-Méndez, S. (2024). Leveraging large language models through natural language processing to provide interpretable machine learning predictions of mental deterioration in real time. *Arab. J. Sci. Eng.* doi: 10.1007/s13369-024-09508-2
- Dekhlyar, M., Braun, E. J., Billot, A., Foo, L., and Kiran, S. (2020). Videoconference administration of the Western aphasia battery–revised: feasibility and validity. *Am. J. Speech Lang. Pathol.* 29, 673–687. doi: 10.1044/2019_AJSLP-19-00023
- Diedrichs, V. A., Jewell, C. C., and Harnish, S. M. (2022). A scoping review of the relationship between nonlinguistic cognitive factors and aphasia treatment response. *Top. Lang. Disord.* 42, 212–235. doi: 10.1097/TL0.0000000000000290
- Dignam, J., Copland, D., O'Brien, K., Burfein, P., Khan, A., and Rodriguez, A. D. (2017). Influence of cognitive ability on therapy outcomes for anomia in adults with chronic poststroke aphasia. *J. Speech Lang. Hear. Res.* 60, 406–421. doi: 10.1044/2016_JSLHR-L-15-0384
- Doedens, W. J., and Meteyard, L. (2020). Measures of functional, real-world communication for aphasia: a critical review. *Aphasiology* 34, 492–514. doi: 10.1080/02687038.2019.1702848
- Edmonds, L. A., and Kiran, S. (2006). Effect of semantic naming treatment on crosslinguistic generalization in bilingual aphasia. *J. Speech Lang. Hear. Res.* 49, 729–748. doi: 10.1044/1092-4388(2006/053)
- Efron, D. (1941). *Gesture and Environment*. New York: King's Crown Press.
- Efstratiadou, E. A., Papathanasiou, I., Holland, R., Archonti, A., and Hilari, K. (2018). A systematic review of semantic feature analysis therapy studies for aphasia. *J. Speech Lang. Hear. Res.* 61, 1261–1278. doi: 10.1044/2018_JSLHR-L-16-0330
- ElHennawy, S. M. (2024). The impact of artificial intelligence (AI) in the assessment and treatment of communication disorders (a review of literature). Egypt: J. Lang. Eng.
- Fleming, V., Brownsett, S., Krasen, A., Maegli, M. A., Coley-Fisher, H., Ong, Y.-H., et al. (2021). Efficacy of spoken word comprehension therapy in patients with chronic aphasia: a cross-over randomised controlled trial with structural imaging. *J. Neurol. Neurosurg. Psychiatry* 92, 418–424. doi: 10.1136/jnnp-2020-324256
- Fridriksson, J., Elm, J., Stark, B. C., Basilakos, A., Rorden, C., Sen, S., et al. (2018). BDNF genotype and tDCS interaction in aphasia treatment. *Brain Stimul.* 11, 1276–1281. doi: 10.1016/j.brs.2018.08.009
- Fridriksson, J., and Hillis, A. E. (2021). Current approaches to the treatment of post-stroke aphasia. *J. Stroke* 23, 183–201. doi: 10.5853/jos.2020.05015
- Fridriksson, J., Rorden, C., Elm, J., Sen, S., George, M. S., and Bonilha, L. (2018). Transcranial direct current stimulation vs sham stimulation to treat aphasia after stroke: a randomized clinical trial. *Arch. Neurol.* 75, 1470–1476. doi: 10.1001/jamaneurol.2018.2287
- Galliers, J., Wilson, S., Marshall, J., Talbot, R., Devane, N., Booth, T., et al. (2017). Experiencing EVA Park, a multi-user virtual world for people with aphasia. *ACM Trans. Access. Comput.* 10, 1–24. doi: 10.1145/3134227
- Gao, F., Cai, Q., Li, Z., Chen, P., Fang, H., Hao, J., et al. (2019). The effect of rTMS over the different targets on language recovery in stroke patients with global aphasia: a randomized sham-controlled study. *Biomed. Res. Int.* 2019, 1–7. doi: 10.1155/2019/4589056
- Gilmore, N., Meier, E. L., Johnson, J. P., and Kiran, S. (2019). Nonlinguistic cognitive factors predict treatment-induced recovery in chronic poststroke aphasia. *Arch. Phys. Med. Rehabil.* 100, 1251–1258. doi: 10.1016/j.apmr.2018.12.024
- Gkintoni, E., and Michou, E. (2024). Advancing neuropsychological rehabilitation in primary progressive aphasia based on principles of cognitive neuroscience: a scoping review and systematic analysis of the data. *Brain Sci.* 14:1234. doi: 10.3390/brainsci14121234
- Gobbo, M., and Marini, A. (2024). Transcranial alternating current stimulation applied to language recovery in persons with aphasia: a scoping review. *Aphasiology* 39, 684–709. doi: 10.1080/02687038.2024.2373135
- Goodglass, H., Kaplan, E., and Weintraub, S. (2001). BDAE: The Boston diagnostic aphasia examination. Philadelphia, PA: Lippincott Williams & Wilkins.
- Goral, M., and Hejazi, Z. (2021). Aphasia in multilingual patients. *Curr. Neurol. Neurosci. Rep.* 21:60. doi: 10.1007/s11910-021-01148-5
- Goral, M., and Lerman, A. (2020). Variables and mechanisms affecting response to language treatment in multilingual people with aphasia. *Behav. Sci.* 10:1. doi: 10.3390/BS10090144
- Goral, M., Norvik, M. I., Antfolk, J., Agrotou, I., and Lehtonen, M. (2023). Cross-language generalization of language treatment in multilingual people with post-stroke aphasia: a meta-analysis. *Brain Lang.* 246:105326. doi: 10.1016/j.bandl.2023.105326
- Greig, C.-A., Harper, R., Hirst, T., Howe, T., and Davidson, B. (2008). Barriers and facilitators to mobile phone use for people with aphasia. *Top. Stroke Rehabil.* 15, 307–324. doi: 10.1310/tsr1504-307
- Grönberg, A., Henriksson, I., Stenman, M., and Lindgren, A. G. (2022). Incidence of aphasia in ischemic stroke. *Neuroepidemiology* 56, 174–182. doi: 10.1159/000524206
- Gu, J., Long, W., Zeng, S., Li, C., Fang, C., and Zhang, X. (2024). Neurologic music therapy for non-fluent aphasia: a systematic review and meta-analysis of randomized controlled trials. *Front. Neurol.* 15:1395312. doi: 10.3389/fneur.2024.1395312
- Haro-Martínez, A. M., García-Concejero, V. E., López-Ramos, A., Maté-Arribas, E., López-Táppero, J., Lubrini, G., et al. (2017). Adaptation of melodic intonation therapy to Spanish: a feasibility pilot study. *Aphasiology* 31, 1333–1343. doi: 10.1080/02687038.2017.1279731
- Haro-Martínez, A. M., Lubrini, G., Madero-Jarabo, R., Díez-Tejedor, E., and Fuentes, B. (2019). Melodic intonation therapy in post-stroke nonfluent aphasia: a randomized pilot trial. *Clin. Rehabil.* 33, 44–53. doi: 10.1177/0269215518791004
- Hatem, R., Simmons, B., and Thornton, J. E. (2023). A Call to Address AI “Hallucinations” and How Healthcare Professionals Can Mitigate Their Risks. *Cureus (Palo Alto, CA)*, 15:e44720. doi: 10.7759/cureus.44720
- Hill, A. J., Theodoros, D. G., Russell, T. G., Ward, E. C., and Wootton, R. (2009). The effects of aphasia severity on the ability to assess language disorders via telerehabilitation. *Aphasiology* 23, 627–642. doi: 10.1080/02687030801909659
- Hinckley, J., and Jayes, M. (2023). Person-centered care for people with aphasia: tools for shared decision-making. *Front. Rehabil. Sci.* 4:1236534. doi: 10.3389/fresc.2023.1236534
- Imaezue, G. C. (2024a). Recursive self-feedback improves spontaneous speech in chronic aphasia within real-world settings. *Aphasiology*, 1–21. doi: 10.1080/02687038.2024.2432024
- Imaezue, G. C. (2024b). Transfer effects of recursive self-feedback on connected speech production in patients with chronic nonfluent aphasia: preliminary results. *Aphasiology* 39, 346–362. doi: 10.1080/02687038.2024.2351010
- Imaezue, G. C., Davis, C., Maram, K. V., Ajayi, D., and Goral, M. (2025). Recursive self-feedback enables target language production in aphasia with pathological language mixing. *Brain Lang* 270:105633. doi: 10.1016/j.bandl.2025.105633
- Imaezue, G. C., and Goral, M. (2025). Effect of feedback type and self-selected practice schedule on speech fluency during aphasia script treatment. *J. Commun. Disord.* 117:106553. doi: 10.1016/j.jcomdis.2025.106553
- Imaezue, G. C., and Marampelly, H. (2025). ABCD: a simulation method for accelerating conversational agents with applications in aphasia therapy. *J. Speech Lang. Hear. Res.* 68, 3322–3336. doi: 10.1044/2025_JSLHR-25-00003
- Imaezue, G. C., Marampelly, H., Azemard, J., Mineo, G., and Salas, C. (2024a). Multilingual AI-Agent for Language Therapy (MALT) in diverse languages: A feasibility study with English, Hindi and Spanish speakers. Presented at the 2nd University of South Florida Artificial Intelligence + X Symposium, Tampa, Florida.
- Imaezue, G. C., Tchernichovski, O., and Goral, M. (2024b). Self-improved language production in nonfluent aphasia through automated recursive self-feedback. *Am. J. Speech Lang. Pathol.* 33, 3343–3357. doi: 10.1044/2024_AJSLP-23-00320
- Imaezue, G. C., Tchernichovski, O., and Goral, M. (2023). Recursive self-feedback improved speech fluency in two patients with chronic nonfluent aphasia. *Aphasiology*, 38, 838–861. doi: 10.1080/02687038.2023.2239511
- Ivanova, M. V., and Hallowell, B. (2013). A tutorial on aphasia test development in any language: key substantive and psychometric considerations. *Aphasiology* 27, 891–920. doi: 10.1080/02687038.2013.805728
- Jacobs, M., Briley, P. M., Fang, X., and Ellis, C. (2021). Telepractice treatment for aphasia: association between clinical outcomes and client satisfaction. *Telemed. Rep.* 2, 118–124. doi: 10.1089/tmr.2020.0024
- Jacobs, M., and Ellis, C. (2025). More than a communication disorder: inequities in the financial toxicity of post-stroke aphasia. *Front. Stroke* 3. doi: 10.3389/fstro.2024.1507025

- Jiang, Z., He, M., Zhang, C., and Chen, X. (2024). The effect of mobile application-based technology on post-stroke aphasia: a systematic review. *Front. Neurol.* 15:1405209. doi: 10.3389/fneur.2024.1405209
- Karim, S., and Sima, K. (2015). Languages of the world: an introduction. *Can. J. Linguist.* 60, 91–94. doi: 10.1017/S0008413100000578
- Kertesz, A. (2007). Western Aphasia Battery–Revised (WAB-R) [Test kit]. San Antonio, TX: PsychCorp (Pearson Clinical).
- Kiran, S., and Iakupova, R. (2011). Understanding the relationship between language proficiency, language impairment and rehabilitation: evidence from a case study. *Clin. Linguist. Phon.* 25, 565–583. doi: 10.3109/02699206.2011.566664
- Kiran, S., and Roberts, P. M. (2010). Semantic feature analysis treatment in Spanish-English and French-English bilingual aphasia. *Aphasiology* 24, 231–261. doi: 10.1080/02687030902958365
- Kiran, S., and Thompson, C. K. (2019). Neuroplasticity of language networks in aphasia: advances, updates, and future challenges. *Front. Neurol.* 10:295. doi: 10.3389/fneur.2019.00295
- Kong, Q., Wang, J., Huang, X., He, J., and Chang, J. (2024). Comparative efficacy of cognitive training for post-stroke aphasia: a systematic review and network meta-analysis. *Neurorehabil. Neural Repair* 38, 863–876. doi: 10.1177/15459683241274755
- Kristinsson, S., and Fridriksson, J. (2022). “Chapter 19- genetics in aphasia recovery” in *Handb. Clin. Neurol.* eds. A. E. Hillis and J. Fridriksson, vol. 185 (Elsevier), 283–296. doi: 10.1016/B978-0-12-823384-9.00015-3
- Kuzmina, E., Goral, M., Norvik, M., and Weekes, B. S. (2019). What influences language impairment in bilingual aphasia? A meta-analytic review. *Front. Psychol.* 10:445. doi: 10.3389/fpsyg.2019.00445
- LaCroix, A. N., James, E., and Rogalsky, C. (2021). Neural resources supporting language production vs. comprehension in chronic post-stroke aphasia: a meta-analysis using activation likelihood estimates. *Front. Hum. Neurosci.* 15:680933. doi: 10.3389/fnhum.2021.680933
- Lambon Ralph, M. A., Snell, C., Fillingham, J. K., Conroy, P., and Sage, K. (2010). Predicting the outcome of anomia therapy for people with aphasia post CVA: both language and cognitive status are key predictors. *Neuropsychol. Rehabil.* 20, 289–305. doi: 10.1080/09602010903237875
- Larkman, C. S., Lanyon, L., and Rose, M. L. (2024). “It’s so complicated”: a qualitative study of interpreters’ experiences working with speech pathologists to support the provision of aphasia rehabilitation. *Disabil. Rehabil.* 47, 3637–3648. doi: 10.1080/09638288.2024.2435518
- Larkman, C. S., Mellahn, K., Han, W., and Rose, M. L. (2023). Aphasia rehabilitation when speech pathologists and clients do not share the same language: a scoping review. *Aphasiology* 37, 635–657. doi: 10.1080/02687038.2022.2035672
- Latif, S., Qadir, J., Qayyum, A., Usama, M., and Younis, S. (2021). Speech technology for healthcare: opportunities, challenges, and state of the art. *IEEE Rev. Biomed. Eng.* 14, 342–356. doi: 10.1109/RBME.2020.3006860
- Lee, J., Fowler, R., Rodney, D., Cherney, L., and Small, S. L. (2010). IMITATE: an intensive computer-based treatment for aphasia based on action observation and imitation. *Aphasiology* 24, 449–465. doi: 10.1080/02687030802714157
- Leonard, C., Rochon, E., and Laird, L. (2008). Treating naming impairments in aphasia: findings from a phonological components analysis treatment. *Aphasiology* 22, 923–947. doi: 10.1080/02687030701831474
- Li, J., and Wilson, L. S. (2013). Telehealth trends and the challenge for infrastructure. *Telemed. J. E-Health* 19, 772–779. doi: 10.1089/tmj.2012.0324
- Marangolo, P., Fiori, V., Sabatini, U., De Pasquale, G., Razzano, C., Caltagirone, C., et al. (2016). Bilateral transcranial direct current stimulation language treatment enhances functional connectivity in the left hemisphere: preliminary data from aphasia. *J. Cogn. Neurosci.* 28, 724–738. doi: 10.1162/jocn_a_00927
- Marian, V., Blumenfeld, H. K., and Kaushanskaya, M. (2007). The language experience and proficiency questionnaire (LEAP-Q): assessing language profiles in bilinguals and multilinguals. *J. Speech Lang. Hear. Res.* 50, 940–967. doi: 10.1044/1092-4388(2007/067)
- Marks, M., and Haupt, C. E. (2023). AI chatbots, health privacy, and challenges to HIPAA compliance. *JAMA* 330, 309–310. doi: 10.1001/jama.2023.9458
- Marshall, J., Best, W., Cocks, N., Cruice, M., Pring, T., Bulcock, G., et al. (2012). Gesture and naming therapy for people with severe aphasia: a group study. *J. Speech Lang. Hear. Res.* 55, 726–738. doi: 10.1044/1092-4388(2011/11-0219)
- Marshall, J., Roper, A., Galliers, J., Wilson, S., Cocks, N., Muscroft, S., et al. (2013). Computer delivery of gesture therapy for people with severe aphasia. *Aphasiology* 27, 1128–1146. doi: 10.1080/02687038.2013.786803
- McCool, J., Dobson, R., Whittaker, R., and Paton, C. (2022). Mobile health (mHealth) in low and middle-income countries. *Annu. Rev. Public Health* 43, 525–539. doi: 10.1146/annurev-publhealth-052620-093850
- Medenica, V., Ivanovic, L., and Milosevic, N. (2024). Applicability of artificial intelligence in neuropsychological rehabilitation of patients with brain injury. *Appl. Neuropsychol. Adult* 1–28, 1–28. doi: 10.1080/23279095.2024.2364229
- Meinzer, M., Djundja, D., Barthel, G., Elbert, T., and Rockstroh, B. (2005). Long-term stability of improved language functions in chronic aphasia after constraint-induced aphasia therapy. *Stroke* 36, 1462–1466. doi: 10.1161/01.STR.0000169941.29831.2a
- Meinzer, M., Elbert, T., Djundja, D., Taub, E., and Rockstroh, B. (2007). Extending the constraint-induced movement therapy (CIMT) approach to cognitive functions: constraint-induced aphasia therapy (CIAT) of chronic aphasia. *Neuro Rehabilitation* 22, 311–318. doi: 10.3233/NRE-2007-22409
- Mellahn, K., Kilkenny, M. F., Siyambalapitiya, S., Lakhani, A., Burns, C., Purvis, T., et al. (2025). Requiring an interpreter influences stroke care and outcomes for people with aphasia during inpatient rehabilitation. *Stroke* 56, 716–724. doi: 10.1161/STROKEAHA.124.047893
- Meteyard, L., and Bose, A. (2018). What does a cue do? Comparing phonological and semantic cues for picture naming in aphasia. *J. Speech Lang. Hear. Res.* 61, 658–674. doi: 10.1044/2017_JSLHR-L-17-0214
- Mooijman, S., Schoonen, R., Roelofs, A., and Ruiter, M. B. (2024). Benefits of free language choice in bilingual individuals with aphasia. *Aphasiology* 38, 1793–1831. doi: 10.1080/02687038.2024.2326239
- Nichol, L., Wallace, S. J., Pitt, R., Rodriguez, A. D., Diong, Z. Z., and Hill, A. J. (2021). People with aphasia share their views on self-management and the role of technology to support self-management of aphasia. *Disabil. Rehabil.* 44, 7399–7412. doi: 10.1080/09638288.2021.1989501
- Nikolaev, V. A., and Nikolaev, A. A. (2022). Recent trends in telerehabilitation of stroke patients: a narrative review. *Neuro Rehabilitation* 51, 1–22. doi: 10.3233/NRE-210330
- Noroozi, F., Corneanu, C. A., Kaminska, D., Sapinski, T., Escalera, S., and Anbarjafari, G. (2021). Survey on emotional body gesture recognition. *IEEE Trans. Affect. Comput.* 12, 505–523. doi: 10.1109/TAFFC.2018.2874986
- Norton, A., Zipse, L., Marchina, S., and Schlaug, G. (2009). Melodic intonation therapy: shared insights on how it is done and why it might help. *Ann. N. Y. Acad. Sci.* 1169, 431–436. doi: 10.1111/j.1749-6632.2009.04859.x
- Paradis, M., and Libben, G. (2014). The assessment of bilingual aphasia (1st ed.). New York, NY, USA & Hove, East Sussex, UK: Psychology Press. doi: 10.4324/9781315802138
- Penaloza, C., Scimeca, M., Gaona, A., Carpenter, E., Mukadam, N., Gray, T., et al. (2021). Telerehabilitation for word retrieval deficits in bilinguals with aphasia: effectiveness and reliability as compared to in-person language therapy. *Front. Neurol.* 12:589330. doi: 10.3389/fneur.2021.589330
- Plowman, E., Hentz, B., and Ellis, C. (2012). Post-stroke aphasia prognosis: a review of patient-related and stroke-related factors. *J. Eval. Clin. Pract.* 18, 689–694. doi: 10.1111/j.1365-2753.2011.01650.x
- Pontus, S., Shuanglan, L., and Leonardo Horn, I. (2025). Usability and accessibility in mHealth stroke apps: an empirical assessment. *Inf. Med. Unlocked* 53:101616. doi: 10.1016/j.imu.2025.101616
- Privitera, A. J., Ng, S. H. S., Kong, A. P.-H., and Weekes, B. S. (2024). AI and aphasia in the digital age: a critical review. *Brain Sci.* 14:383. doi: 10.3390/brainsci14040383
- Pulvermuller, F., Neininger, B., Elbert, T., Mohr, B., Rockstroh, B., Koebbel, P., et al. (2001). Constraint-induced therapy of chronic aphasia after stroke. *Stroke* 32, 1621–1626. doi: 10.1161/01.STR.32.7.1621
- Python, G., Durand, E., and Masson-Trottier, M. (2025). A systematic review of phonological components analysis therapy studies for aphasia. *Brain Res. Bull.* 223:111269. doi: 10.1016/j.brainresbull.2025.111269
- Quike, Y. M., Evans, W. S., Ortega-Llebaria, M., Zipse, L., and Dickey, M. W. (2022). Get in sync: active ingredients and patient profiles in scripted-sentence learning in Spanish speakers with aphasia. *J. Speech Lang. Hear. Res.* 65, 1478–1493. doi: 10.1044/2021_jslhr-21-00060
- Raymer, A. M., and Johnson, R. K. (2024). Effectiveness of transcranial direct current stimulation as an adjuvant to aphasia treatment following stroke: evidence from systematic reviews and meta-analyses. *Am. J. Speech Lang. Pathol.* 33, 3431–3443. doi: 10.1044/2024_AJSLP-23-00312
- Ray, P. P., and Majumder, P. (2023). The potential of ChatGPT to transform healthcare and address ethical challenges in artificial intelligence-driven medicine. *J. Clin. Neuro.* 19, 509–511. doi: 10.3988/jcn.2023.0158
- Raymer, A. M., and Roitschb, J. (2023). Effectiveness of constraint-induced language therapy for aphasia: evidence from systematic reviews and meta-analyses. *Am. J. Speech Lang. Pathol.* 32, 2393–2401. doi: 10.1044/2022_AJSLP-22-00248
- Raymond, M. J., Christie, L. J., Kramer, S., Malaguti, C., Mok, Z., Gardner, B., et al. (2024). Delivery of allied health interventions using telehealth modalities: a rapid systematic review of randomized controlled trials. *Healthcare (Basel)* 12:1217. doi: 10.3390/healthcare12121217
- Repetto, C., Paolillo, M. P., Tuena, C., Bellinzona, F., and Riva, G. (2021). Innovative technology-based interventions in aphasia rehabilitation: a systematic review. *Aphasiology* 35, 1623–1646. doi: 10.1080/02687038.2020.1819957
- Rizzolatti, G., and Craighero, L. (2004). The mirror-neuron system. *Annu. Rev. Neurosci.* 27, 169–192. doi: 10.1146/annurev.neuro.27.070203.144230
- Robey, R. R. (1998). A meta-analysis of clinical outcomes in the treatment of aphasia. *J. Speech Lang. Hear. Res.* 41, 172–187. doi: 10.1044/jslhr.4101.172
- Rodriguez-Fornells, A., Kraemer, U. M., Lorenzo-Seve, U., Festman, J., and Munte, T. F. (2012). Self-assessment of individual differences in language switching. *Front. Psychol.* 3:388. doi: 10.3389/fpsyg.2011.00388

- Rose, M., Douglas, J., and Matyas, T. (2002). The comparative effectiveness of gesture and verbal treatments for a specific phonologic naming impairment. *Aphasiology* 16, 1001–1030. doi: 10.1080/02687030143000825
- Rose, M. L., Mok, Z., Carragher, M., Katthagen, S., and Attard, M. (2016). Comparing multi-modality and constraint-induced treatment for aphasia: a preliminary investigation of generalisation to discourse. *Aphasiology* 30, 678–698. doi: 10.1080/02687038.2015.1100706
- Rose, M. L., Raymer, A. M., Lanyon, L. E., and Attard, M. C. (2013). A systematic review of gesture treatments for post-stroke aphasia. *Aphasiology* 27, 1090–1127. doi: 10.1080/02687038.2013.805726
- Salter, K., Jutai, J., Foley, N., Hellings, C., and Teasell, R. (2006). Identification of aphasia post stroke: a review of screening assessment tools. *Brain Inj.* 20, 559–568. doi: 10.1080/02699050600744087
- Saxena, S., and Hillis, A. E. (2017). An update on medications and noninvasive brain stimulation to augment language rehabilitation in post-stroke aphasia. *Expert. Rev. Neurother.* 17, 1091–1107. doi: 10.1080/14737175.2017.1373020
- Schuell, H., Carroll, V., and Street, B. S. (1955). Clinical treatment of aphasia. *J. Speech Hear. Disord.* 20, 43–53. doi: 10.1044/jshd.2001.43
- Scimeca, M., Penalzoa, C., and Kiran, S. (2024). Multilevel factors predict treatment response following semantic feature-based intervention in bilingual aphasia. *Biling. (Camb. Engl.)* 27, 246–262. doi: 10.1017/S1366728923000391
- Seńiów, J., Litwin, M., and Leńniak, M. (2009). The relationship between non-linguistic cognitive deficits and language recovery in patients with aphasia. *J. Neurol. Sci.* 283, 91–94. doi: 10.1016/j.jns.2009.02.315
- Shah-Basak, P. P., Sivaratnam, G., Teti, S., Francois-Nienaber, A., Yossofzai, M., Armstrong, S., et al. (2020). High definition transcranial direct current stimulation modulates abnormal neurophysiological activity in post-stroke aphasia. *Sci. Rep.* 10:19625. doi: 10.1038/s41598-020-76533-0
- Shapiro, N. (2005). Sounds in the world: multicultural influences in music therapy in clinical practice and training. *Music. Ther. Perspect.* 23, 29–35. doi: 10.1093/mtp/23.1.29
- Sharma, R. K., Cowan, A., Gill, H., and Miyagawa, L. A. (2023). Understanding the role of caseworker-cultural mediators in addressing healthcare inequities for patients with limited-English proficiency: a qualitative study. *J. Gen. Intern. Med.* 38, 1190–1199. doi: 10.1007/s11606-022-07816-7
- Sheppard, S. M., and Sebastian, R. (2021). Diagnosing and managing post-stroke aphasia. *Expert. Rev. Neurother.* 21, 221–234. doi: 10.1080/14737175.2020.1855976
- Skipper, J. I., Goldin-Meadow, S., Nusbaum, H. C., and Small, S. L. (2007). Speech-associated gestures, Broca's area, and the human mirror system. *Brain Lang.* 101, 260–277. doi: 10.1016/j.bandl.2007.02.008
- Sparks, R. W., and Holland, A. L. (1976). Method: melodic intonation therapy for aphasia. *J. Speech Hear. Disord.* 41, 287–297. doi: 10.1044/jshd.41.03.287
- Swiderski, A. M., Quique, Y. M., Dickey, M. W., and Hula, W. D. (2021). Treatment of underlying forms: a Bayesian meta-analysis of the effects of treatment and person-related variables on treatment response. *J. Speech Lang. Hear. Res.* 64, 4308–4328. doi: 10.1044/2021_JSLHR-21-00131
- Taiebne, M., and Keegan, L. C. (2024). Telepractice for bilingual and multilingual people with aphasia: a scoping review. *Aphasiology* 1–23, 1–23. doi: 10.1080/02687038.2024.2373434
- Themistocleous, C. (2024). Open brain AI and language assessment. *Front. Hum. Neurosci.* 18:1421435. doi: 10.3389/fnhum.2024.1421435
- Thompson, C. K. (1996). Linguistic-specific sentence production treatment for agrammatic aphasia. *Top. Stroke Rehabil.* 3, 60–85. doi: 10.1080/10749357.1996.11754114
- Thompson, C. K., and Shapiro, L. P. (2005). Treating agrammatic aphasia within a linguistic framework: treatment of underlying forms. *Aphasiology* 19, 1021–1036. doi: 10.1080/02687030544000227
- Turkeltaub, P. E. (2015). Brain stimulation and the role of the right hemisphere in aphasia recovery. *Curr. Neurol. Neurosci. Rep.* 15:72. doi: 10.1007/s11910-015-0593-6
- Vaezipour, A., Campbell, J., Theodoros, D., and Russell, T. (2020). Mobile apps for speech-language therapy in adults with communication disorders: review of content and quality. *JMIR Mhealth Uhealth* 8:e18858. doi: 10.2196/18858
- Vaismoradi, M., Rae, J., Turunen, H., and Logan, P. A. (2024). Specialized nurses' role in ensuring patient safety within the context of telehealth in home care: a scoping review. *Digit. Health* 10:20552076241287272. doi: 10.1177/20552076241287272
- Van der Meulen, I., Van de Sandt-Koenderman, M. W. M. E., Heijnenbrok, M. H., Visch-Brink, E., and Ribbers, G. M. (2016). Melodic intonation therapy in chronic aphasia: evidence from a pilot randomized controlled trial. *Front. Hum. Neurosci.* 10:533. doi: 10.3389/fnhum.2016.00533
- van der Meulen, I., van de Sandt-Koenderman, W. M. E., Heijnenbrok-Kal, M. H., Visch-Brink, E. G., and Ribbers, G. M. (2014). The efficacy and timing of melodic intonation therapy in subacute aphasia. *Neurorehabil. Neural Repair* 28, 536–544. doi: 10.1177/1545968313517753
- van V, S., Matussevych, Y., and Tsiwah, F. (2024). Generating completions for fragmented Broca's aphasic sentences using large language models. Ithaca: Cornell University Library.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2023). Attention is all you need. Ithaca: Cornell University library.
- Wadams, A., Suting, L., Lindsey, A., and Mozeiko, J. (2022). Metacognitive treatment in acquired brain injury and its applicability to aphasia: a systematic review. *Front. Rehabil. Sci.* 3:813416. doi: 10.3389/fresc.2022.813416
- Wallace, S. E., Patterson, J., Purdy, M., Knollman-Porter, K., and Coppens, P. (2022). Auditory comprehension interventions for people with aphasia: a scoping review. *Am. J. Speech Lang. Pathol.* 31, 2404–2420. doi: 10.1044/2022_AJSLP-21-00297
- Wang, C.-P., Hsieh, C.-Y., Tsai, P.-Y., Wang, C.-T., Lin, F.-G., and Chan, R.-C. (2014). Efficacy of synchronous verbal training during repetitive transcranial magnetic stimulation in patients with chronic aphasia. *Stroke* 45, 3656–3662. doi: 10.1161/STROKEAHA.114.007058
- Webster, J., Morris, J., Connor, C., Horner, R., McCormac, C., and Potts, A. (2013). Text level reading comprehension in aphasia: what do we know about therapy and what do we need to know? *Aphasiology* 27, 1362–1380. doi: 10.1080/02687038.2013.825760
- Williams, E. E. R., Sghirripa, S., Rogasch, N. C., Hordacre, B., and Attrill, S. (2024). Non-invasive brain stimulation in the treatment of post-stroke aphasia: a scoping review. *Disabil. Rehabil.* 46, 3802–3826. doi: 10.1080/09638288.2023.2259299
- Yari, A. H., and Koto, F. (2025). Unveiling cultural blind spots: Analyzing the limitations of mLLMs in procedural text comprehension. arXiv:2502.14315. doi: 10.48550/arXiv.2502.14315
- Youmans, G., Holland, A., Munoz, M., and Bourgeois, M. (2005). Script training and automaticity in two individuals with aphasia. *Aphasiology* 19, 435–450. doi: 10.1080/02687030444000877
- Zhong, X. (2024). AI-assisted assessment and treatment of aphasia: a review. *Front. Public Health* 12:1401240. doi: 10.3389/fpubh.2024.1401240
- Zhou, R., Wan, G., Gabriel, S., Li, S., Gates, A. J., Sap, M., et al. (2025). Disparities in llm reasoning accuracy and explanations: A case study on African American English. arXiv:2503.04099. doi: 10.48550/arXiv.2503.04099



OPEN ACCESS

EDITED BY

Thomas L. Spalding,
University of Alberta, Canada

REVIEWED BY

Hasan Alisoy,
Nakhchivan State University, Azerbaijan
Mirela Cunha Cardoso Ramacciotti,
University of São Paulo, Brazil

*CORRESPONDENCE

Sergey Kiselev
✉ s.j.kiselev@urfu.ru
Ningkun Xiao
✉ fellones.zero@gmail.com

RECEIVED 15 July 2025

ACCEPTED 13 August 2025

PUBLISHED 29 August 2025

CITATION

Pu L, Kiselev S and Xiao N (2025) Language and cognitive function in children: a narrative review of neural, behavioral, and developmental evidence.
Front. Psychol. 16:1666719.
doi: 10.3389/fpsyg.2025.1666719

COPYRIGHT

© 2025 Pu, Kiselev and Xiao. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Language and cognitive function in children: a narrative review of neural, behavioral, and developmental evidence

Ling Pu¹, Sergey Kiselev ^{1*} and Ningkun Xiao ^{2*}

¹Laboratory for Brain and Neurocognitive Development, Department of Psychology, Institution of Humanities, Ural Federal University, Yekaterinburg, Russia, ²Department of Immunochimistry, Institution of Chemical Engineering, Ural Federal University, Yekaterinburg, Russia

Language is not merely a conduit for thought—it plays an active, constitutive role in shaping cognitive development. This narrative review synthesizes interdisciplinary findings across bilingualism, theory of mind, developmental disorders (DLD and ASD), and cross-cultural studies to propose a dynamic, context-sensitive model of the language-cognition relationship. We argue that language functions not only as a cognitive tool but as a cognitive architect, influencing the structure and function of neural networks supporting executive function and social cognition. Evidence from behavioral and neuroimaging studies reveals bidirectional and developmentally contingent interactions between language and cognition, moderated by linguistic structure, developmental timing, and sociocultural context. By examining both typical and atypical populations, we challenge modular and unidirectional models, advocating instead for integrative frameworks that capture the diversity and plasticity of human cognition. We conclude with a roadmap for future research, emphasizing longitudinal, cross-linguistic, and translational approaches. This work calls for a rethinking of language's role—not as a passive reflection of mind, but as its formative force.

KEYWORDS

language, cognition, cross-culture, cognitive development, language cognition

1 Introduction

Language is one of the most distinctive achievements of human cognition. Beyond serving as a medium for social communication, it functions as a central tool in shaping thought, organizing experience, and regulating behavior (Lupyan, 2016). Traditional views in cognitive science have often framed language as a modular system subject to top-down regulation by domain-general Executive Function (EF) (Shokrkon and Nicoladis, 2022). However, a growing body of neuroscientific, behavioral, and developmental research has challenged this unidirectional control model. Instead, evidence increasingly supports a bidirectional and deeply interwoven relationship between language and cognitive functions (Perszyk and Waxman, 2018; Goldin-meadow et al., 2014).

Early linguistic experience—encompassing the quantity, quality, and diversity of input (e.g., monolingual, bilingual, and multilingual environments)—exerts profound and lasting influence on the trajectory of cognitive development, including executive function, working memory, and social cognition (Shokrkon and Nicoladis, 2022; Garfield et al., 2001). This implies that language is not merely a communicative tool but a core architect of cognitive development. Understanding how language evolves from a passive medium

into an active constructor of cognition is essential for uncovering the developmental foundations of the human mind.

In this review, we aim to synthesize recent progress on the interaction between language and cognition during childhood, moving beyond modular frameworks toward a dynamic, integrative perspective. Specifically, we seek to:

1. Examine the co-developmental relationship between language and executive function;
2. Use the “bilingual advantage” debate to highlight methodological and moderating factors in the study of language-cognition interactions;
3. Investigate the scaffolding role of language—particularly syntax—in the development of higher-order social cognition such as theory of mind;
4. Analyze patterns of co-occurring deficits in language and cognition in populations with developmental language disorder (DLD) and autism spectrum disorder (ASD) to elucidate their deep interdependence;
5. Emphasize the importance of cross-linguistic and cross-cultural research in testing the universality of language-cognition theories.

By reviewing these key domains, we aim to provide a scientific foundation for improving educational policy (e.g., bilingual education), advancing clinical interventions for language and cognitive disorders, and deepening theoretical insights into the architecture of human cognition.

2 The Core engine: the co-development of language and executive function

The relationship between language and EF represents a foundational topic in cognitive science. EF refers to a set of higher-order regulatory processes that support goal-directed behavior, including working memory, inhibitory control, and cognitive flexibility (Friedman and Robbins, 2022; Zelazo, 2015). Traditional models have tended to conceptualize EF as a central controller that exerts top-down regulation on domain-specific systems, including language (Fedorenko, 2014). However, emerging longitudinal and neuroimaging evidence suggests a more reciprocal, co-developer relationship in which language and EF evolve interactively throughout development (Shokrkon and Nicoladis, 2022; Gooch et al., 2016).

These dynamic systems view call for a reassessment of how these domains interact across developmental time and offer a more integrated framework for understanding both typical and atypical cognitive development.

2.1 Beyond unidirectionality: toward a bidirectional developmental framework

A growing number of longitudinal studies support a bidirectional influence between language and EF. Some findings

indicate that early language abilities can predict later improvements in EF. For instance, Kuhn et al. found that the rate of growth in expressive language abilities (vocabulary size and syntactic complexity) between 15 and 36 months significantly predicted gains in EF from 36 to 60 months, as well as EF levels at age five (Kuhn et al., 2016). These results align with Vygotskian theory, which posits that early language serves as a symbolic tool for the development of self-regulation and executive control (Valloton and Ayoub, 2011).

Conversely, there is also evidence that early advances in EF facilitate subsequent language growth (Ekerim and Selcuk, 2018). Improvements in inhibitory control during toddlerhood, (e.g.), have been linked to greater vocabulary gains, and performance on interference suppression tasks (e.g., the Stroop task) at age five has been found to predict syntactic proficiency (Bub et al., 2006; Ibbotson and Kearvell-white, 2015). These findings support a developmental cascade model, in which the two systems grow in parallel and mutually reinforce one another over time.

Within this cascade framework, progress in one domain creates scaffolding for advancement in the other. For example, rapid vocabulary acquisition may provide children with representational tools for performing complex rule-based tasks, while enhanced inhibitory control may free cognitive resources needed for mastering hierarchical syntactic structures. Such intertwined developmental trajectories imply that interventions should be dynamically aligned with children’s developmental stages. When language and EF are reciprocally reinforcing, isolated intervention in one domain may be insufficient; instead, adaptive strategies that jointly target both systems may yield more robust outcomes.

2.2 Shared neural foundations: overlapping networks and co-activation

The bidirectional relationship between language and EF, evident at the behavioral level, is also mirrored in the brain (Shokrkon and Nicoladis, 2022; Lehtonen et al., 2023). Rather than operating as distinct, modular systems, language and EF rely on partially overlapping and tightly interconnected neural networks. This perspective departs from classical modular models of language, such as the Broca–Wernicke framework (Nasios et al., 2019), and aligns with modern network-based views of brain organization (Chang et al., 2015).

One particularly relevant system is the multiple-demand (MD) network, which is recruited across a wide range of cognitively demanding tasks (Cai et al., 2024). Neuroimaging studies have shown that early balanced bilinguals exhibit stronger activation of the MD network during executive processing tasks compared to monolinguals (Malik-moraleda et al., 2021; Grundy et al., 2017). For example, a large-scale fMRI study revealed that bilinguals demonstrated significantly greater MD system activation during spatial working memory tasks, directly linking bilingual experience to the neural architecture of cognitive control (Malik-moraleda et al., 2021). These findings provide neurobiological evidence for how language experience can shape executive systems.

Moreover, during complex linguistic tasks—such as syntactic ambiguity resolution—or language tasks that require cognitive

control, traditional language-related regions (e.g., the left inferior frontal gyrus) and prefrontal executive regions are co-activated (Ye and Zhou, 2009; Coderre et al., 2016). This overlapping activation pattern suggests that language processing is not supported by a fixed set of cortical areas but rather by dynamically assembled functional alliances. For relatively simple and automatic language tasks, activation is confined largely to perisylvian regions; however, for more complex or control-demanding language operations—such as code-switching or interpreting puns—the brain recruits broader resources from the MD system.

Lesion-symptom mapping studies in individuals with aphasia also support this non-modular account. Damage to Broca's area is frequently associated not only with impaired language production but also with deficits in nonverbal cognitive control (Fedorenko et al., 2012). This indicates that Broca's area contains subregions with distinct functions: some are highly specialized for language, while others participate in domain-general control processes (Bulut, 2023). Studies have visually demonstrated this distinction, showing that language-selective and domain-general regions are juxtaposed within Broca's area (Fedorenko et al., 2012; Fedorenko and Blank, 2020). The former are selectively activated during sentence comprehension, whereas the latter are engaged across various cognitive tasks (Fedorenko et al., 2012).

Thus, from a neural standpoint, language processing and executive control are not cleanly separable; in many contexts, the controlled use of language is itself an act of executive function (Ye and Zhou, 2009). This also explains why lesions in traditionally defined “language areas” sometimes lead to EF impairments. For instance, patients with focal damage to Broca's area often show not only expressive aphasia but also marked difficulties in task switching, inhibitory control, and verbal working memory—key components of executive functioning (Irani et al., 2025; Fedorenko et al., 2012). These findings suggest that parts of Broca's area contribute to domain-general control processes that extend beyond language, reinforcing the view that language and executive control rely on shared neural substrates rather than distinct modular systems. From an evolutionary perspective, language-related brain regions such as Broca's and Wernicke's areas exhibit distinct anatomical specializations in humans, including broader and more widely spaced cortical minicolumns (Rilling, 2008; Buxhoeveden et al., 2001), as well as increased horizontal connectivity and cytoarchitectonic complexity (Zimmermann et al., 2024; Galakhova et al., 2022). These structural features suggest that the evolution of language was accompanied by a reorganization of existing domain-general neural systems, supporting the view that language functions emerge from, and are integrated with, broader cognitive control architecture.

2.3 Working memory: a central interface between language and cognition

Working Memory (WM), particularly verbal working memory, represents a critical interface through which language interacts with other cognitive processes (Baddeley, 2003). It is important to recognize that WM capacity and efficiency are not merely

properties of fixed “cognitive hardware” but are significantly shaped by an individual's linguistic knowledge (Miyake and Friedman, 2013).

Empirical studies show that verbal WM performance leverages long-term representations of phonological, lexical, and semantic patterns (Schwering and Macdonald, 2020). Participants consistently recall real words more accurately than pseudo words and perform better with pseudo words that resemble real words in sound structure than with those that do not (Messer et al., 2010). In other words, working memory is not a passive buffer; it actively recruits stored language knowledge to enhance performance (Messer et al., 2010).

This phenomenon has significant implications for bilingual assessments. Bilinguals often have smaller vocabularies in each language compared to monolinguals, which can disadvantage them in verbal memory tasks administered in a single language (Bialystok, 2009; Kaushanskaya et al., 2011). However, this does not indicate weaker cognitive capacity—rather, it reflects unfair demands on their lexical resources (Bialystok, 2009). Indeed, studies have found that bilinguals may underperform monolinguals in rote repetition tasks but perform equivalently in tasks that require cognitive control (Bialystok, 2009). This suggests that lower performance on traditional verbal tests may reflect vocabulary differences rather than true disparities in memory ability.

Research on developmental language disorder (DLD) further highlights the role of verbal WM. One of the hallmark cognitive impairments in DLD is a persistent and severe deficit in verbal working memory (Niu et al., 2024). Children with DLD typically underperform on tasks involving digit span, word span, and sentence repetition (Larson and Ellis Weismer, 2022). This WM deficit is regarded as a defining cognitive feature of DLD. In contrast, findings regarding nonverbal EF in DLD are more heterogeneous: while some studies report difficulties in attention control and task switching, others do not, and considerable variability exists within this population (Montgomery, 2002; Contiramsden et al., 2015). This imbalance has led to a consistent pattern: verbal WM is reliably impaired, whereas nonverbal EF is variable and less consistently affected.

These findings challenge single-deficit models of DLD. If EF deficits were the root cause, nearly all children with DLD should exhibit such impairments, yet many demonstrate typical performance on nonverbal WM and executive tasks. A more plausible account is that language and EF are mutually influential developmental systems, and their dysregulation jointly contributes to the cognitive phenotype observed in DLD (Niu et al., 2024).

Taken together, these lines of evidence support the notion that language and EF are embedded in a recursive feedback loop from early development. Language experience shapes the architecture of cognitive control networks. Conversely, executive functions—such as attentional control, cognitive flexibility, and inhibitory regulation—can both support and limit language development, depending on their efficiency and maturation (Diamond, 2013; Jurado and Rosselli, 2007). For example, sustained attention and working memory capacity facilitate vocabulary acquisition and sentence parsing, while deficits in inhibitory control may impede the suppression of irrelevant linguistic information or hinder discourse planning (Novick et al., 2010; Korko and Williams,

2017). In this way, executive capacities provide the scaffolding for language to develop efficiently—but when underdeveloped, they may also impose constraints. Rather than running on parallel tracks, language and executive function form an interwoven, co-regulated system. As we emphasize throughout this review, language is not merely involved in cognition—it helps construct it.

3 Deconstructing the “bilingual executive function advantage” debate

The bilingual advantage hypothesis proposes that managing two languages on a daily basis strengthens domain-general executive function (EF) through continuous practice in control, inhibition, and selection (Paap et al., 2025). Over the past two decades, this idea has evolved into one of the most intensely debated and methodologically scrutinized topics in cognitive science (Malik-moraleda et al., 2021).

Several empirical studies have reported enhanced performance among bilinguals in tasks involving attentional control and conflict resolution, such as the Simon task and the Attention Network Test (Bialystok et al., 2004; Costa et al., 2008), and several meta-analyses have concluded moderate advantages for bilinguals under certain conditions (Adesope et al., 2010; Gunnerud et al., 2020). Proponents argue that bilingual language use enhances domain-general EF by repeatedly engaging mechanisms for monitoring, switching, and inhibition (Lehtonen et al., 2023; Wu et al., 2021; Bialystok, 2011).

However, a growing body of evidence has challenged the robustness and generalizability of these effects. Critics emphasize inconsistent replication, task impurity, and potential confounds such as socioeconomic status, immigration status, and publication bias (Paap and Greenberg, 2013; Lehtonen et al., 2018; De bruin et al., 2015). Other meta-analyses have found negligible or null effects across diverse cognitive tasks (Donnelly et al., 2019).

These conflicting findings have shifted the focus of the field from a binary question of presence or absence of a bilingual advantage to a more nuanced consideration of boundary conditions, measurement reliability, and individual difference factors (Grundy and Timmer, 2017; Valian, 2015). As such, the debate increasingly reflects a deeper methodological and theoretical divide within developmental cognitive neuroscience.

3.1 The rise and re-evaluation of the bilingual advantage hypothesis

The bilingual advantage hypothesis gained early momentum with pioneering studies demonstrating that managing two languages may bolster domain-general EF—such as inhibitory control and task switching—through ongoing cognitive practice (Bialystok et al., 2004; Costa et al., 2008). Supportive findings extended across development: bilingual children outperformed monolingual peers in conflict-monitoring tasks (Costa et al., 2008; Novitskiy et al., 2019), while lifelong bilingualism appeared to delay the onset of age-related cognitive decline in older adults (Bialystok, 2011; Bialystok et al., 2016).

As the literature grew, meta-analytic efforts began to probe the robustness of the advantage. Some syntheses reported small to moderate bilingual benefits under certain conditions—for instance, Adesope et al. (Adesope et al., 2010) observed modest overall EF gains, and De bruin et al. (2021) found a global EF advantage. However, accumulating evidence also challenged the generalizability of the advantage. Critics identified methodological shortcomings, including publication bias and confounding variables like socioeconomic status and immigration history (De bruin et al., 2015; Paap and Greenberg, 2013; Lehtonen et al., 2018). Other meta-analyses reported negligible or null effects across age groups (Donnelly et al., 2019; Gunnerud et al., 2020).

Taken together, the field has evolved from a simple “yes/no” framing of the bilingual advantage to recognizing a context-dependent phenomenon. Now, advantage appears contingent on factors such as participant age, task type, proficiency level, and study quality (Grundy and Timmer, 2017; Ware et al., 2020). Consequently, the debate has shifted toward addressing methodological rigor and boundary conditions, rather than proving or disproving a universal effect. And the bilingual advantage debate has evolved into a nuanced scientific inquiry where both supportive and contradictory findings have advanced our understanding of the interplay between bilingual experience and executive function. It emphasizes that bilingualism may support executive control under specific circumstances—particularly in tasks with high conflict demands or in populations with greater cognitive plasticity (e.g., children, older adults)—but it should not be assumed as a general cognitive enhancer. Future research must refine measurement paradigms, control for confounds, and delineate when and for whom bilingual experience yields EF benefits.

3.2 Identifying moderators: the “when, who, and what” of bilingual advantage

A growing consensus holds that any putative bilingual advantage is contingent rather than universal, and it varies along the when—who—what dimensions. When bilingual experience relates to EF appears to depend on developmental stage and everyday interactional contexts. Reported group differences are more frequent in older adults than in young adults—consistent with cognitive-reserve or adaptation accounts—whereas findings in children and young adults are typically small or null unless tasks impose substantial conflict-monitoring or switching demands (Bialystok et al., 2012, 2024; Ware et al., 2020). Moreover, interactional context matters: more integrated dual-language use and higher switching frequency are associated with reduced switch costs and modest benefits on control-demanding paradigms, in line with the Adaptive Control Hypothesis and usage-based perspectives (Green and Abutalebi, 2013; Prior and Gollan, 2011; Hartanto and Yang, 2016).

Who appears to benefit reflects individual profiles, but not in a threshold-like way. Higher proficiency across both languages and more frequent dual-language use have been associated with slightly larger effects in some studies; however, converging work suggests that what better predicts outcomes is the intensity and diversity

of control demands—indexed by integrated usage patterns, habitual language switching, and language-use entropy—rather than proficiency alone (Iniesta et al., 2025; Pliatsikas, 2020; Van den berg et al., 2022). Background factors such as socioeconomic status and immigration history can attenuate, inflate, or even reverse apparent group differences (Morton and Harper, 2007; Calvo and Bialystok, 2014). Importantly, within a bidirectional framework, monolingual language experience (e.g., lexical growth, syntactic proficiency) also correlates with EF components (e.g., interference control, updating, controlled parsing), supporting a general experience-dependent adaptation account whereby language use—mono- or bilingual—shapes attention/control systems through practice (Novick et al., 2010; De abreu et al., 2011).

What is observed depends on EF components, task features, and measurement quality. Moderator analyses and bias-corrected syntheses indicate that small, relatively stable differences sometimes appear for interference control (e.g., Flanker/Stroop), whereas purported effects on response inhibition (e.g., Go/No-Go) and working memory/updating are less reliable and often vanish after correcting for publication bias (Paap and Greenberg, 2013; Donnelly et al., 2019; Paap et al., 2020; Ware et al., 2020; Lehtonen et al., 2018). Some overviews further suggest that group differences may reflect processing-speed differences rather than core EF *per se*, and concerns about task impurity, psychometric reliability, and analytic flexibility constrain generalization (Lehtonen et al., 2018; Hedge et al., 2018). Taken together, where bilingual—monolingual differences are detected, they appear more likely under specific developmental and usage contexts, for particular EF components, and with appropriately sensitive and reliable tasks—patterns that align more closely with experience-dependent adaptation to specific control demands than with a broad, domain-general enhancement.

In sum, current evidence suggests that potential bilingual advantages are contingent on a complex interplay of factors, including developmental stage, task type, executive function domain, proficiency level, and patterns of language use. Such advantages seem to emerge most consistently when multiple facilitating conditions converge, rather than as a universal by-product of bilingualism.

Yet, even when carefully considering these moderating variables, the observed variability in outcomes remains striking. This raises a critical question: to what extent do the mixed results in the literature reflect true heterogeneity in the underlying cognitive effects, and to what extent might they be shaped by the way research is disseminated and reported? Addressing this question requires examining structural influences on the evidence base—most notably, the role of publication bias.

3.3 The shadow of publication bias

An often under examined factor in the bilingual advantage debate is publication bias—the tendency for studies reporting significant effects to be more likely published than those reporting null results. Several studies have documented this skew in the

bilingual advantage literature (Lehtonen et al., 2018; De bruin et al., 2015; Donnelly et al., 2019). For example, De bruin et al. (2015) surveyed conference abstracts and found that studies with null results were substantially less likely to appear in peer-reviewed journals.

While this pattern suggests that the published record may represent only the “visible tip” of a larger body of mixed or null findings, it is important to acknowledge recent structural shifts in dissemination practices that actively counteract this bias. The growth of preprint servers and open data repositories has enabled the rapid sharing of null and replication results, increasing transparency and reducing the dependence on traditional journal gatekeeping (Tracz and Lawrence, 2016; Smart, 2022).

Consequently, any assessment of the bilingual advantage hypothesis must consider publication bias not merely as a methodological artifact but as a shaping force in the scientific narrative—one that is itself evolving. The critical question may not be whether bilingualism confers cognitive benefits in principle, but how our research ecosystem, including incentives, publication practices, and open-science initiatives, determines which outcomes are amplified, which are muted, and how the balance between them changes over time.

3.4 From “advantage” to “adaptation”: a neurocognitive perspective

In light of the variability and ambiguity in behavioral results, many scholars have moved beyond the binary “advantage or no advantage” question to adopt a more mechanistic view: bilingual experience as a driver of neurocognitive adaptation. This perspective aligns with the well-established notion of language as a potent neuroplasticity-inducing factor—an idea long recognized in cognitive neuroscience (Baum and Titone, 2014; Bialystok et al., 2014; Del maschio et al., 2018; Korenar et al., 2023). “Cognitive architect” is used as a metaphor to describe how bilingualism, through sustained engagement with two linguistic systems, continuously remodels neural circuitry to meet the unique demands of bilingual communication.

Neuroimaging studies show that bilinguals often display structural and functional differences in regions of the MD network, with greater recruitment during executive EF tasks compared to monolinguals (Pliatsikas and Luk, 2016). Electrophysiological evidence similarly reveals distinct event-related potential (ERP) patterns during conflict monitoring and resolution (Botezatu et al., 2021), pointing to changes in neural efficiency or processing strategies. Crucially, such adaptations are robust and replicable even when behavioral differences are small or absent, suggesting that neural reconfiguration may precede, or occur independently of, overt performance changes.

This neurocognitive adaptation framework helps to resolve a long-standing paradox: the frequent observation of neural differences in the absence of strong behavioral effects. The explanation lies in the reorganization of cognitive control systems to handle challenges such as language switching, interference suppression, and cross-linguistic monitoring. These adaptations

may optimize brain function for bilingual contexts without necessarily producing generalized EF advantages on laboratory tasks (Bialystok, 2017).

Thus, the central question is not whether bilinguals are universally “better,” but how bilingual experience interacts with neuroplastic mechanisms to shape the architecture of cognitive control. In this view, bilingualism represents a powerful experiential context that, much like other forms of intensive cognitive engagement, continuously shapes neural and cognitive landscapes over the lifespan.

4 How language shapes higher-order cognition

One of the hallmarks of human cognition is the ability to reason about one's own and others' “mental states—a capacity known as Theory of Mind (ToM) (Siegal and Varley, 2002). The ability to understand others' beliefs, desires, and intentions represents a key milestone in the development of social cognition (Frith and Frith, 2012). Yet, how children acquire ToM has long been debated. While some accounts emphasize innate modules or conceptual leaps, others increasingly highlight the constructive role of language in scaffolding the development of social understanding (De villiers and De villiers, 2025).

4.1 From words to worlds: the bidirectional link between language and social cognition

The development of ToM does not occur in isolation; rather, it is intricately intertwined with language acquisition (Garfield et al., 2001). From infancy onward, this relationship is bidirectional. On the one hand, early ToM skills—such as understanding others' intentions—can support language learning. For instance, infants who attend to intentional actions (e.g., purposeful gestures or pointing) are better at mapping words to referents (Tomasello et al., 2005). Research has shown that infants can differentiate between intentional and accidental actions and preferentially imitate goal-directed behaviors (Elsner, 2007). This indicates that early mentalizing abilities are already guiding vocabulary acquisition.

On the other hand, as children mature, language begins to serve as a powerful tool for navigating and reasoning about the social world. Children can already use linguistic cues to infer a speaker's knowledge state or epistemic certainty (Hübscher et al., 2019). For example, toddlers interpret “I know X” as more confident than “I think X,” suggesting that children use subtle semantic distinctions to infer others' beliefs and attitudes.

Overall, language and social cognition co-develop through a dynamic, reciprocal process. Language provides a medium for expressing and rehearsing mental state reasoning, while social cognition directs attention and motivation during language learning. The development of perspective-taking and intention understanding is thus closely tied to the growth of linguistic competence.

4.2 The syntax key: unlocking false-belief reasoning

Why do children only reliably pass standard false-belief tasks around age four (Hughes et al., 2000)? One influential account emphasizes the acquisition of a specific syntactic structure—sentential complements (De villiers, 2005). These embedded clauses allow the expression of beliefs that may diverge from reality, e.g., “Mary thinks the ball is in the box.” Mastering such constructions has been linked to the ability to represent others' mistaken beliefs, with longitudinal and training studies showing that early competence in complement clauses predicts earlier success in false-belief reasoning (De villiers and De villiers, 2025; Boeg Thomsen et al., 2021; Mo et al., 2014). This association remains robust even after controlling for vocabulary size, memory span, and general language ability (Lind and Bowler, 2009).

The prevailing interpretation is that complement syntax provides the representational format needed to mentally simulate propositions that contradict reality, thereby scaffolding mature Theory of Mind (ToM). However, alternative interpretations remain viable. As some scholars argue, the use of such syntax may function less as a causal enabler and more as a behavioral marker of underlying cognitive readiness for ToM (Baillargeon et al., 2010). Language does not necessarily create the capacity to distinguish between personal knowledge and others' misbeliefs but serves as an observable artifact—a socially available tool that makes such distinctions communicable and testable in experimental contexts.

Recognizing both perspectives helps clarify the current debate. On one hand, complement syntax may facilitate the mental representation of false beliefs; on the other, its emergence in children's speech could simply coincide with, and reveal, a pre-existing maturation of ToM-related cognitive architecture. Distinguishing between these possibilities remains an open empirical challenge, which points to the importance of integrative approaches that combine linguistic, cognitive, and neurodevelopmental evidence.

4.3 New methods for tracking neurodevelopmental trajectories

Recent advances in child-friendly neuroimaging methods—such as functional near-infrared spectroscopy (fNIRS) and high-density EEG—have opened a new window into the neural underpinnings of language and higher-order cognition in early development.

fNIRS has proven to be a sensitive and non-invasive tool for measuring cortical activation in infants and young children during exposure to various sensory and social stimuli (Quaresima and Ferrari, 2019). For example, converging evidence from developmental neuroscience shows that even in the first year of life, infants' temporal and frontal cortices respond selectively to socially relevant cues such as eye gaze, facial expressions, and infant-directed speech (Grossmann and Johnson, 2007; Grossmann, 2015). These neural responses are thought to support early social attention and may form a foundation upon which more

complex communicative skills later emerge, although the link to communicative intent *per se* requires careful, task-specific testing.

Moreover, a longitudinal fNIRS study of 4-year-olds found that resting-state functional connectivity between the prefrontal cortex and other regions predicted individual differences in executive function months later (Kerr-german et al., 2022). These findings indicate that the maturity of early brain networks plays a role in shaping later cognitive development.

Moving forward, integrating multiple neuroimaging modalities holds promise for a more comprehensive understanding of the language-cognition interface. Combining the high temporal resolution of EEG with the moderate spatial resolution of fNIRS, for example, enables researchers to simultaneously track the timing and localization of brain activity during cognitive tasks. This multimodal approach can help identify the precise moments and neural circuits involved when children deploy language to reason about others' minds.

Such methods offer exciting potential for testing causal hypotheses—such as whether learning sentential complements leads to measurable changes in neural activation during ToM tasks. By capturing the dynamic interplay between linguistic input and cognitive development, these tools will allow researchers to map how the “neural architecture of social cognition” is constructed in real time.

5 Insights from the margins: atypical development as a window into cognitive architecture

One powerful method for investigating the architecture of the cognitive system is to observe how it manifests under atypical developmental conditions. Developmental disorders—such as DLD and Autism Spectrum Disorder (ASD)—present not only clinical challenges but also serve as valuable “natural experiments.” These populations often display non-canonical patterns of interaction or dissociation between language and cognition, offering rare opportunities to reveal connections that are otherwise tightly intertwined and difficult to tease apart in typical development.

DLD is characterized by persistent difficulties in language acquisition (Montgomery, 2002); ASD is marked by deficits in social communication and restricted, repetitive behaviors (Xiao et al., 2024). Yet, both are cognitively complex conditions in which impairments in language and EF frequently co-occur. By comparing these distinct atypical profiles, we can test the limits of modular accounts and gain deeper insight into the developmental interdependence of language and cognition.

5.1 DLD: a language deficit, a cognitive deficit, or both?

Formerly known as Specific Language Impairment (SLI), the condition now referred to as Developmental Language Disorder (DLD) has undergone substantial terminological refinement in recent years. A major step in this process—a consensus study led by Bishop et al. (Bishop et al., 2017a,b) and involving an international

panel of experts—which recommended adopting the term DLD to improve diagnostic clarity and cross-disciplinary communication. The change was motivated by persistent challenges in defining the disorder solely in terms of exclusionary criteria (e.g., absence of hearing loss, intellectual disability) and by the recognition that language impairments often co-occur with other cognitive difficulties (Bishop et al., 2017b,a).

Historically, DLD was conceptualized as an isolated deficit in language (Montgomery, 2002). However, accumulating evidence reveals that DLD is rarely a “pure” language impairment—it often coexists with broader cognitive difficulties, particularly in verbal memory and processing (Montgomery, 2002; Clegg et al., 2005). Children with DLD consistently perform poorly on verbal working memory (WM) tasks across languages and ages (Jackson et al., 2021). These deficits are so robust and stable that they are often considered cognitive hallmarks of the disorder (Niu et al., 2024).

In contrast, findings on nonverbal executive functions (e.g., visuospatial WM, planning, set-shifting) in DLD are highly variable. Some children show clear EF deficits; others perform within normal ranges (Conti-ramsdén et al., 2015). This variability challenges simple causal models. If EF deficits were the primary cause of DLD, then all children with DLD should exhibit impaired EF. Yet, many do not. This suggests a more nuanced relationship in which language and EF may interact and potentially compensate for each other during development. Some children may mitigate language impairments via intact EF resources; others may not, depending on their cognitive profile (Pauls and Archibald, 2016).

Indeed, while both DLD and ASD involve disruptions in language and cognition, the nature and profile of these disruptions differ markedly. DLD is marked by pronounced verbal memory deficits, while ASD is characterized more by social cognition and broad EF challenges. These divergent profiles underscore our central argument: language and cognition are deeply interwoven, and their relationship must be understood within specific developmental contexts. For children with DLD, weakened language skills may hinder the use of language as a cognitive tool, leading to secondary difficulties in memory and learning. However, children with relatively intact EF may develop compensatory strategies. In contrast, ASD children may struggle to engage in language-mediated social learning due to underlying EF and mentalizing impairments—though in some cases, multilingual exposure appears to be beneficial.

5.2 ASD: the interplay of language, executive function, and social cognition

ASD is often described as involving a “triple burden” of language impairment, executive dysfunction, and theory of mind deficits. Importantly, these domains do not operate independently but are tightly interrelated in complex ways. Studies have found significant correlations between language abilities (e.g., vocabulary, syntax, pragmatic understanding) and EF performance (e.g., working memory, cognitive flexibility, inhibitory control) in children with ASD (Friedman and Sterling, 2019). The direction of causality is unclear—deficits in one domain may cascade

into the others, or shared underlying factors (e.g., atypical brain maturation) may drive co-impairments.

One particularly intriguing line of research has investigated the impact of multilingual environments on children with ASD. Surprisingly, several studies report that bilingual or multilingual children with ASD perform better than their monolingual counterparts on parent-rated measures of EF (e.g., inhibition, shifting) and perspective-taking skills (Ratto et al., 2020; Andreou et al., 2020). They also show fewer core ASD symptoms, such as reduced stereotyped behaviors.

These findings challenge longstanding assumptions. Rather than exacerbating symptoms, multilingual exposure may actually strengthen cognitive control systems and indirectly enhance social functioning. While the current data are largely correlational and based on parent reports, they suggest that language experience can modulate core cognitive features of neurodevelopmental disorders. Importantly, they counter outdated advice to limit language input in ASD children (“Don’t confuse them with two languages”), instead pointing toward the potential utility of bilingual environments as part of broader intervention strategies (Hamby and Fombonne, 2012).

5.3 Toward integrated interventions

Given the close interdependence of language and cognitive functions in both DLD and ASD, interventions targeting a single domain may be insufficient. Instead, integrative approaches that address both language and EF capacities may be more effective.

For children with DLD, language interventions should go beyond vocabulary and syntax, incorporating supports for attention and memory—such as visual scaffolds or chunking techniques to ease WM load (Montgomery et al., 2024). Similarly, social skills training for children with ASD may be more successful when paired with exercises targeting EF (e.g., turn-taking, rule-switching, response inhibition).

Neuroimaging tools are beginning to play a role in shaping individualized interventions. For example, fNIRS studies in children who stutter have identified atypical blood flow patterns in left frontal language areas, potentially serving as early biomarkers for persistent stuttering (Walsh et al., 2017). Similar approaches could be extended to children with DLD—using fNIRS or EEG to identify weak neural responses to speech sounds or syntactic anomalies. These markers could inform the design of targeted training programs and provide real-time feedback on neural changes.

As the field of educational neuroscience advances, neurobiologically informed interventions are becoming more feasible. Computational modeling, for instance, may help match language games to specific cognitive deficits, optimizing training effects across domains.

In sum, atypical development reminds us that language and cognition are two wings of the same bird. Strengthening one in isolation rarely enables flight. What we need are cross-disciplinary collaborations that unite speech-language therapy, cognitive training, and enriched language environments. Only then can we design interventions that truly harness the brain’s capacity for growth—and help every child reach their full potential.

6 Cross-cultural and cross-linguistic diversity: expanding the boundaries of language-cognition research

For decades, psychological and cognitive science research has relied disproportionately on WEIRD populations—those who are Western, Educated, Industrialized, Rich, and Democratic (Henrich et al., 2010). This sampling bias raises serious questions about the universality of our theories. Are the observed links between language and cognition truly species-wide mechanisms, or are they artifacts of specific cultural-linguistic environments? To answer this, the field must break free from its reliance on English-speaking, monolingual children and systematically investigate a broader range of linguistic and cultural contexts.

The structural features of different languages (e.g., morphological complexity, writing systems) and the cultural norms of language use can shape cognitive strategies in unique ways. For example, polysynthetic languages such as Inuktitut, which frequently use long compound words, may place different demands on memory and attention compared to analytic languages like Mandarin. Similarly, parenting styles and educational values—such as emphasis on storytelling vs. rule-following—may influence the developmental trajectory and prioritization of specific cognitive skills. We argue that building truly universal theories requires integrating cross-linguistic and cross-cultural diversity into the research agenda.

6.1 Commonalities and variability in the language–cognition link

At a macro level, the relationship between language and cognition reflects a fundamental human commonality: across all languages, the acquisition of language provides a framework and set of tools for thought (Ellis, 2019). This principle holds regardless of language type. However, its specific manifestations can vary depending on the structural features of the language and the sociocultural context in which it is used.

For instance, the morphosyntactic features of a language may shape WM demands and processing strategies (Chávez and Auza Benavides, 2017). In highly inflected languages (e.g., Turkish), where words are packed with multiple suffixes encoding nuanced meanings, sentence comprehension may rely more heavily on maintaining and manipulating morphological information—thus training different WM-attention systems. By contrast, in morphologically sparse languages (e.g., Mandarin), children may develop context-sensitive interpretive strategies earlier due to the greater reliance on word order and discourse context.

Cultural values also modulate cognitive development (Wang, 2018). Cultures that prioritize narrative skills may foster children’s abilities in memory organization and causal reasoning, whereas cultures that emphasize rule compliance and behavioral restraint may shape inhibitory control (Raver and Blair, 2016). Cross-linguistic studies have found both universality and variability in disorders like DLD. The core impairment in verbal WM is observable across languages (universality) (Niu et al., 2024), but

performance on nonverbal EF tasks can differ by language group (variability) (Montgomery, 2002).

Taken together, these findings suggest that while some cognitive effects of language are broadly shared, others are contingent on culturally and linguistically specific experiences. These contextual influences do not undermine the bidirectional relationship between language and cognition; rather, they highlight that this relationship is dynamically shaped by the interplay between universal mechanisms and local adaptations.

6.2 Language distance and cognitive control

Not all bilingual experiences are created equal—language distance, defined as the degree of structural similarity or dissimilarity between two languages, can substantially influence the cognitive demands placed on bilingual speakers (Deluca et al., 2019). From a neurocognitive perspective, managing structurally distant languages requires greater engagement of control-related networks—particularly the dorsolateral prefrontal cortex, anterior cingulate cortex, and basal ganglia—due to increased demands on language selection and conflict monitoring (Abutalebi and Green, 2016; Li et al., 2014).

Recent lifespan studies have shown that speakers of more distant language pairs often exhibit enhanced monitoring and attentional switching, especially in aging populations, potentially reflecting sustained recruitment of domain-general executive resources over decades of bilingual experience (Carthery-goulart et al., 2023; Bialystok, 2017). For example, older adult bilinguals managing typologically distant languages demonstrate superior performance in task-switching paradigms compared with those speaking closely related languages, suggesting that long-term exposure to high-conflict language environments may bolster cognitive resilience.

However, language distance interacts with other experiential variables such as code-switching frequency, degree of language mixing, and sociolinguistic context. In immersion settings with high daily switching demands, structural distance may amplify flexibility in task-set shifting, whereas in low-switching contexts, inhibitory control processes may dominate (Wu et al., 2021). Neuroimaging evidence further indicates that distant-language bilinguals show more segregated neural representations in the language network, while close-language bilinguals exhibit greater overlap, potentially influencing control strategies (Pliatsikas, 2020).

Another emerging direction is the study of multilinguals, where the acquisition of a third or fourth language creates a natural gradient of language distances. Research indicates that the proximity of a new language to one's existing languages influences not only vocabulary acquisition but also adaptive changes in working memory capacity and control mechanisms (Engel de abreu and Gathercole, 2012). This perspective extends beyond traditional binary bilingual comparisons, emphasizing the complex, multidimensional nature of language experience in shaping cognitive control.

By situating language distance within a broader neurocognitive and experiential framework, future research can move beyond

simple dichotomies and toward integrative models that capture how structural, functional, and sociocultural factors jointly modulate the bilingual cognitive profile.

6.3 Methodological challenges and opportunities in cross-cultural research

Comparing cognition across cultures is like measuring with different rulers—unless carefully calibrated, results may be misleading. Traditional standardized assessments often embed cultural biases. For instance, vocabulary tests are highly sensitive to formal education and cultural exposure, putting children from rural or marginalized backgrounds at a disadvantage.

To mitigate this, some researchers advocate for processing-dependent measures that minimize the influence of acquired knowledge (Quinn et al., 2011). Tasks such as digit span and non-word repetition are less tied to cultural content and may allow for fairer cross-group comparisons.

However, even these measures are not entirely culture-free. For example, the number of commonly used digits in a language and the phonological familiarity of nonwords can subtly influence memory performance (Grivol and Hage, 2011). This is known as the “lexicalization effect”—pseudo-words that resemble real words or high-frequency phoneme sequences are easier to recall (Martínez-tomás et al., 2025). Thus, in cross-cultural assessments, researchers must both minimize cultural loading and statistically account for any residual biases.

A promising future direction lies in developing truly cross-linguistic cognitive assessment tools. For instance, using abstract pattern-based tasks (that don't require language) or creating language-specific test versions calibrated for difficulty equivalence. Large-scale international collaborations can help normalize these tools across diverse samples, setting global benchmarks.

In summary, addressing cultural bias is not about abandoning comparisons but about making them smarter. By acknowledging the cognitive diversity of humanity while seeking shared underlying dimensions, we can build theories that are both broadly generalizable and sensitive to local realities.

7 General discussion and future directions

Language is not merely a passive tool stored within the cognitive system—it is an active, dynamic force. It serves as an “architect” of cognition, shaping the very mechanisms it helps enact throughout development. In this review, we conceptualize language as a bi directionally interactive and culturally embedded cognitive system, wherein structural and experiential factors jointly remodel the brain's architecture over the lifespan (Bialystok, 2017; Li et al., 2014). This perspective calls for a shift away from modular, unidirectional models of causality toward more integrative, dynamic, and context-sensitive frameworks in understanding the neurobiological foundations of language.

Importantly, culture is positioned here as a contextual moderator—not as an alternative causal direction—but as a

factor that shapes the magnitude and manifestation of language–cognition links. For instance, differences in communicative norms, educational practices, and societal bilingualism policies can modulate the cognitive benefits associated with language experience (Nuri, 2024).

Building on the integrative framework presented here, we propose several promising directions for future investigation:

- 1) **Developmental Cascade and Causal Testing** Longitudinal cross-lagged or experimental studies are needed to clarify the causal pathways between language and EF. Pioneering studies have already demonstrated partial causal links in early childhood, providing a methodological blueprint (Engel de abreu and Gathercole, 2012). For example, an intervention could contrast a language-enrichment group (enhanced input in vocabulary and syntax) with an EF-training group (memory, inhibition, shifting games), tracking untrained domains over time to test co-development hypotheses. Prior studies indicate that language-focused interventions in preschoolers can yield transfer effects to EF measures, particularly in working memory and cognitive flexibility (Engel de abreu and Gathercole, 2012). These findings support—but do not yet fully specify—the developmental timing and mechanisms of transfer, leading to larger-scale, stage-specific trials.
- 2) **Neurointerventions and Network Plasticity** Noninvasive neuromodulation (e.g., tDCS, TMS) can probe causal links in the brain. For instance, stimulating or inhibiting subregions of the left inferior frontal gyrus (IFG) may reveal dual effects on language and domain-general control tasks. Evidence from noninvasive stimulation studies shows that targeting the dorsolateral prefrontal cortex (DLPFC) modulates syntactic processing efficiency and sentence comprehension accuracy, supporting its causal role in integrating language and executive functions (Hertrich et al., 2021; Hussey et al., 2015). Combining these protocols with functional neuroimaging and machine learning could track how intensive bilingual or syntactic training reorganizes brain networks, detecting shifts in the topology of the MD and language networks.
- 3) **Language as a Pillar of High-Level Cognition** To validate the “syntax supports ToM” hypothesis, neurodevelopmental studies should integrate EEG and eye-tracking to examine neural responses when children process embedded mental-state sentences. Foundational work has illuminated how syntactic complexity interfaces with neural systems for higher-order cognition, including ToM and discourse processing (Bourguignon, 2023). Further, studies in sign language users and minimally verbal children with ASD can test whether alternative modalities (e.g., signed complement clauses, visual-symbol systems) facilitate ToM acquisition, informing whether syntax is necessary or merely sufficient for representing others’ minds.
- 4) **Integrative Cross-Linguistic Models** Large-scale international collaborations should examine cognition under comparable paradigms across typologically and culturally diverse languages (e.g., English, Mandarin, Turkish, Spanish). By incorporating multilevel modeling, we can partition

variance due to linguistic structure, cultural practices, and socioeconomic status, clarifying universal versus culture-specific patterns. This approach also directly operationalizes our theoretical stance on culture as a moderator, enabling empirical tests of how sociocultural context shapes the strength of bidirectional language–cognition links.

- 5) **Real-World Applications in Education and Clinical Practice** Theory must inform practice. “Bilingual–EF fusion curricula” could integrate language learning with EF-enhancing tasks in schools, while clinical programs for DLD could combine speech–language therapy with EF game-based interventions. Neuroimaging could track both behavioral and neural changes, testing whether integrated programs outperform traditional methods.

In sum, we are at the cusp of a more profound understanding of the language–mind relationship. We acknowledge the rich theoretical history and neuroscientific work in this domain and view our synthesis as a bridge between these traditions and emerging data-driven models. Through interdisciplinary collaboration, cross-cultural engagement, and rigorous empirical testing, we can move closer to a comprehensive picture of how this “architect” builds the magnificent edifice of the human mind.

Author contributions

LP: Conceptualization, Investigation, Writing – original draft, Writing – review & editing. SK: Investigation, Supervision, Validation, Writing – review & editing. NX: Methodology, Project administration, Supervision, Validation, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Abutalebi, J., and Green, D. W. (2016). Neuroimaging of language control in bilinguals: neural adaptation and reserve. *Biling. Lang. Cogn.* 19, 689–698. doi: 10.1017/S1366728916000225
- Adesope, O. O., Lavin, T., Thompson, T., and Ungerleider, C. (2010). A systematic review and meta-analysis of the cognitive correlates of bilingualism. *Rev. Educ. Res.* 80, 207–245. doi: 10.3102/0034654310368803
- Andreou, M., Tsimpli, I. M., Durrleman, S., and Peristeri, E. (2020). Theory of mind, executive functions, and syntax in bilingual children with autism spectrum disorder. *Languages* 5:67. doi: 10.3390/languages5040067
- Baddeley, A. (2003). Working memory and language: an overview. *J. Commun. Disord.* 36, 189–208. doi: 10.1016/S0021-9924(03)00019-4
- Baillargeon, R., Scott, R. M., and He, Z. (2010). False-belief understanding in infants. *Trends Cogn. Sci.* 14, 110–118. doi: 10.1016/j.tics.2009.12.006
- Baum, S., and Titone, D. (2014). Moving toward a neuroplasticity view of bilingualism, executive control, and aging. *Appl. Psycholing.* 35, 857–894. doi: 10.1017/S0142716414000174
- Bialystok, E. (2009). Bilingualism: the good, the bad, and the indifferent. *Biling. Lang. Cogn.* 12, 3–11. doi: 10.1017/S1366728908003477
- Bialystok, E. (2011). Reshaping the mind: the benefits of bilingualism. *Can. J. Exp. Psychol.* 65:229. doi: 10.1037/a0025406
- Bialystok, E. (2017). The bilingual adaptation: How minds accommodate experience. *Psychol. Bull.* 143:233. doi: 10.1037/bul0000099
- Bialystok, E., Abutalebi, J., Bak, T. H., Burke, D. M., and Kroll, J. F. (2016). Aging in two languages: implications for public health. *Ageing Res. Rev.* 27, 56–60. doi: 10.1016/j.arr.2016.03.003
- Bialystok, E., Craik, F. I., Klein, R., and Viswanathan, M. (2004). Bilingualism, aging, and cognitive control: evidence from the Simon task. *Psychol. Aging* 19, 290–303. doi: 10.1037/0882-7974.19.2.290
- Bialystok, E., Craik, F. I., and Luk, G. (2012). Bilingualism: consequences for mind and brain. *Trends Cogn. Sci.* 16, 240–250. doi: 10.1016/j.tics.2012.03.001
- Bialystok, E., Craik, F. I., and Luk, G. (2024). “Bilingualism: consequences for mind and brain,” *Where Language Meets Thought*. Routledge. p. 26. doi: 10.4324/9781003412427-12
- Bialystok, E., Poarch, G., Luo, L., and Craik, F. I. (2014). Effects of bilingualism and aging on executive function and working memory. *Psychol. Aging* 29:696. doi: 10.1037/a0037254
- Bishop, D. V., Snowling, M. J., Thompson, P. A., Greenhalgh, T., Consortium, C., Adams, C., et al. (2017a). Phase 2 of CATALISE: a multinational and multidisciplinary Delphi consensus study of problems with language development: terminology. *J. Child Psychol. Psychiatry* 58, 1068–1080. doi: 10.1111/jcpp.12721
- Bishop, D. V., Snowling, M. J., Thompson, P. A., Greenhalgh, T., and Klee, T. (2017b). CATALISE: a multinational and multidisciplinary Delphi consensus study of problems with language development. Phase 2. *J. Child Psychol. Psychiatry* 58, 1068–1080. doi: 10.7287/peerj.preprints.2484v2
- Boeg Thomsen, D., Theakston, A., Kandemirci, B., and Brandt, S. (2021). Do complement clauses really support false-belief reasoning? A longitudinal study with English-speaking 2-to 3-year-olds. *Dev. Psychol.* 57:1210. doi: 10.1037/dev0001012
- Botezatu, M. R., Miller, C. A., Johnson, J., and Misra, M. (2021). Event-related potentials reveal that bilinguals are more efficient in resolving conflict than monolinguals. *Neuroreport* 32, 721–726. doi: 10.1097/WNR.00000000000001645
- Bourguignon, N. J. (2023). The emergence of language in the human mind and brain—Insights from the neurobiology of language, thought and action. *Psychol. Rev.* 130:1544. doi: 10.1037/rev0000387
- Bub, D. N., Masson, M. E., and Lalonde, C. E. (2006). Cognitive control in children: stroop interference and suppression of word reading. *Psychol. Sci.* 17, 351–357. doi: 10.1111/j.1467-9280.2006.01710.x
- Bulut, T. (2023). Domain-general and domain-specific functional networks of Broca's area underlying language processing. *Brain Behav.* 13:e3046. doi: 10.1002/brb3.3046
- Buxhoeveden, D. P., Switala, A. E., Litaker, M., Roy, E., and Casanova, M. F. (2001). Lateralization of minicolumns in human planum temporale is absent in nonhuman primate cortex. *Brain Behav. Evol.* 57, 349–358. doi: 10.1159/000047253
- Cai, W., Taghia, J., and Menon, V. (2024). A multi-demand operating system underlying diverse cognitive tasks. *Nat. Commun.* 15:2185. doi: 10.1038/s41467-024-46511-5
- Calvo, A., and Bialystok, E. (2014). Independent effects of bilingualism and socioeconomic status on language ability and executive functioning. *Cognition* 130, 278–288. doi: 10.1016/j.cognition.2013.11.015
- Carthey-goulart, M. T., Privitera, A. J., and Weekes, B. S. (2023). Does language distance modulate the contribution of bilingualism to cognitive reserve in seniors? A systematic review. *Am. J. Alzheimer's Dis. Other Dement.* 38:15333175231167223. doi: 10.1177/15333175231167223
- Chang, E. F., Raygor, K. P., and Berger, M. S. (2015). Contemporary model of language organization: an overview for neurosurgeons. *J. Neurosurg.* 122, 250–261. doi: 10.3171/2014.10.JNS132647
- Chávez, A., and Auza Benavides, A. (2017). “Working memory and morphosyntax in children with specific (primary) language impairment,” *Language Development and Disorders in Spanish-Speaking Children*. New York, NY: Springer. doi: 10.1007/978-3-319-53646-0_14
- Clegg, J., Hollis, C., Mawhood, L., and Rutter, M. (2005). Developmental language disorders—a follow-up in later adult life. cognitive, language and psychosocial outcomes. *J. Child Psychol. Psychiatry* 46, 128–149. doi: 10.1111/j.1469-7610.2004.00342.x
- Coderre, E. L., Smith, J. F., Van heuven, W. J., and Horwitz, B. (2016). The functional overlap of executive control and language processing in bilinguals. *Biling. Lang. Cogn.* 19, 471–488. doi: 10.1017/S1366728915000188
- Conti-ramsdén, G., Ullman, M. T., and Lum, J. A. (2015). The relation between receptive grammar and procedural, declarative, and working memory in specific language impairment. *Front. Psychol.* 6:1090. doi: 10.3389/fpsyg.2015.01090
- Costa, A., Hernández, M., and Sebastián-gallés, N. (2008). Bilingualism aids conflict resolution: evidence from the ANT task. *Cognition* 106, 59–86. doi: 10.1016/j.cognition.2006.12.013
- De abreu, P. M. J. E., Gathercole, S. E., and Martin, R. (2011). Disentangling the relationship between working memory and language: the roles of short-term storage and cognitive control. *Learn. Individ. Differ.* 21, 569–574. doi: 10.1016/j.lindif.2011.06.002
- De bruin, A., Dick, A. S., and Carreiras, M. (2021). Clear theories are needed to interpret differences: perspectives on the bilingual advantage debate. *Neurobiol. Lang.* 2, 433–451. doi: 10.1162/nol_a_00038
- De bruin, A., Treccani, B., and Della sala, S. (2015). Cognitive advantage in bilingualism: an example of publication bias? *Psychol. Sci.* 26, 99–107. doi: 10.1177/0956797614557866
- De villiers, J. G. (2005). “Can language acquisition give children a point of view? *Why Language Matters for Theory of Mind* eds. J. W. Astington and J. A. Baird (Oxford, U.K: Oxford University Press), 186–219. doi: 10.1093/acprof:oso/9780195159912.003.0010
- De villiers, J. G., and De villiers, P. (2025). The contributions of language and inhibitory control to false belief reasoning over time. *Front Psychol.* 15:1455941. doi: 10.3389/fpsyg.2024.1455941
- Del maschio, N., Sulpizio, S., Gallo, F., Fedeli, D., Weekes, B. S., and Abutalebi, J. (2018). Neuroplasticity across the lifespan and aging effects in bilinguals and monolinguals. *Brain Cogn.* 125, 118–126. doi: 10.1016/j.bandc.2018.06.007
- Deluca, V., Rothman, J., Bialystok, E., and Pliatsikas, C. (2019). Redefining bilingualism as a spectrum of experiences that differentially affects brain structure and function. *Proc. Nat. Acad. Sci.* 116, 7565–7574. doi: 10.1073/pnas.1811513116
- Diamond, A. (2013). Executive functions. *Ann. Rev. Psychol.* 64, 135–168. doi: 10.1146/annurev-psych-113011-143750
- Donnelly, S., Brooks, P. J., and Homer, B. D. (2019). Is there a bilingual advantage on interference-control tasks? A multiverse meta-analysis of global reaction time and interference cost. *Psychon. Bull. Rev.* 26, 1122–1147. doi: 10.3758/s13423-019-01567-z

- Ekerim, M., and Selcuk, B. (2018). Longitudinal predictors of vocabulary knowledge in Turkish children: the role of maternal warmth, inductive reasoning, and children's inhibitory control. *Early Educ. Dev.* 29, 324–341. doi: 10.1080/10409289.2017.1407607
- Ellis, N. C. (2019). Essentials of a theory of language cognition. *Mod. Lang. J.* 103, 39–60. doi: 10.1111/modl.12532
- Elsner, B. (2007). Infants' imitation of goal-directed actions: the role of movements and action effects. *Acta psychol.* 124, 44–59. doi: 10.1016/j.actpsy.2006.09.006
- Engel de abreu, P. M., and Gathercole, S. E. (2012). Executive and phonological processes in second-language acquisition. *J. Educ. Psychol.* 104:974. doi: 10.1037/a0028390
- Fedorenko, E. (2014). The role of domain-general cognitive control in language comprehension. *Front. Psychol.* 5:335. doi: 10.3389/fpsyg.2014.00335
- Fedorenko, E., and Blank, I. A. (2020). Broca's area is not a natural kind. *Trends Cogn. Sci.* 24, 270–284. doi: 10.1016/j.tics.2020.01.001
- Fedorenko, E., Duncan, J., and Kanwisher, N. (2012). Language-selective and domain-general regions lie side by side within Broca's area. *Curr. Biol.* 22, 2059–2062. doi: 10.1016/j.cub.2012.09.011
- Friedman, L., and Sterling, A. (2019). A review of language, executive function, and intervention in autism spectrum disorder. *Semin. Speech Lang.* 40, 291–304. doi: 10.1055/s-0039-1692964
- Friedman, N. P., and Robbins, T. W. (2022). The role of prefrontal cortex in cognitive control and executive function. *Neuropsychopharmacology* 47, 72–89. doi: 10.1038/s41386-021-01132-0
- Frith, C. D., and Frith, U. (2012). Mechanisms of social cognition. *Ann Rev. Psychol.* 63, 287–313. doi: 10.1146/annurev-psych-120710-100449
- Galakhova, A. A., Hunt, S., Wilbers, R., Heyer, D. B., De kock, C. P., Mansvelder, H. D., et al. (2022). Evolution of cortical neurons supporting human cognition. *Trends Cogn. Sci.* 26, 909–922. doi: 10.1016/j.tics.2022.08.012
- Garfield, J. L., Peterson, C. C., and Perry, T. (2001). Social cognition, language acquisition and the development of the theory of mind. *Mind Lang.* 16, 494–541. doi: 10.1111/1468-0017.00180
- Goldin-meadow, S., Levine, S. C., Hedges, L. V., Huttenlocher, J., Raudenbush, S. W., and Small, S. L. (2014). New evidence about language and cognitive development based on a longitudinal study: hypotheses for intervention. *Am. Psychol.* 69:588. doi: 10.1037/a0036886
- Gooch, D., Thompson, P., Nash, H. M., Snowling, M. J., and Hulme, C. (2016). The development of executive function and language skills in the early school years. *J. Child Psychol. Psychiatry* 57, 180–187. doi: 10.1111/jcpp.12458
- Green, D. W., and Abutalebi, J. (2013). Language control in bilinguals: the adaptive control hypothesis. *J. Cogn. Psychol.* 25, 515–530. doi: 10.1080/20445911.2013.796377
- Grivol, M. A., and Hage, S. R. D. V. (2011). Phonological working memory: a comparative study between different age groups. *J. Soc. Bras. Fonoaudiol.* 23, 245–251. doi: 10.1590/S2179-64912011000300010
- Grossmann, T. (2015). The development of social brain functions in infancy. *Psychol. Bull.* 141:1266. doi: 10.1037/bul0000002
- Grossmann, T., and Johnson, M. H. (2007). The development of the social brain in human infancy. *Eur. J. Neurosci.* 25, 909–919. doi: 10.1111/j.1460-9568.2007.05379.x
- Grundy, J. G., Anderson, J. A., and Bialystok, E. (2017). Neural correlates of cognitive processing in monolinguals and bilinguals. *Ann. New York Acad. Sci.* 1396, 183–201. doi: 10.1111/nyas.13333
- Grundy, J. G., and Timmer, K. (2017). Bilingualism and working memory capacity: a comprehensive meta-analysis. *Second Lang. Res.* 33, 325–340. doi: 10.1177/0267658316678286
- Gunnerud, H. L., Ten braak, D., Reikerås, E. K. L., Donolato, E., and Melby-lervåg, M. (2020). Is bilingualism related to a cognitive advantage in children? A systematic review and meta-analysis. *Psychol. Bull.* 146:1059. doi: 10.1037/bul0000301
- Hamblly, C., and Fombonne, E. (2012). The impact of bilingual environments on language development in children with autism spectrum disorders. *J. Autism Dev. Disord.* 42, 1342–1352. doi: 10.1007/s10803-011-1365-z
- Hartanto, A., and Yang, H. (2016). Disparate bilingual experiences modulate task-switching advantages: a diffusion-model analysis of the effects of interactional context on switch costs. *Cognition* 150, 10–19. doi: 10.1016/j.cognition.2016.01.016
- Hedge, C., Powell, G., and Sumner, P. (2018). The reliability paradox: why robust cognitive tasks do not produce reliable individual differences. *Behav. Res. Methods* 50, 1166–1186. doi: 10.3758/s13428-017-0935-1
- Henrich, J., Heine, S. J., and Norenzayan, A. (2010). Beyond WEIRD: towards a broad-based behavioral science. *Behav. Brain Sci.* 33:111. doi: 10.1017/S0140525X10000725
- Hertrich, I., Dietrich, S., Blum, C., and Ackermann, H. (2021). The role of the dorsolateral prefrontal cortex for speech and language processing. *Front. Hum. Neurosci.* 15:645209. doi: 10.3389/fnhum.2021.645209
- Hübscher, I., Vincze, L., and Prieto, P. (2019). Children's signaling of their uncertain knowledge state: prosody, face, and body cues come first. *Lang. Learn. Dev.* 15, 366–389. doi: 10.1080/15475441.2019.1645669
- Hughes, C., Adlam, A., Happé, F., Jackson, J., Taylor, A., and Caspi, A. (2000). Good test-retest reliability for standard and advanced false-belief tasks across a wide range of abilities. *J. Child Psychol. Psychiatry* 41, 483–490. doi: 10.1111/1469-7610.00633
- Hussey, E. K., Ward, N., Christianson, K., and Kramer, A. F. (2015). Language and memory improvements following tDCS of left lateral prefrontal cortex. *PLoS ONE* 10:e0141417. doi: 10.1371/journal.pone.0141417
- Ibbotson, P., and Kearvell-white, J. (2015). Inhibitory control predicts grammatical ability. *PLoS ONE* 10:e0145030. doi: 10.1371/journal.pone.0145030
- Iniesta, A., Yang, M., Beatty-martinez, A. L., Itzhak, I., Gullifer, J. W., and Titone, D. (2025). Leveraging social network data to ground multilingual background measures: the case of general and socially based language entropy. *Can. J. Exp. Psychol.* 79, 15–27. doi: 10.1037/cep0000352
- Irani, D., Purmohammad, M., and Abbasian, S. (2025). A comparative analysis of executive functions in Broca's aphasia and healthy adults. *Appl. Neuropsychol. Adult* 1–10. doi: 10.1080/23279095.2025.2515127
- Jackson, E., Leitão, S., Claessen, M., and Boyes, M. (2021). Word learning and verbal working memory in children with developmental language disorder. *Autism Dev. Lang. Impair.* 6:23969415211004109. doi: 10.1177/23969415211004109
- Jurado, M. B., and Rosselli, M. (2007). The elusive nature of executive functions: a review of our current understanding. *Neuropsychol. Rev.* 17, 213–233. doi: 10.1007/s11065-007-9040-z
- Kaushanskaya, M., Blumenfeld, H. K., and Marian, V. (2011). The relationship between vocabulary and short-term memory measures in monolingual and bilingual speakers. *Int. J. Biling.* 15, 408–425. doi: 10.1177/1367006911403201
- Kerr-german, A., Namuth, A., Santosa, H., Buss, A. T., and White, S. (2022). To snack or not to snack: using fNIRS to link inhibitory control to functional connectivity in the toddler brain. *Dev. Sci.* 25:e13229. doi: 10.1111/desc.13229
- Korenar, M., Treffers-daller, J., and Pliatsikas, C. (2023). Dynamic effects of bilingualism on brain structure map onto general principles of experience-based neuroplasticity. *Sci. Rep.* 13:3428. doi: 10.1038/s41598-023-30326-3
- Korko, M., and Williams, S. A. (2017). Inhibitory control and the speech patterns of second language users. *Br. J. Psychol.* 108, 43–72. doi: 10.1111/bjop.12176
- Kuhn, L. J., Willoughby, M. T., Vernon-feagans, L., Blair, C. B., and Investigators, F. L. P. K. (2016). The contribution of children's time-specific and longitudinal expressive language skills on developmental trajectories of executive function. *J. Exp. Child Psychol.* 148, 20–34. doi: 10.1016/j.jecp.2016.03.008
- Larson, C., and Ellis weismer, S. (2022). Working memory performance in children with developmental language disorder: the role of domain. *J. Speech Lang. Hear. Res.* 65, 1906–1920. doi: 10.1044/2022_JSLHR-21-00420
- Lehtonen, M., Fyndanis, V., and Jylkkä, J. (2023). The relationship between bilingual language use and executive functions. *Nat. Rev. Psychol.* 2, 360–373. doi: 10.1038/s44159-023-00178-9
- Lehtonen, M., Soveri, A., Laine, A., Järvenpää, J., De bruin, A., and ANTfolk, J. (2018). Is bilingualism associated with enhanced executive functioning in adults? A meta-analytic review. *Psychol. Bull.* 144:394. doi: 10.1037/bul0000142
- Li, P., Legault, J., and Litcofsky, K. A. (2014). Neuroplasticity as a function of second language learning: Anatomical changes in the human brain. *Cortex* 58, 301–324. doi: 10.1016/j.cortex.2014.05.001
- Lind, S. E., and Bowler, D. M. (2009). Language and theory of mind in autism spectrum disorder: the relationship between complement syntax and false belief task performance. *J. Autism Dev. Disord.* 39, 929–937. doi: 10.1007/s10803-009-0702-y
- Lupyan, G. (2016). The centrality of language in human cognition. *Lang. Learn.* 66, 516–553. doi: 10.1111/lang.12155
- Malik-moraleda, S., Cucu, T., Lipkin, B., and Fedorenko, E. (2021). The domain-general multiple demand network is more active in early balanced bilinguals than monolinguals during executive processing. *Neurobiol. Lang.* 2, 647–664. doi: 10.1162/nol_a_00058
- Martínez-tomás, C., Baciero, A., Lázaro, M., and Hinojosa, J. A. (2025). What do pseudowords tell us about word processing? An overview. *Front. Lang. Sci.* 4:1504770. doi: 10.3389/flang.2025.1504770
- Messer, M. H., Leseman, P. P., Boom, J., and Mayo, A. Y. (2010). Phonotactic probability effect in nonword recall and its relationship with vocabulary in monolingual and bilingual preschoolers. *J. Exp. Child Psychol.* 105, 306–323. doi: 10.1016/j.jecp.2009.12.006
- Miyake, A., and Friedman, N. P. (2013). *Individual Differences in Second Language Proficiency: Working Memory as Language Aptitude*. Foreign language learning. New York, NY: Psychology Press.
- Mo, S., Su, Y., Sabbagh, M. A., and Jiaming, X. (2014). Sentential complements and false belief understanding in Chinese Mandarin-speaking preschoolers: a training study. *Cogn. Dev.* 29, 50–61. doi: 10.1016/j.cogdev.2013.11.001

- Montgomery, J. W. (2002). Understanding the language difficulties of children with specific language impairments. *Am. J. Speech-Lang. Pathol.* 11, 77–91. doi: 10.1044/1058-0360(2002/009)
- Montgomery, J. W., Gillam, R. B., and Plante, E. (2024). Enhancing syntactic knowledge in school-age children with developmental language disorder: the promise of syntactic priming. *Am. J. Speech Lang. Pathol.* 33, 580–597. doi: 10.1044/2023_AJSLP-23-00079
- Morton, J. B., and Harper, S. N. (2007). What did Simon say? Revisiting the bilingual advantage. *Dev. Sci.* 10, 719–726. doi: 10.1111/j.1467-7687.2007.00623.x
- Nasios, G., Dardiotis, E., and Messinis, L. (2019). From Broca and Wernicke to the neuromodulation era: insights of brain language networks for neurorehabilitation. *Behav. Neurol.* 2019:9894571. doi: 10.1155/2019/9894571
- Niu, T., Wang, S., Ma, J., Zeng, X., and Xue, R. (2024). Executive functions in children with developmental language disorder: a systematic review and meta-analysis. *Front. Neurosci.* 18:1390987. doi: 10.3389/fnins.2024.1390987
- Novick, J. M., Trueswell, J. C., and Thompson-schill, S. L. (2010). Broca's area and language processing: evidence for the cognitive control connection. *Lang. Ling. Compass* 4, 906–924. doi: 10.1111/j.1749-818X.2010.00244.x
- Novitskiy, N., Shtyrov, Y., and Myachykov, A. (2019). Conflict resolution ability in late bilinguals improves with increased second-language proficiency: ANT evidence. *Front. Psychol.* 10:2825. doi: 10.3389/fpsyg.2019.02825
- Nuri, A. (2024). Exploring bilingualism: cognitive benefits and cultural challenges. *Acta Glob. Hum. Ling.* 1, 71–81. doi: 10.69760/aghel.024053
- Paap, K. R., and Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cogn. Psychol.* 66, 232–258. doi: 10.1016/j.cogpsych.2012.12.002
- Paap, K. R., Majoubi, J., Anders-jefferson, R. T., Iosilevsky, R., and Tate, C. U. (2025). Do bilingual advantages in domain-general executive functioning occur in everyday life and/or when performance-based measures have excellent psychometric properties? *Psychol. Rev.* doi: 10.1037/rev0000537
- Paap, K. R., Mason, L., Zimiga, B., Ayala-silva, Y., and Frost, M. (2020). The alchemy of confirmation bias transmutes expectations into bilingual advantages: a tale of two new meta-analyses. *Q. J. Exp. Psychol.* 73, 1290–1299. doi: 10.1177/1747021819900098
- Pauls, L. J., and Archibald, L. M. (2016). Executive functions in children with specific language impairment: a meta-analysis. *J. Speech Lang. Hear. Res.* 59, 1074–1086. doi: 10.1044/2016_JSLHR-L-15-0174
- Perszyk, D. R., and Waxman, S. R. (2018). Linking language and cognition in infancy. *Ann. Rev. Psychol.* 69, 231–250. doi: 10.1146/annurev-psych-122216-011701
- Pliatsikas, C. (2020). Understanding structural plasticity in the bilingual brain: the dynamic restructuring model. *Biling. Lang. Cogn.* 23, 459–471. doi: 10.1017/S1366728919000130
- Pliatsikas, C., and Luk, G. (2016). Executive control in bilinguals: a concise review on fMRI studies. *Biling. Lang. Cogn.* 19, 699–705. doi: 10.1017/S1366728916000249
- Prior, A., and Gollan, T. H. (2011). Good language-switchers are good task-switchers: evidence from Spanish-English and Mandarin-English bilinguals. *J. Int. Neuropsychol. Soc.* 17, 682–691. doi: 10.1017/S1355617711000580
- Quaresima, V., and Ferrari, M. (2019). Functional near-infrared spectroscopy (fNIRS) for assessing cerebral cortex function during human behavior in natural/social situations: a concise review. *Organ. Res. Methods* 22, 46–68. doi: 10.1177/1094428116658959
- Quinn, A., Leahy, M. M., and Mctiernan, K. (2011). Evaluating language assessments with members of the travelling community. *J. Clin. Speech Lang. Stud.* 18, 41–56. doi: 10.3233/ACS-2011-18106
- Ratto, A. B., Potvin, D., Pallathra, A. A., Saldana, L., and Kenworthy, L. (2020). Parents report fewer executive functioning problems and repetitive behaviors in young dual-language speakers with autism. *Child Neuropsychol.* 26, 917–933. doi: 10.1080/09297049.2020.1733512
- Raver, C. C., and Blair, C. (2016). Neuroscientific insights: attention, working memory, and inhibitory control. *Future of Child.* 26, 95–118. doi: 10.1353/foc.2016.0014
- Rilling, J. K. (2008). Neuroscientific approaches and applications within anthropology. *Am. J. Phys. Anthropol. Suppl.* 47, 2–32. doi: 10.1002/ajpa.20947
- Schwering, S. C., and Macdonald, M. C. (2020). Verbal working memory as emergent from language comprehension and production. *Front. Hum. Neurosci.* 14:68. doi: 10.3389/fnhum.2020.00068
- Shokrkon, A., and Nicoladis, E. (2022). The directionality of the relationship between executive functions and language skills: a literature review. *Front. Psychol.* 13:848696. doi: 10.3389/fpsyg.2022.848696
- Siegal, M., and Varley, R. (2002). Neural systems involved in 'theory of mind'. *Nat. Rev. Neurosci.* 3, 463–471. doi: 10.1038/nrn844
- Smart, P. (2022). The evolution, benefits, and challenges of preprints and their interaction with journals. *Sci. Ed.* 9, 79–84. doi: 10.6087/kcse.269
- Tomasello, M., Carpenter, M., Call, J., Behne, T., and Moll, H. (2005). Understanding and sharing intentions: the origins of cultural cognition. *Behav. Brain Sci.* 28, 675–691. doi: 10.1017/S0140525X05000129
- Tracz, V., and Lawrence, R. (2016). Towards an open science publishing platform. *F1000Res.* 5:130. doi: 10.12688/f1000research.7968.1
- Valian, V. (2015). Bilingualism and cognition. *Biling. Lang. Cogn.* 18, 3–24. doi: 10.1017/S1366728914000522
- Vallotton, C., and Ayoub, C. (2011). Use your words: the role of language in the development of toddlers' self-regulation. *Early child. Res. Q.* 26, 169–181. doi: 10.1016/j.jecresq.2010.09.002
- Van den berg, F., Brouwer, J., Tienkamp, T. B., Verhagen, J., and Keijzer, M. (2022). Language entropy relates to behavioral and pupil indices of executive control in young adult bilinguals. *Front. Psychol.* 13:864763. doi: 10.3389/fpsyg.2022.864763
- Walsh, B., Tian, F., Tourville, J. A., Yücel, M., Kuczek, T., and Bostian, A. (2017). Hemodynamics of speech production: an fNIRS investigation of children who stutter. *Sci. Rep.* 7:4034. doi: 10.1038/s41598-017-04357-6
- Wang, Q. (2018). Studying cognitive development in cultural context: a multi-level analysis approach. *Dev. Rev.* 50, 54–64. doi: 10.1016/j.dr.2018.03.002
- Ware, A. T., Kirkovski, M., and Lum, J. A. G. (2020). Meta-analysis reveals a bilingual advantage that is dependent on task and age. *Front. Psychol.* 11:1458. doi: 10.3389/fpsyg.2020.01458
- Wu, Y. J., Chen, M., Thierry, G., Fu, Y., Wu, J., and Guo, T. (2021). Inhibitory control training reveals a common neurofunctional basis for generic executive functions and language switching in bilinguals. *BMC Neurosci.* 22:36. doi: 10.1186/s12868-021-00640-5
- Xiao, N., Bagayi, V., Dandan, Y., Huang, X., Zhong, L., Kiselev, S., et al. (2024). Effectiveness of animal-assisted activities and therapies for autism spectrum disorder: a systematic review and metaanalysis. *Front. Vet. Sci.* 11:1403527. doi: 10.3389/fvets.2024.1403527
- Ye, Z., and Zhou, X. (2009). Executive control in language processing. *Neurosci. Biobehav. Rev.* 33, 1168–1177. doi: 10.1016/j.neubiorev.2009.03.003
- Zelazo, P. D. (2015). Executive function: Reflection, iterative reprocessing, complexity, and the developing brain. *Dev. Rev.* 38, 55–68. doi: 10.1016/j.dr.2015.07.001
- Zimmermann, S., Sakreida, K., Bludau, S., Camilleri, J. A., Hoffstaedter, F., Pelzer, D. I., et al. (2024). Asymmetry, cytoarchitectonic morphology and genetics associated with Broca's area in schizophrenia. *Nat. Ment. Health* 2, 310–319. doi: 10.1038/s44220-023-00200-2



OPEN ACCESS

EDITED BY

Thomas L. Spalding,
University of Alberta, Canada

REVIEWED BY

Schnell Zsuzsanna,
University of Pécs, Hungary
Veronika Svindt,
HUN-REN Hungarian Research Centre for
Linguistics, Hungary

*CORRESPONDENCE

Melanie Glenwright
✉ Melanie.Glenwright@umanitoba.ca

RECEIVED 23 July 2025

ACCEPTED 27 August 2025

PUBLISHED 11 September 2025

CITATION

Smith J and Glenwright M (2025) The
importance of intonation for children's
understanding of verbal irony.
Front. Psychol. 16:1672104.
doi: 10.3389/fpsyg.2025.1672104

COPYRIGHT

© 2025 Smith and Glenwright. This is an
open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](#). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or reproduction
is permitted which does not comply with
these terms.

The importance of intonation for children's understanding of verbal irony

Jordanna Smith and Melanie Glenwright*

Department of Psychology, University of Manitoba, Winnipeg, MB, Canada

Verbal irony refers to any utterance in which the speaker's words mean something different from their intended meaning (e.g., "You're really on top of things" said to a disorganized person). For children who are just learning to recognize verbal irony, a crucial cue to the ironist's intended meaning is their intonation. In this narrative review, we describe research methods for examining how intonation influences children's understanding of verbal irony and the task demands researchers need to consider when designing these studies. Next, we examine how children weigh different cues to verbal irony as they grow older, and we identify cross-linguistic factors that can impact children's use of intonation for irony comprehension. We offer suggestions for planning future studies on this topic by stressing the importance of reducing task demands, acoustically analyzing directional frequency changes, examining children's intonation consideration in languages other than English, and comparing across tonal and non-tonal languages.

KEYWORDS

sarcasm, prosody, task demands, cognitive, linguistic, cross-linguistic

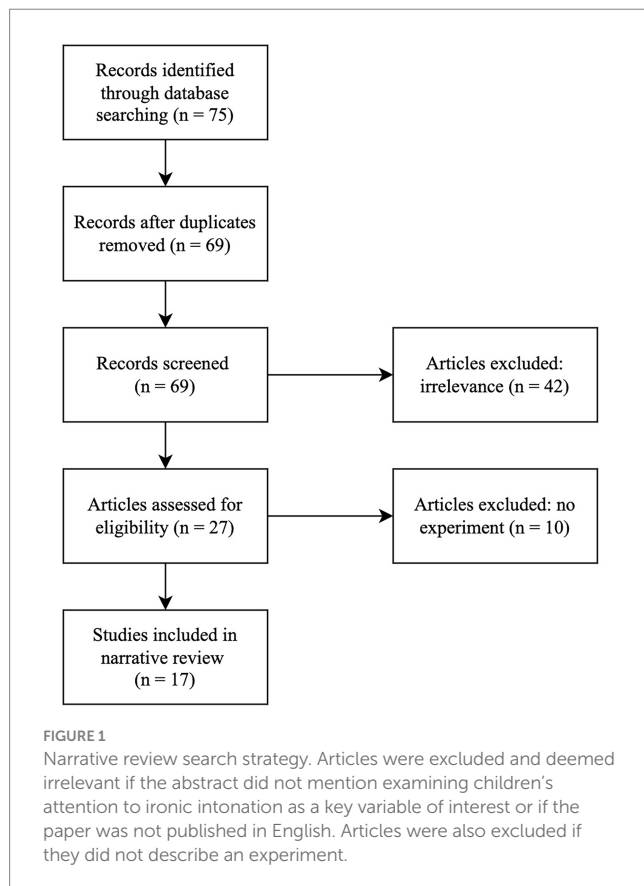
1 Introduction

When a teacher comments "*Fantastic weather we are having*" as she looks out the classroom window at an increasingly worrisome blizzard, how will her Kindergarten students interpret this message? Although these 5-year-old children are just learning to recognize verbal irony, they can rely on various cues to discern the teacher's intended meaning, even though it is different from her literal message. These cues include the fact that the teacher's utterance clearly contradicts the context (e.g., it is snowing heavily), the children's world knowledge (e.g., a blizzard cannot be "fantastic" from a child's point of view), and the speaker's ironic tone of voice. Here, the teacher's intonation can help children in this age group to recognize her use of verbal irony. In this narrative review paper, we provide an overview of research methods used to study children's use of intonation as a cue to verbal irony, and we examine how research design choices can impact task demands on children. Next, we review findings on how children's consideration of intonation and relevant cues to verbal irony changes with development. Finally, we provide an analysis of how these cues might differ between languages, and we suggest directions for future research in this field.

Verbal irony encompasses a broad category of figurative language in which the speaker's words do not directly convey their intended meaning. Sarcasm is one form of verbal irony where the speaker uses nonliteral language to convey a critical attitude toward a person, object, or situation (Kreuz and Glucksberg, 1989). However, there is disagreement in the literature regarding the definitions of verbal irony and sarcasm, which has likely contributed to inconsistencies in findings (Burgers et al., 2011). As much of the research concerning children's understanding of sarcasm has referred to sarcasm as verbal irony, we will do the same to maintain clarity in this paper.

Research with adult participants has established that speakers use verbal irony to criticize, to tease, to mock, and to be funny (Jorgensen, 1996). For the social functions of verbal irony to be served, the listener must recognize the irony and accurately interpret the speaker's intended meaning. This is why ironic speakers usually provide the listener with some paralinguistic cues such as an ironic intonation, exaggerated body language (e.g., eye rolling), and expressions of mirth (e.g., laughing). Compared to a native English speaker's usual sincere prosody, an ironic intonation tends to be slower, lower in pitch, and louder (Rockwell, 2000). Although intonation is one of the most studied cues for children's irony understanding, this body of research has produced mixed results. These divergent findings are likely influenced by differences in theoretical assumptions about verbal irony, differences in research questions, differences in research methods and the demands posed by various methods, and differences in sociocultural functions and implications of verbal irony.

When collecting articles for this review, we searched Google Scholar and our university library database system, which include the following databases: Web of Science, PsycARTICLES, PsycINFO, Scopus, Research Gate, and Directory of Open Access Journals. We included experimental papers that had the following terms in the title or abstract: children, sarcasm or verbal irony, intonation or prosody or tone of voice. Papers were excluded if they did not describe an experiment, if the abstract did not mention examining children's attention to ironic intonation as a key variable of interest, or if the paper was not published in English (i.e., irrelevant; see Figure 1).



1.1 Verbal irony theories

Many of the methodological differences discussed in this narrative review arise from differing theoretical assumptions about the nature of verbal irony. In the first study to examine children's use of ironic intonation as a cue, Ackerman (1983) tested an informal theory suggesting that both contextual incongruity and ironic intonation are cues that signal verbal irony. He manipulated these cues to see which more strongly enhanced children's comprehension. Children aged 6–8 years heard remarks that were either contextually congruent or incongruent, and that were delivered with either an ironic/stressed intonation or a neutral/unstressed intonation. Ackerman predicted that children would understand verbal irony better when it was spoken with ironic intonation and when the statement conflicted with the surrounding context. Since this influential paper, many researchers have adopted a similar approach - manipulating cues to verbal irony to assess their relative effectiveness in improving children's comprehension.

Although several verbal irony theories developed for adults have been tested with children, only the echoic mention theory (Wilson and Sperber, 2012) has been used to examine how children consider intonation as a cue to verbal irony. According to this theory, listeners identify verbal irony when they detect both a critical attitude from the speaker and an “echo” of a previously expressed comment, thought, social norm, or cultural expectation. Researchers have tested this theory's predictions by comparing children's comprehension of verbal irony across experimental conditions that vary in the presence or absence of ironic intonation (Keenan and Quigley, 1999), explicit echoes of prior statements (Keenan and Quigley, 1999), or implicit references to a thought, social norm, or cultural expectation (Köder and Falkum, 2021). These studies collectively contribute to understanding the specific cues children use to recognize verbal irony and how those cues interact in shaping comprehension.

Other verbal irony theories have been adapted to study children's comprehension of irony in ways that do not focus on ironic intonation. The allusional pretense theory (Kumon-Nakamura et al., 1995) proposes that ironic remarks, whether implicit or explicit, refer to failed expectations by violating norms of pragmatic sincerity. According to this theory, speakers use irony to highlight the gap between an unmet expectation and their actual attitude. Researchers have tested this theory by investigating whether children understand ironic criticisms better when they are explicitly incongruent with expectations than when they are implicitly incongruent with social norms (Hancock et al., 2000). They have also examined whether children can detect the insincerity of an ironic speaker who refers to someone else's violation of a socially expected behavior (Creusere, 2000).

Cognitive pragmatics theory (Bara, 2010) proposes that communication follows an interaction pattern, or “behavioral game,” mutually understood by speaker and listener. Comprehension requires the listener to recognize this shared expectation. Simple communicative acts directly reference the game, whereas complex acts involve multiple, conflicting mental representations. The theory predicts a developmental sequence: children first understand and produce sincere statements (i.e., no mental conflict), then lies (i.e., one level of conflict), and finally verbal irony (i.e., multiple conflicts, making it the most difficult). Bosco et al. (2013) found that children under 7 years of age struggled with irony in both simple and complex

scenarios. They concluded that irony comprehension demands more advanced inferencing about speaker intent than lies or sincere statements, emerging around 8 years of age [Bosco et al. \(2013\)](#) suggested that their findings support the theory's stepwise model, with sincere statements acquired first, followed by lies, and lastly irony as the most complex form.

The Parallel Constraint Satisfaction (PCS) framework ([Pexman, 2008](#)) views irony comprehension as a rapid, simultaneous integration of multiple cues as speech unfolds. For example, if you spill coffee on yourself and a friend says with ironic intonation, “*You’re so graceful*,” PCS treats understanding this remark as an ambiguity resolution process. Cues such as the negative event outcome (e.g., spilled coffee), positive statement valence (e.g., “*You’re so graceful*”), and ironic intonation are all activated in the cognitive system, which weighs their relevance until a stable interpretation—literal compliment, ironic criticism, or white lie—emerges. According to the PCS account, a child given these cues should correctly identify the remark as ironic. This has been tested by showing children response objects representing literal (e.g., a nice duck) or ironic (e.g., a mean shark) interpretations and tracking their eye movements in real time ([Climie and Pexman, 2008](#)). Gaze direction and duration toward each object reveal how quickly children consider the speaker’s intent as nice or mean. PCS predicts early gaze shifts toward the ironic interpretation. Studies with children aged 5–8 show support for this model, demonstrating that they use cues such as the speaker’s personality traits ([Climie and Pexman, 2008](#)) and the speaker–listener relationship ([Whalen et al., 2020](#)) when interpreting irony. The PCS framework has motivated widespread use of online, time-sensitive measures to capture how children process irony in real time, not just their final interpretations.

Many of the methodological differences highlighted in this narrative review stem from divergent theoretical assumptions about what verbal irony is, how it is processed, and which cognitive and social abilities are required for its comprehension. These differing assumptions influence the design of research tasks, the measures used to assess understanding, and the interpretation of findings.

1.2 Nature and function of different ironic intonations

In studying the influence of intonation on children’s verbal irony comprehension, an operational definition of ironic intonation is crucial to make valid and generalizable claims. Research done with English-speaking adults suggests that ironic intonation has certain acoustic features that can be verified with a computerized signal editor. Some researchers argue that relative to sincere statements in English, ironic statements tend to be voiced slower, louder, and with a lower pitch ([Rockwell, 2000](#); [Cheang and Pell, 2008](#)). However, research done with English-speaking children has characterized the ironic intonation in a variety of ways. Some have described their narrator’s ironic remarks as having a high pitch, strong energy, and slow rate ([Milosky and Ford, 1997](#)), or a negative, mocking, and nasal tone ([Dews et al., 1996](#)). Others have described this ironic intonation as a mocking and contemptuous tone ([Winner et al., 1987](#)), an exaggerated and mocking tone ([Keenan and Quigley, 1999](#)), or a sarcastic and mocking tone ([Capelli et al., 1990](#)).

Despite these various conceptualizations of ironic intonation, only a few studies have acoustically analyzed intonational stimuli with

software (e.g., [Milosky and Ford, 1997](#); [Laval and Bert-Erboul, 2005](#); [Köder and Falkum, 2021](#)). For example, [Glenwright et al. \(2013\)](#) compared different levels of intonation by adjusting the pitch or fundamental frequency (F0) in which the ironic and literal statements were presented to English-speaking participants. For both 5- to 6-year-old children and adults, utterances with large F0 reductions including both ironic and literal statements were perceived as meaner than those with medium and small F0 reductions. However, given that intonation did not influence humor ratings for either age group ([Glenwright et al., 2013](#)), future research should explore what kinds of intonation can cue children and adults to an ironic speaker’s humorous intent.

In addition to examining the acoustic features of ironic intonation, the [Glenwright et al. \(2013\)](#) study represents another focus in this area of research: the specific functions served by different intonations. Specifically, prosody has been shown to influence children’s inferences of the ironic speaker’s intent or emotion. For instance, [Dews et al. \(1996\)](#) asked 5- to 9-year-old English-speaking children to rate various ironic utterances and found that those vocalized in a sincere (i.e., described as positive and pleasant) tone were rated mean less often and funny more often compared to utterances voiced in a deadpan (i.e., described as flat) or ironic (i.e., described as negative, mocking, and nasal) tone. The researchers concluded that a sincere intonation serves the purposes of muting the intensity and increasing the humor of the criticism, whereas a deadpan intonation makes the ironic comment seem meaner ([Dews et al., 1996](#)). Furthermore, [Milosky and Ford \(1997\)](#) asked adult participants to rate the prosody of various pre-recorded utterances on a 7-point scale ranging from *complimentary* to *sarcastic*. Using these stimuli, the researchers found that relative to an ambiguous intonation (i.e., rated at the midpoint of the scale 85–90% of the time) or a complimentary prosody, an ironic prosody increased the likelihood of English-speaking 6- to 9-year-old children inferring that a negative event had occurred ([Milosky and Ford, 1997](#)). These findings suggest that children can distinguish types of intonation to infer the speaker’s attitude.

Although the studies described all highlight the importance of intonation as a cue to children’s verbal irony understanding, these inconsistencies in conceptualizing and defining ironic intonation have likely contributed to the mixed findings reviewed in the present paper. It is thus important that, when planning future studies, researchers verify their stimuli with acoustical analyses to allow for subsequent replications and comparisons between studies.

1.3 Research methods for studying children’s consideration of ironic intonation

When examining children’s use of intonation as a cue to verbal irony, researchers have manipulated whether ironic remarks are voiced with an ironic, neutral or ambiguous intonation to assess the influence of this change on children’s interpretation of the irony. Aside from cases in which researchers analyzed the acoustic features of their ironic intonation with a computerized signal editor (e.g., [Milosky and Ford, 1997](#); [Laval and Bert-Erboul, 2005](#); [Glenwright et al., 2013](#); [Köder and Falkum, 2021](#)), experimenters often verify that their narrator’s intonation sounds ironic by having adults perceptually

judge their stimuli before collecting data from children (e.g., Milosky and Ford, 1997; Keenan and Quigley, 1999).

Although intonation is one of the most important cues to verbal irony for children, it rarely works in isolation and is often supplemented by other cues. Thus, most studies examining ironic intonation also incorporate some other type of cue to verbal irony, such as contextual information, violation of norms, or the ironic speaker's facial expressions. Sometimes children are presented with visual cues such as illustrations, pictures, or puppet show props to highlight whether the speaker's remark is congruent or incongruent with story context.

In most research described in this narrative review, story narrations were pre-recorded, and children were tested individually. These studies have used standardized methods such as presenting children with audiotaped stories, computerized delivery of stories with pre-recorded audio and photos or illustrations, videos of puppet shows, or videos of actors in conversation. These stories or conversations typically depict a scenario in which an unexpected and negative event occurs (e.g., Chris spills his juice; Zajączkowska, 2016). A speaker then produces an ironic remark that is incongruent with the scenario and voiced in an ironic intonation, such as "Well done!" (Zajączkowska, 2016). As a control condition, children are additionally shown stories where a positive event goes as expected (e.g., Chris is very careful and does not spill his juice) followed by a speaker producing a literal compliment or literal statement that is congruent with the scenario and voiced in a sincere, positive and pleasant intonation (e.g., "Well done!"; Zajączkowska, 2016). The story narration and final ironic or literal remark are usually pre-recorded by the same speaker to maintain consistency of intonation across participants and scenarios.

After each scenario is presented, researchers determine children's comprehension of each remark by asking them questions about the facts of the story, the speaker's belief, the speaker's emotions, or the speaker's intended meaning. Children's appreciation of each remark can also be assessed by their responses to questions about the speaker's attitude and the speaker's intentions. After an ironic sentence (or literal sentence) has been provided, children can answer test questions by verbalizing their responses, by writing them, or by pointing to markers on rating scales. Additionally, some researchers have examined children's online responses as the ironic sentence unfolds in real-time by recording and timing their spontaneous eye gazes toward illustrations on a computer screen. Next, we will examine the task demands of these design features and how they can influence children's successful irony comprehension.

2 Task demands in methods assessing children's verbal irony comprehension

When designing studies on children's consideration of intonation as a cue to verbal irony, researchers must be cognizant that participating in this research poses cognitive processing demands on children. Each design choice—how verbal irony is presented, how questions are asked, and how children are required to respond—can raise or lower the threshold at which children successfully comprehend verbal irony. Researchers should aim to minimize this threshold by making design choices that place low demands on children's attention, working memory, world knowledge, and linguistic skills.

2.1 Verbal irony presentation modes

When aiming to choose methods for presenting verbal irony with low task demands, researchers should maintain children's attention and provide them with visual aids. An ideal method with low demands on attentional resources is to have children participate in the context of the ironic event so they can see a live presentation of the ironic speaker's intonation and facial expression (e.g., Rattray and Tolmie, 2008). If this is not feasible, providing children with moving puppet shows or videos of moving people or puppets can maintain their attention better than still illustrations (Hill et al., 2016). Importantly, stories presented with illustrations are less cognitively demanding with stories presented without illustrations. Illustrations allow children to see the story events, so they are not required to construct internal representations of these events (Milosavljevic, 2024). To minimize cognitive demands, these visualizations should be simple and contain only relevant parts of the story (Köder and Falkum, 2021). Illustrations or visual aids can help children focus their attention on the stimuli, and they can provide information about the context of the ironic statement. Visual aids can also reduce linguistic demands on children by providing them with images for words that might not be in their vocabulary. If visual aids can remain in the child's sight when they are required to answer offline test questions after the story has completed, this can help them remember the story context and the context valence with reduced demands on their working memory. Engaging methods with visual aids can increase the availability of children's cognitive resources for processing and responding to test questions, thereby reducing the threshold for children's verbal irony comprehension.

Researchers should also use simple language in devising verbal irony stories and aim to keep the task duration short as possible. These stories should be constructed with a basic vocabulary using short sentences to minimize linguistic demands (Banasik-Jemielniak and Bokus, 2019). It is also recommended that these stories are constructed with three or four sentences to keep the working memory demands low (Milosavljevic, 2024). Lastly, the task duration should be short (Banasik-Jemielniak and Bokus, 2019) because children can experience cognitive fatigue when they participate in several successive experimental trials without taking a rest, and this can increase their threshold for verbal irony comprehension. Next, we will consider how the format of test questions influences task demands for children.

2.2 Question formats

Researchers also need to consider the cognitive load of the test questions they choose. Verbal irony test questions usually refer to the speaker's belief or intentions to gauge children's comprehension and appreciation. To answer these questions correctly, the child must process the test question, story context, and verbal irony cues to infer the speaker's mental state and then maintain all this information in working memory while formulating a response. Shorter questions with fewer response options pose lower task demands. For example, Köder and Falkum (2021) asked children forced-choice questions with binary single-word response options (e.g., How is Mum feeling inside? Is she happy or angry?). Other researchers have asked children longer questions such as "Show me how funny or serious Jasmine was

trying to be when she said, “*You picked a great spot for hide-and-seek*” then provided children with a 6-point face scale (Glenwright et al., 2013). Here, the syntactic structure of this lengthy question can pose linguistic and working memory demands on children (Fuchs, 2023), and this increases the threshold for children’s verbal irony comprehension. Compared to short options, lengthy response options for forced-choice questions require children to maintain more information in working memory to formulate a response. Furthermore, responding to several questions before answering a key comprehension question increases working memory demands (Falkum and Köder, 2024), and this can increase the threshold for children’s irony comprehension.

On the other hand, forced-choice questions can be problematic when they present children with options they would not generate or deem suitable, leading them to rely on elimination strategies rather than genuine understanding (Cassels and Birch, 2014). For instance, asking “Is [the speaker] happy or angry?” may be insufficient in cases where the speaker is expressing more subtle emotions than happiness or anger (e.g., disappointment about the weather). Children may select a response option despite feeling unsure, and in turn their forced-choice response options may not be a clear measure of their understanding. This reduces their ecological validity compared to open-ended questions, which better approximate natural conversation (e.g., “Why do you think they said that?”; Zajackowska, 2016). Assessing verbal irony often requires children to draw on their understanding of language, infer the speaker’s beliefs or intentions, and articulate responses that reflect pragmatic awareness and metacognitive reasoning. Evidence indicates that 4- to 8-year-olds demonstrate more advanced perspective-taking and emotion recognition when assessed with open-ended than with forced-choice formats (Cassels and Birch, 2014), suggesting that open-ended questions might also be a more sensitive measure of verbal irony comprehension.

In summary, when aiming to optimize children’s success on verbal irony tasks, researchers should avoid offering children long response options because they increase the threshold for irony comprehension. Rather, they should aim to choose short forced-choice questions with minimal, carefully tailored response options or short open-ended questions. These methods reduce the risk of misinterpretation and can provide clearer insight into children’s irony comprehension processes.

2.3 Response modes

Verbal irony task demands also vary according to how children are required to respond to test questions. Most studies examining intonation as a cue to verbal irony have required children to respond verbally (e.g., Winner et al., 1987; Milosky and Ford, 1997; Keenan and Quigley, 1999; Ratray and Tolmie, 2008; Zajackowska, 2016; Yu et al., 2024). When presented with these test questions, children younger than 5 years of age have difficulty formulating verbal responses that reflect their own thoughts and judgments because their structural language skills and vocabularies are limited (Milosavljevic, 2024). Questions requiring verbal responses place demands on children’s verbal explanatory skills and metapragmatic reasoning (Falkum and Köder, 2024), and this can increase the threshold for their ability to produce correct responses.

Less commonly, researchers have used questions where children could choose to respond verbally or nonverbally with a motor

response by circling an icon in writing or by pointing to an icon (e.g., Dews et al., 1996; Glenwright et al., 2013). For example, some researchers have presented children with icons on a printed rating scale or a computer touchscreen, and asked children a forced-choice question such as “How funny was the turtle being? Circle one of the faces” (Dews et al., 1996). Because verbal forced-choice and open-ended responses rely on verbal ability (Cassels and Birch, 2014), using icons (e.g., emojis) as nonverbal response options can reduce the linguistic task demands compared to questions requiring a verbal response, in turn lowering the threshold for correct responses.

It is also important for researchers to recognize that their selection of response options can influence children’s verbal irony comprehension and the specificity of study results. Both the type of response icons (e.g., faces, animals, etc.) and their labels (e.g., nice, mean, angry, funny, serious, etc.) affect how well children understand the task. For example, Glenwright et al. (2013) showed 5–6-year-olds puppet show videos containing ironic criticisms, literal criticisms, and literal compliments. To assess their understanding of the speaker’s attitude, children were asked, “*Show me how nice or mean Jasmine was trying to be when she said, ‘You picked a great spot for hide-and-seek.’*” Responses were given on a 6-point Nice/Mean Scale with illustrated faces ranging from “very nice” (i.e., a joyful expression, relaxed eyebrows, broad smile) to “very mean” (i.e., furrowed brows, downturned mouth). To a 5-year-old, these faces might simply represent “very happy” to “very angry,” regardless of the provided labels. This can cause confusion, leading children to choose faces they like, or to base their choice on their own feelings or those of the addressee puppet, rather than the speaker’s intended attitude. Furthermore, labels like nice or mean may oversimplify the intended meaning—in this case, the speaker’s attitude might be more accurately described as critical or dismissive. This example shows that tailoring response-option icons and labels to children’s abilities can reduce measurement specificity and mask performance on verbal irony tasks, producing results that reflect how children interpret the options rather than their actual understanding.

All question formats described so far are offline questions where children are required to allow the experimenter to finish verbalizing the question before formulating a response. More recently, experimenters have been examining children’s spontaneous reactions to verbal irony stories as they unfold in real time. Köder and Falkum (2021) tracked children’s eye gaze locations and durations as they were presented with stories ending with ironic or literal statements that varied according to intonation and social norm information. Children’s implicit eye gaze behavior showed they were considering the ironic icon online as they heard an ironic criticism spoken, before the test question was asked. This contrasted with 6-year-old children’s explicit nonverbal pointing responses to an offline question, which indicated irony comprehension levels below a chance rate. In this study, the offline question was a direct, elicited-response question, and thus demanded a response. To answer this question correctly, a child must process the story context and verbal irony cues to infer the speaker’s mental state, maintain this information in working memory, process the test question, and formulate a response (Glenwright et al., 2021). However, when Köder and Falkum (2021) examined 3-year-old children’s online eye gaze reactions to ironic criticisms before a test question was asked, they discovered that children’s spontaneous responses could reveal some recognition of the features of verbal irony (i.e., intonation and contextual cues) that was not accounted for

by the elicited-response measures. Infant cognition researchers contend that this occurs because elicited-response tasks pose considerable cognitive and verbal demands whereas spontaneous-response tasks do not (Scott and Baillargeon, 2009). These kinds of online measures assess children's cognitive processes through their spontaneous eye gaze reactions, in turn revealing their real-time processing of a scene as it unfolds without the need to invoke a response-selection process (Scott and Baillargeon, 2009; Glenwright et al., 2021). The contrast between 3-year-olds' spontaneous eye gazes toward the correct ironic icon and their below-chance pointing responses to offline questions in Köder and Falkum (2021) supports Scott and Baillargeon's response-selection account. However, this account has not yet been directly tested in verbal irony research, presenting an avenue for future investigation. Nonetheless, online spontaneous-response measures such as eye-tracking methods pose significantly lower cognitive and linguistic task demands than offline elicited-response measures and can thus provide additional insight into the cognitive processes that occur when children as young as 3 interpret verbal irony.

We encourage researchers to conduct more online studies on children's spontaneous attention to intonation as a cue to verbal irony. In the next section, we describe how children's ability to consider cues to verbal irony changes throughout development.

3 Development of children's use of intonation and other cues to understand verbal irony

Most researchers agree that children typically detect verbal irony by 6 years of age (de Groot et al., 1995; Dews et al., 1996; Fuchs, 2023; Pexman, 2023; Olkonien et al., 2025). However, some studies have reported recognition of some of the features of verbal irony in children as young as 3 years old (Ratray and Tolmie, 2008; Köder and Falkum, 2021). For instance, Ratray and Tolmie (2008) found that English-speaking 3-year-old children can recognize an ironic intonation, however, 4-year-olds are able to integrate a variety of cues to ironic intent (Ratray and Tolmie, 2008). Similarly, Köder and Falkum (2021) found that Norwegian-speaking 3-year-olds' eye movements show sensitivity to some of verbal irony's features. Relative to a deadpan (i.e., flat, slower tempo, lower pitch level, and greater intensity) intonation, a parodic (i.e., imitative and exaggerated) tone of voice led children to look more at an angry emoticon at the offset of an ironic utterance (Köder and Falkum, 2021). The authors argue that the exaggerated nature of the parodic intonation makes it easier for children to recognize the speaker's negative attitude (Köder and Falkum, 2021). However, when asked about the speaker's meaning after the utterance was processed, 3-year-olds incorrectly interpreted ironic utterances literally. These results suggest that this early sensitivity to intonation cues is not sufficient for successful interpretation of verbal irony (Köder and Falkum, 2021).

Despite these findings showing early recognition of intonation cues, the ability to effectively use these cues for verbal irony interpretation seems to develop a couple of years later. Some researchers have argued that by age 6, children's interpretation of verbal irony is improved when it is uttered with an ironic intonation (de Groot et al., 1995; Zajączkowska, 2016), while others have argued that children this age are able to use intonation cues as effectively as

adults (Dews et al., 1996; Nakassis and Snedeker, 2002). However, other researchers have suggested that the ability to recognize and interpret intonation cues to verbal irony continues to improve as children age beyond 6 years old. For instance, Milosky and Ford (1997) found that English-speaking 6-year-olds had more difficulty identifying ironic intonation from an ambiguous intonation than from a complimentary (i.e., to the statement, as rated by participants) intonation. This suggests that although they were able to recognize non-literal intent signaled by intonation, it was not always sufficient for correct interpretation of the irony. Furthermore, Ackerman (1983) found that English-speaking 8- to 9-year-olds were better than 6- to 7-year-olds at using intonation and context to recognize verbal irony, noting that the older children's use of intonation cues was comparable to adults (Ackerman, 1983). Indeed, there are mixed results regarding development of the ability to use intonation as a cue to irony.

3.1 Children's use of ironic intonation relative to other cues

There is also a lack of consensus regarding the order in which children recognize different cues to verbal irony. In particular, contextual incongruity cues, which are based on situational information that does not align with the literal form of the ironic statement, have received much attention in the literature (e.g., Ackerman, 1983; Winner et al., 1987; Capelli et al., 1990). Some researchers argue that children first recognize intonation cues to verbal irony, and the ability to appreciate contextual cues follows shortly after. For instance, Zajączkowska (2016) found that Polish-speaking 5-year-old children rely on intonation when interpreting irony, as contextual discrepancy cues alone are not sufficient. However, by 6 years old, children seem to rely more on contextual cues as their ability to interpret irony develops (Zajączkowska, 2016). Similarly, Laval and Bert-Erboul (2005) found that some level of understanding emerges around age 5 in French-speaking children, but only when ironic intonation was present. By 7 years old, contextual cues were equally important as intonation. This was shown by the 7-year-olds' strategic use of both cues, which involved basing their interpretation on one cue when the other was lacking, and their hindered performance when both cues were lacking. These researchers posit that children rely more heavily on intonation at first and develop a more encompassing use of various cues to understand irony as they develop (Laval and Bert-Erboul, 2005). Other researchers hold that this primary reliance on intonation cues continues into middle childhood. Ackerman (1983) reported that English-speaking 8- to 9-year-olds use intonation as effectively as adults, but they are less sensitive to contextual discrepancy cues. He suggested that children might rely on intonation as a cue prior to developing information processing skills that allow them to integrate contextual information across sentences (Ackerman, 1983). Similarly, Capelli et al. (1990) found that 8- to 12-year-old English-speaking children rely much more heavily on ironic intonation than context in irony interpretation.

In contrast to findings highlighting children's preference for intonation cues, there are reports that children rely more on context than intonation to understand verbal irony. For example, Stiles and Nadler (2013) compared consideration of cues in 5- to 9-year-old English-speaking children with hearing loss and with normal hearing.

They found that children with hearing loss interpreted utterances as ironic and generally relied on intonation less often than children with normal hearing. However, they found that children in both groups relied more heavily on context than intonation (Stiles and Nadler, 2013), suggesting that children as young as 5 rely more on contextual cues than on intonation cues. Other researchers have taken this argument even further by claiming that 8-year-old English-speaking (Winner et al., 1987) and French-speaking (Aguert et al., 2010) children do not take intonation into account when interpreting speaker intentions. Clearly, there is a lack of consensus in the literature on the developmental sequence in children's recognition of intonation and contextual cues to verbal irony.

Although contextual cues have received the most attention in the literature, some researchers have examined how other types of verbal irony cues work in conjunction with intonation cues. For instance, echoic mention refers to knowledge that the listener has from previous statements made by the speaker. One study examining echoic mention found increased comprehension in Norwegian-speaking 4- to 5-year-old children when a moral norm, as opposed to a social norm, was violated (Köder and Falkum, 2021). The researchers suggest that the children's preference for the echoic mention cues may indicate that for children, intonation is less reliable than other cues in verbal irony comprehension. Similarly, Keenan and Quigley (1999) found that 6- to 10-year-old English-speaking children's understanding of irony is best when various sources of information are available (i.e., echoic mention and ironic intonation), but that echoic mention alone can be used as a cue when necessary. Keenan and Quigley (1999) claimed that, despite the utility of intonation, it is not a necessary feature of verbal irony. In addition to echoic mention cues, researchers have identified facial expressions such as expressive mouth movements and eyebrow flashes as cues to verbal irony (Aguert, 2021). Yu et al. (2024) found that 3- to 7-year-old Mandarin-speaking children's comprehension of ironic criticisms was improved with facial expression and intonation cues. Although intonation is often a useful cue to children learning to understand verbal irony, it is most effective when used in conjunction with various other cues such as facial expressions or echoic references to previous utterances by the ironic speaker.

Research on children's use of cues to ironic intent has produced wildly divergent results which are likely due to diversity in research questions, research methods, and the task demands of these methods. It is also important to note that studies on children's verbal irony comprehension across different languages reveal sociocultural differences that shape how children interpret irony (discussed in section 4.2). Next, we describe how the ironic intonation differs across languages, and we offer suggestions for how these differences might impact children's understanding of verbal irony.

4 Cross-linguistic differences in ironic intonation

In the studies we have reviewed, children were presented with an ironic intonation in English, Cantonese, French, Norwegian, Mandarin, or Polish. Research with adults has demonstrated that the acoustic features of the ironic intonation vary across languages (Bettelli et al., 2024). In this section, we will review these features and suggest how these features might influence children's understanding of verbal irony.

4.1 Differences in ironic intonation due to linguistic structure

One intonational feature of verbal irony that may differ between languages is the pitch variation used to signal ironic intent (Rockwell, 2000; Glenwright et al., 2013; Bettelli et al., 2024). For example, relative to literal speech, verbal irony is often spoken with a higher pitch in French, Italian, and Cantonese, whereas it is often spoken in a lower pitch in English and German (Bettelli et al., 2024). In line with this argument, studies have shown that English-speaking children and adults are more likely to recognize verbal irony when the speaker uses a lower pitch as measured by reduced F0 (Cheang and Pell, 2008; Glenwright et al., 2013). Furthermore, González-Fuente et al. (2016) tested the effect of pitch variation on French-speaking adults' verbal irony comprehension and found that ironic utterances were produced with a higher F0 than literal utterances. In support of Bettelli et al. (2024) argument, some results show the opposite pattern for English speakers (Glenwright et al., 2013).

In addition to pitch variation, the pitch range used to signal verbal irony may also function differently across languages. For instance, verbal irony spoken in languages including English, German, and Cantonese is often produced within a reduced pitch range (i.e., relative to literal speech), whereas the pitch range is generally expanded for verbal irony in languages such as French and Italian (Bettelli et al., 2024). However, this does not mean that pitch range variation is necessarily prioritized equally across languages. For instance, González-Fuente et al. (2016) found that while F0 change was a significant predictor of ironic intent in both English and French, word lengthening was a more powerful cue to verbal irony than pitch range expansion for French speakers. These findings may have important implications regarding the cues recognized by French-speaking children relative to their English-speaking counterparts, as they may rely more on rate of speech than pitch range in identifying irony. However, one similarity noted by Bettelli et al. (2024) across all six of the languages they examined (i.e., Italian, French, German, Dutch, English, and Cantonese) is the rate of speech: ironic utterances tend to have a longer duration than literal speech. This suggests that although English- and French- speakers may prioritize intonational features (i.e., pitch range, rate of speech) differently, children learning to interpret verbal irony in both languages rely on these features to some extent.

In addition to the varying features of ironic intonation across different languages, there are languages in which intonation is less flexible and thus may be more difficult for children to use as a cue to irony. For instance, 60–70% of languages (e.g., many Asian, African, and Indigenous American languages; Best, 2019) use lexical tones to distinguish words with different meanings. Although they function uniquely between different languages, these lexical tones generally restrict the pitch variation that an ironic speaker can use, since the tonal change itself alters the meaning of words. Li et al. (2013) demonstrated this constraint in a study of 8- to 12-year-old Cantonese-speaking children with and without ASD. These researchers incorporated intonation and sentence-final particles (SFPs), a linguistic feature of Cantonese that do not have a direct counterpart in English and can be used as a cue to verbal irony (Li et al., 2013). Interestingly, their results showed that for typically developing children, intonation alone was not sufficient for irony comprehension. Rather, performance was best in the other two conditions: SFPs only, and both intonation

and SFPs. Li et al. (2013) argued that the discrepancy between their findings and findings from English-speaking participants (e.g., Ackerman, 1983; Capelli et al., 1990) is likely due to the tonal nature of Cantonese. The authors suggest this could be due to a mutual compensation in which greater use of sentence particles lessens the significance of intonation patterns in a language, and vice versa (Li et al., 2013). Thus, intonation is likely less prioritized as a cue for children who speak tonal as opposed to non-tonal languages. This argument is further supported by the finding that Mandarin-speaking children integrate facial expression with intonation cues for verbal irony interpretation (Yu et al., 2024), a strategy that may result from restrictions on lexical tones in Mandarin. However, the researchers incorporated this additional cue in consideration of the children with hearing loss in this study, who likely rely more heavily on non-prosodic cues than children with typical hearing.

We recommend that future studies expand on the current literature highlighting the different features of ironic intonation across languages. Furthermore, more research is necessary regarding the differences in cues used by children to recognize verbal irony in tonal and non-tonal languages. This would allow for a better understanding of specific language features that children attend to at different developmental stages.

4.2 Differences in ironic intonation due to cultural variation

Some researchers have argued that despite the social-cognitive processes involved in verbal irony use and comprehension, socio-cultural context is often neglected in studies on verbal irony (Banasik-Jemielniak and Kałowski, 2022). One of the most important factors in irony use and comprehension is shared knowledge between the ironic speaker and the listener, often resulting from similar experiences or attitudes (Averbeck and Hample, 2008). This implies that verbal irony may be used to serve different social functions depending on the socio-cultural context. For instance, cultures differ in their use of non-literal language, such that individualistic cultures generally prioritize direct communication while collectivistic cultures tend to value indirect meaning (Holtgraves, 1997). However, some findings have shown that people from individualistic cultures are more likely to use verbal irony than those from collectivistic cultures (Rockwell and Theriot, 2001), suggesting that verbal irony may be serving different social functions in different cultural contexts. One study examined irony perceptions across participants from three countries varying in levels of individualism/collectivism and power distance (i.e., societal emphasis placed on hierarchy): China (i.e., high collectivism and high power distance), the United States (i.e., high individualism, low power distance) and Mexico (i.e., in between China and the United States; Blasko et al., 2021). The researchers found higher rates of verbal irony use in participants from the United States and Mexico than from Chinese participants, explaining that verbal irony may pose higher social risks in China due to social dynamics of harmony and respect in power hierarchies (Blasko et al., 2021).

In consideration of the cross-linguistic variation of the social functions served by intonation, it is important to recognize that more than one ironic tone of voice has been identified for English, which is likely the case for other languages as well. These include a

pretense-based, enthusiastic intonation and a more contemptuous, deadpan one (Sperber, 1984). Indeed, some studies have shown the influence of irony spoken with different types of ironic intonation on verbal irony comprehension in both 5- to 9-year-old children and adults (Dews et al., 1996). Specifically, an ironic (i.e., negative, mocking) tone seems to imply that the speaker is annoyed, whereas a deadpan (i.e., flat) intonation implies playfulness or superiority (Dews et al., 1996). In contrast, a study done with Norwegian-speaking children reported no significant differences in irony processing between utterances spoken in parodic and deadpan tones (Köder and Falkum, 2021). However, Köder and Falkum (2021) did report that children were more likely to look at an angry emoticon at the offset of an ironic utterance spoken in a parodic (i.e., relative to deadpan) tone. These findings suggest that both English- and Norwegian-speaking children use intonation cues to interpret a speaker's attitude, even if the verbal irony is not consistently and accurately understood. Furthermore, Aguert et al. (2010) suggest that for French-speaking children, a speaker's intonation can provide cues to both their ironic intent and their emotional state.

Clearly, there are social implications of various ironic intonations that influence children's interpretations of verbal irony and likely vary across languages. However, due to the limited number of studies on ironic intonation in languages besides English, more research is needed to provide a clearer picture of its cross-linguistic effects on children's understanding of verbal irony.

5 Directions for future research

In this section, we will synthesize suggestions for future experimental investigations of children's consideration of ironic intonation. Milosky and Ford (1997) argued that irony comprehension depends on many different cues, and their relative strengths are determined by the specific situation. Thus, rather than continuing to ask whether intonation is a cue at all, it is worth examining when intonation is a strong cue to understanding irony. Simulating the nature of real-life situations involving irony would provide a clearer picture of how intonation interacts with various other information sources. In turn, this research would present a more thorough understanding of irony comprehension development in children (Milosky and Ford, 1997). Similarly, Nakassis and Snedeker (2002) suggested future research would benefit from examining the generalization strategy that children develop to detect and interpret irony. This would allow for a better understanding of just how broad this generalization skill is, and how it may coincide with other developmental changes.

Taking a more theoretical approach, Keenan and Quigley (1999) suggested that future studies should incorporate measures of children's understanding of both the speaker's intended meaning and the speaker's communicative intention (i.e., whether the speaker meant to be understood as ironic), as this would allow for a more complete model of how irony comprehension develops. Stiles and Nadler (2013) suggested that future research should examine the interaction between intonation and context in children's interpretation of irony to gain a deeper understanding of these processes. They further suggested the incorporation of other, less-studied cues such as speaker familiarity and facial expression.

We have reviewed only one study that highlights the importance of facial expression as a cue to children's irony understanding (i.e., Yu et al., 2024) and the research field would certainly benefit from future

research on this topic. Furthermore, Zajęczkowska (2016) recommended that future researchers conduct experiments examining the effects of intonation on more complex forms of irony than sarcasm (e.g., hyperbole or understatement) to provide a more complete picture of how intonation is used as a cue in irony comprehension. We agree with the need for more breadth in figurative language forms in future research, and we additionally suggest that more research needs to be done with children who speak languages other than English. Finally, future research is needed to study online measures in this field, given that eye tracking methods and eye gaze latencies can give us a more detailed examination of how children consider intonation when deciphering a speaker's meaning and intentions.

6 Conclusion

We suggest that researchers carefully consider the task demands of methods when planning future studies on children's consideration of intonation as a cue to verbal irony, and we suggest that online methods can provide valuable information regarding the cognitive processes of verbal irony interpretation because they pose the lowest demands on children. In future studies, researchers need to acoustically analyze directional F0 changes in their intonation stimuli, include child participants who speak languages other than English, and make direct comparisons across languages, especially tonal and non-tonal ones.

Author contributions

JS: Writing – original draft, Writing – review & editing, Data curation, Investigation. MG: Writing – review & editing, Supervision, Writing – original draft, Conceptualization.

References

- Ackerman, B. P. (1983). Form and function in children's understanding of ironic utterances. *J. Exp. Child Psychol.* 35, 487–508. doi: 10.1016/0022-0965(83)90023-1
- Agurt, M. (2021). Paraverbal expression of verbal irony: vocal cues matter and facial cues even more. *J. Nonverb. Behav.* 46, 45–70. doi: 10.1007/s10919-021-00385-z
- Agurt, M., Laval, V., Le Bigot, L., and Bernicot, J. (2010). Understanding expressive speech acts: the role of prosody and situational context in French-speaking 5- to 9-year-olds. *J. Speech Lang. Hear. Res.* 53, 1629–1641. doi: 10.1044/1092-4388(2010/08-0078)
- Averbeck, J. M., and Hample, D. (2008). Ironic message production: how and why we produce ironic messages. *Commun. Monogr.* 75, 396–410. doi: 10.1080/03637750802512389
- Banasik-Jemielniak, N., and Bokus, B. (2019). Children's comprehension of irony: studies on polish-speaking preschoolers. *J. Psycholinguist. Res.* 48, 1217–1240. doi: 10.1007/s10936-019-09654-x
- Banasik-Jemielniak, N., and Kałowski, P. (2022). Socio-cultural and individual factors in verbal irony use and understanding: what we know, what we don't know, what we want to know. *Rev. Commun. Res.* 10, 80–113. doi: 10.12840/issn.2255-4165.036
- Bara, B. G. (2010). Cognitive pragmatics: the mental processes of communication. 1st Edn. Cambridge, MA: MIT Press.
- Best, C. T. (2019). The diversity of tone languages and the roles of pitch variation in non-tone languages: considerations for tone perception research. *Front. Psychol.* 10:364. doi: 10.3389/fpsyg.2019.00364
- Bettelli, G., Giustolisi, B., and Panzeri, F. (2024). Cross-linguistic recognition of irony through visual and acoustic cues. *J. Psycholinguist. Res.* 53:73. doi: 10.1007/s10936-024-10111-7
- Blasko, D. G., Kazmerski, V. A., and Dawood, S. S. (2021). Saying what you don't mean: a cross-cultural study of perceptions of sarcasm. *Can. J. Exp. Psychol.* 75, 114–119. doi: 10.1037/cep0000258
- Bosco, F. M., Angeleri, R., Colle, L., Katiucsi, S., and Bara, B. G. (2013). Communicative abilities in children: an assessment through different phenomena and expressive means. *J. Child Lang.* 40, 741–778. doi: 10.1017/S0305000913000081
- Burgers, C., van Mulken, M., and Schellens, P. J. (2011). Finding irony: an introduction of the verbal irony procedure (VIP). *Metaphor Symbol.* 26, 186–205. doi: 10.1080/10926488.2011.583194
- Capelli, C. A., Nakagawa, N., and Madden, C. M. (1990). How children understand sarcasm: the role of context and intonation. *Child Dev.* 61, 1824–1841. doi: 10.1111/j.1467-8624.1990.tb03568.x
- Cassels, T. G., and Birch, S. A. J. (2014). Comparisons of an open-ended vs. forced-choice 'mind reading' task: implications for measuring perspective-taking and emotion recognition. *PLoS One* 9, 1–20. doi: 10.1371/journal.pone.0093653
- Cheang, H. S., and Pell, M. D. (2008). The sound of sarcasm. *Speech Comm.* 50, 366–381. doi: 10.1016/j.specom.2007.11.003
- Climie, E. A., and Pexman, P. M. (2008). Eye gaze provides a window on children's understanding of verbal irony. *J. Cogn. Dev.* 9, 257–285. doi: 10.1080/15248370802247939
- Creusere, M. A. (2000). A developmental test of theoretical perspectives on the understanding of verbal irony: children's recognition of allusion and pragmatic insincerity. *Metaphor Symbol.* 15, 29–45. doi: 10.1207/S15327868MS151&2_3
- de Groot, A., Kaplan, J., Rosenblatt, E., Dews, S., and Winner, E. (1995). Understanding versus discriminating nonliteral utterances: evidence for a dissociation. *Metaphor Symbol. Act.* 10, 255–273. doi: 10.1207/s15327868ms1004_2
- Dews, S., Winner, E., Kaplan, J., Rosenblatt, E., Hunt, M., Lim, K., et al. (1996). Children's understanding of the meaning and functions of verbal irony. *Child Dev.* 67, 3071–3085. doi: 10.2307/1131767
- Falkum, I. L., and Köder, F. (2024). "Investigating irony comprehension in children: methods, challenges, and ways forward" in Studying verbal irony and sarcasm:

Funding

The author(s) declare that no financial support was received for the research and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The authors declare that no Gen AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Methodological perspectives from communication studies and beyond. eds. N. Banasik-Jemielniak, P. Kałowski and M. Zajęczkowska (Cham: Palgrave Macmillan).
- Fuchs, J. (2023). 40 years of research into children's irony comprehension. *Pragmat. Cogn.* 30, 1–30. doi: 10.1075/pc.22015.fuc
- Glenwright, M., Parackel, J., Cheung, K. R. J., and Nilsen, E. (2013). Intonation influences how children and adults interpret sarcasm. *J. Child Lang.* 41, 472–484. doi: 10.1017/S0305000912000773
- Glenwright, M., Scott, R. M., Bilevicius, E., Pronovost, M., and Hanlon-Dearman, A. (2021). Children with autism spectrum disorder can attribute false beliefs in a spontaneous-response preferential-looking task. *Front. Commun.* 6, 1–12. doi: 10.3389/fcomm.2021.669985
- González-Fuente, S., Prieto, P., and Noveck, I. (2016). A fine-grained analysis of the acoustic cues involved in verbal irony recognition in French. Boston: International Speech Communication Association.
- Hancock, J. T., Dunham, P. J., and Purdy, K. (2000). Children's comprehension of critical and complimentary forms of verbal irony. *J. Cogn. Dev.* 1, 227–248. doi: 10.1207/S15327647JCD010204
- Hill, L. J. B., Coats, R. O., Mushtaq, F., Williams, J. J. H., Aucott, L. S., and Mon-Williams, M. (2016). Moving to capture children's attention: developing a methodology for measuring visuomotor attention. *PLoS One* 11:e0159543. doi: 10.1371/journal.pone.0159543
- Holtgraves, T. (1997). Styles of language use: individual and cultural variability in conversational indirectness. *J. Pers. Soc. Psychol.* 73, 624–637. doi: 10.1037//0022-3514.73.3.624
- Jorgensen, J. (1996). The functions of sarcastic irony in speech. *J. Pragmat.* 26, 613–634. doi: 10.1016/0378-2166(95)00067-4
- Keenan, T. R., and Quigley, K. (1999). Do young children use echoic information in their comprehension of sarcastic speech? A test of echoic mention theory. *Br. J. Dev. Psychol.* 17, 83–96. doi: 10.1348/026151099165168
- Köder, F., and Falkum, I. L. (2021). Irony and perspective-taking in children: the roles of norm violations and tone of voice. *Front. Psychol.* 12:624604. doi: 10.3389/fpsyg.2021.624604
- Kreuz, R. J., and Glucksberg, S. (1989). How to be sarcastic: the echoic reminder theory of verbal irony. *J. Exp. Psychol. Gen.* 118, 374–386. doi: 10.1037/0096-3445.118.4.374
- Kumon-Nakamura, S., Glucksberg, S., and Brown, M. (1995). How about another piece of pie: the allusional pretense theory of discourse irony. *J. Exp. Psychol. Gen.* 124, 3–21. doi: 10.1037/0096-3445.124.1.3
- Laval, V., and Bert-Erboul, A. (2005). French-speaking children's understanding of sarcasm: the role of intonation and context. *J. Speech Lang. Hear. Res.* 48, 610–620. doi: 10.1044/1092-4388(2005/042)
- Li, J. P., Law, T., Lam, G. Y., and To, C. K. (2013). Role of sentence-final particles and prosody in irony comprehension in Cantonese-speaking children with and without autism spectrum disorders. *Clin. Linguist. Phon.* 27, 18–32. doi: 10.3109/02699206.2012.734893
- Milosavljevic, A. (2024). “Experiments on the development of irony: walking through a methodological maze” in Studying verbal irony and sarcasm: Methodological perspectives from communication studies and beyond. eds. N. Banasik-Jemielniak, P. Kałowski and M. Zajęczkowska (Cham: Palgrave Macmillan).
- Milosky, L. M., and Ford, J. A. (1997). The role of prosody in children's inferences of ironic intent. *Discourse Process.* 23, 47–61. doi: 10.1080/01638539709544981
- Nakassis, C., and Snedeker, J. (2002). Beyond sarcasm: intonation and context as relational cues in children's recognition of irony. In: Proceedings of the 26th Boston University Conference on Language Development. Somerville: Cascadia Press.
- Olkoniemi, H., Häikiö, T., Merinen, M., Manninen, J., Laine, M., and Pexman, P. M. (2025). Learning irony in school: effects of metapragmatic training. *J. Child Lang.* 52, [online] 18 Feb, 1–22. doi: 10.1017/S0305000925000054
- Pexman, P. M. (2008). It's fascinating research. *Curr. Dir. Psychol. Sci.* 17, 286–290. doi: 10.1111/j.1467-8721.2008.00591.x
- Pexman, P. M. (2023). “Irony and thought: developmental insights” in The Cambridge handbook of irony and thought. eds. Gibbs R. W. Jr. and H. L. Colston (Cambridge: Cambridge University Press). (Cambridge Handbooks in Psychology)
- Rattray, C., and Tolmie, A. (2008). Young children's detection and decoding of ironic intonation. *Psychol. Lang. Commun.* 12, 29–54. doi: 10.2478/v10057-008-0002-1
- Rockwell, P. (2000). Lower, slower, louder: vocal cues of sarcasm. *J. Psycholinguist. Res.* 29, 483–495.
- Rockwell, P., and Theriot, E. M. (2001). Culture, gender, and gender mix in encoders of sarcasm: a self-assessment analysis. *Commun. Res. Rep.* 18, 44–52. doi: 10.1080/08824090109384781
- Scott, R. M., and Baillargeon, R. (2009). Which penguin is this? Attributing false beliefs about object identity at 18 months. *Child Dev.* 80, 1172–1196. doi: 10.1111/j.1467-8624.2009.01324.x
- Sperber, D. (1984). Verbal irony: pretense or echoic mention? *J. Exp. Psychol. Gen.* 113, 130–136. doi: 10.1037//0096-3445.113.1.130
- Stiles, D., and Nadler, L. (2013). Sarcasm recognition in children with hearing loss: the role of context and intonation. *J. Educ. Audiol.* 19, 3–11.
- Whalen, J. M., Doyle, A., and Pexman, P. M. (2020). Sarcasm between siblings: children's use of relationship information in processing ironic remarks. *J. Pragmat.* 156, 149–159. doi: 10.1016/j.pragma.2019.05.005
- Wilson, D., and Sperber, D. (2012). “Explaining irony” in Meaning and relevance. eds. D. Wilson and D. Sperber (Cambridge, MA: Cambridge University Press).
- Winner, E., Windmueller, G., Rosenblatt, E., Bosco, L., Best, E., and Gardner, H. (1987). Making sense of literal and nonliteral falsehood. *Metaphor Symbol. Act.* 2, 13–32. doi: 10.1207/s15327868ms0201_2
- Yu, Q., Li, H., Li, S., and Tang, P. (2024). Prosodic and visual cues facilitate irony comprehension by mandarin-speaking children with cochlear implants. *J. Speech Lang. Hear. Res.* 67, 2172–2190. doi: 10.1044/2024_JSLHR-23-00701
- Zajęczkowska, M. K. (2016). Influence of voice intonation on understanding irony by polish-speaking preschool children. *Psychol. Lang. Commun.* 20, 278–291. doi: 10.1515/plc-2016-0017



OPEN ACCESS

EDITED BY

Antonio Bova,
Catholic University of the Sacred Heart, Italy

REVIEWED BY

Ammar A. A. Al Abdely,
University of Anbar, Iraq
Yizhou Lan,
Shenzhen University, China

*CORRESPONDENCE

Hongming Zhu
✉ zhuhongming@hrbu.edu.cn

RECEIVED 24 March 2025

ACCEPTED 08 July 2025

PUBLISHED 17 September 2025

CITATION

Zhu H (2025) Effect of the consonant context on the corner vowel produced by native and Chinese speakers: based on AESOP corpora. *Front. Psychol.* 16:1598904. doi: 10.3389/fpsyg.2025.1598904

COPYRIGHT

© 2025 Zhu. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Effect of the consonant context on the corner vowel produced by native and Chinese speakers: based on AESOP corpora

Hongming Zhu*

School of Foreign Languages, Harbin University, Harbin, China

This study investigated the differences in the consonant context on the four corner vowels (/i/, /u/, /æ/, and /ɑ/) between native and Chinese speakers using the AESOP-ILAS 1 corpus dataset. Ninety-six test subjects with manually adjusted data comprising of 12 native speakers (the control group) and 84 non-native speakers (the targeted group), with 58 Chinese speakers (even gender distribution), 22 Taiwanese speakers (quasi-equal gender distribution), and 4 Hakka-speaking women, were chosen for the analysis. By adopting continuous speech initially, a general comparison of the vowel space for the native and the non-native speakers is presented. Next, the vowel space of the liquid and glide sonorants in the onset and the coda position are compared. The nasal sonorants are compared in their subsections. Finally, the study analyzed the obstrusive sounds and compared the vowel spaces of different pronunciation sites (alveolar, labial, posterior alveolar, and palatal sounds). Since male speakers typically having longer vocal tracts, their vowels may be more centralized or lower in pronunciation compared to females. In order to comprehensively understand the spatial distribution of vowels, it is valuable to further analyze whether there are differences in vowel production patterns between male and female speakers in the native and non-native language groups. Compared to native speakers, Chinese English learners have a larger range of vowel spaces, which may be due to the fact that the corpus is collected from sentences rather than isolated words. Chinese learners exhibit lower F1 and F2 values on corner vowels, with particularly significant differences between /ɑ/ and /u/. These differences are influenced by adjacent phonemes, such as the/w/sound at the beginning and end of syllables. The study suggest that improving the stress distribution of Chinese learners in sentences will significantly enhance their pronunciation level.

KEYWORDS

formant F1 and F2, Chinese, English pronunciation, AESOP ILAS 1, vowel space, consonant context

1 Introduction

Vowels play a crucial role in English pronunciation, especially in Second Language Acquisition (SLA; [Celce-Murcia et al., 1996](#)). Pronunciation issues often arise for Chinese speakers (L1) learning English, especially when producing English vowels ([Flege et al., 1997](#)). Although existing acoustic research has focused on these phonetic challenges, most of them are concentrated in controlled experimental environments, making it difficult to reflect real language usage scenarios. As [Chen et al. \(2001\)](#) pointed out, the vowel space of second language learners is usually more constrained than that of native speakers, which means that their range of vowel sounds and perceptual abilities are relatively limited.

Standard American English is typically described as having around 10–12 distinct vowels (Peterson and Barney, 1952; Ladefoged and Johnson, 2015), which are categorized based on tongue position: front, central, or back. Four key corner vowels—/i/ (high-front), /u/ (high-back), /æ/ (low-front), and /ɑ/ (low-back)—establish the vowel space for a speaker, varying acoustically due to differences in vocal tract anatomy and individual articulation patterns (Hillenbrand et al., 1995). Research in second language (L2) acquisition has shown that the production of these vowels is influenced by multiple factors, including age of acquisition (Flege et al., 1999), first language (L1) phonological transfer (Best and Tyler, 2007), and phonetic training experience (Bradlow et al., 1997). For Chinese English learners, their pronunciation difficulties are not only due to differences in the phoneme system between their native language and the target language (Flege, 1995), but also involve fundamental differences in the acoustic implementation of similar vowels. Although the basic vowel system proposed by Jones (1960) still has reference value, contemporary research emphasizes the reshaping effect of perceptual training and pronunciation learning on the production of second language vowels, especially for phonemes that do not exist in the native language phonology (Escudero, 2005; Munro and Derwing, 2008).

Cardinal vowels serve as reference points for vowel articulation, with each language's vowels mapped relative to these idealized positions. However, the exact proximity of these vowels in both English and Chinese may differ. For example, the /i/ sound in “heed” is closer to cardinal vowel 1, which may not align perfectly with the Chinese /i/ in “?” (meaning: clothe). This discrepancy highlights the challenges faced by Chinese learners, who may produce vowels with significantly different acoustic properties from native speakers, even when articulating the same sounds. Understanding these acoustic differences is essential for improving pronunciation accuracy.

Although early research focused mainly on experimental data, there is still a research gap regarding the accuracy of English pronunciation by Chinese learners, especially in terms of pronunciation performance in natural language streams. The acoustic and pronunciation characteristics of vowels are influenced by multiple linguistic factors, including the phonological environment, speed, and production context of vowels. Among them, adjacent consonants (especially consonants before and after the target vowel) will significantly alter their spectral characteristics. For instance, Stevens and House (1963) demonstrated that consonant context can significantly alter the formant values of vowels, with F2 shifts varying based on the surrounding consonants. Further research by Hillenbrand et al. (2001) confirmed the influence of phonetic environments on vowel identification, though the role of consonants in vowel intelligibility is less pronounced.

Given these findings, it is crucial to examine the acoustic differences in vowel production between native and non-native speakers. This study focuses on the pronunciation of the four corner vowels (/i/, /u/, /æ/, and /ɑ/) by 488 Taiwanese Chinese (L1) speakers from the AESOP-ILAS (Asian English Speech Corpus Project) dataset. Specifically, the study analyzes first formant (F1) and second formant (F2) frequencies of these vowels produced

by 96 test subjects, including 12 native English speakers (the control group) and 84 non-native speakers (the target group). The results are compared across different consonant contexts, including sonorants, nasals, and obstruents, as well as the place of articulation (alveolar, labial, post-alveolar, and velar) to better understand the effects of consonantal environment on vowel production.

The research questions of this study are as follows:

1. Compared with the vowels produced by American native English speakers, how different are the formant F1 and F2 values produced by Chinese learners?
2. How does the vowel space differ in size and shape for the Chinese learners as compared with the Americans?
3. To what extent does the phonological environment influence the production of the target monophthongs?

2 Methodology

2.1 Data description

This study is based on the AESOP-ILAS (Asian English Speech Corpus Project—Institute of Linguistics, Academia Sinica) corpus, which is composed of the speech of Taiwanese learners of English. The data is separated into two parts: AESOP-ILAS 1 and AESOP-ILAS 2. The AESOP-ILAS 1 corpus data is adopted in this study as it contains a more diverse dataset. This enables better separation of segmental and suprasegmental characteristics. AESOP-ILAS 1 is 8.64 GB in size and has 500 hours of speech recordings, including L1 English speech data by 12 American English native speakers and L2 English speech by 488 Taiwanese Chinese speakers coming from 12 universities or institutes located in Taiwan. The recording time of each speaker is approximately 1 hour. The recording time of each L1 speaker is approximately 5.25 and 8.7 hours for each L2 speaker.

2.2 Test subjects

The dataset has 12 native English speakers and 488 non-native speakers. For the native speakers, the dataset is evenly distributed with six males and six females. As for the non-native speakers, 58 spoke Chinese, with a nearly even distribution between men and women; 22 spoke Taiwanese, with a quasi-equal number of men and women; and four women spoke Hakka. Since nearly 70% of the non-native speakers are Chinese speakers, the background language will not influence the vowel production. Most of the non-native men, 15 in total, had between 5 and 10 years of English experience; 8 had between 10 and 15 years of experience; 6 had between 15 and 20 years of experience; and 1 had more than 20 years of experience. As for women, 13 had between 5 and 10 years of experience, 13 had between 10 and 15 years of experience, and 2 had between 15 and 20 years of experience.

All recordings were made in a quiet room such as a classroom or in the office of the instructor, using a Sennheiser PC 155 headset with a unidirectional microphone. The capture and digitalization were made with the TWNAESOP Recording software from the

CUHK-SIAT (the Chinese University of Hong Kong and Shenzhen Institutes of Advanced Technology). The labels were automatically generated using the Hidden Markov Toolkit and some of them were manually adjusted.

Among the 500 subjects in the AESOP1 database, 96 had manually adjusted data, comprising of 12 native speakers, who are referred to as the control group, and 84 non-native speakers, who are referred to as the targeted group. Each of the 96 test subjects was asked to finish seven reading tasks and one picture description task. With the exception of the last picture description task, all of the seven reading tasks have been transcribed. Consequently, our four target corner vowels /i/, /u/, /æ/, and /a/ are extracted from the sentences used in these seven tasks.

The recording of AESOP-ILAS 1 was conducted from 2009 to 2012 and it contains eight recorded tasks comprising six elicited read speech tasks, one fully aided computer-prompted dialogue task, and one partially aided picture description task. All of the transcriptions of the recordings are presented in [Appendices A–G](#), where an asterisk (*) indicates that there are no corner vowels appear in that sentence.

2.3 Formant measurements

The four targeted corner vowels are compared based on the averaged formant (F1 and F2) values from the control group (L1) with that of the target group (L2). Instead of comparing the formants (F1 and F2) at the middle of the target vowel as in [Chen \(1999\)](#), this study proposes to measure them at three locations: the start, middle, and end portion of the vowel and then take their average. These targeted corner vowels were obtained from the sentences in tasks 1–7. For example, the /a/ from the word “Apartment” [əpɑːtmənt], “supermarket” [supəˈmɑːkɪt], and so on. From the dataset, 4,982 tokens for the /a/ sound, 7,753 tokens for the /æ/ sound, 11,803 tokens for the /i/ sound, and 6,806 tokens for the /u/ sound were extracted. From these tokens, the relevant formant values are extracted.

The procedure is based on using Praat Software, version 6.2.19, running on a Microsoft Windows PC. It is as follows: for a given speaker, the audio sample file for the task was loaded together with a labeled file containing the time segments of each phoneme. This allowed the localization of the desired corner vowel in the

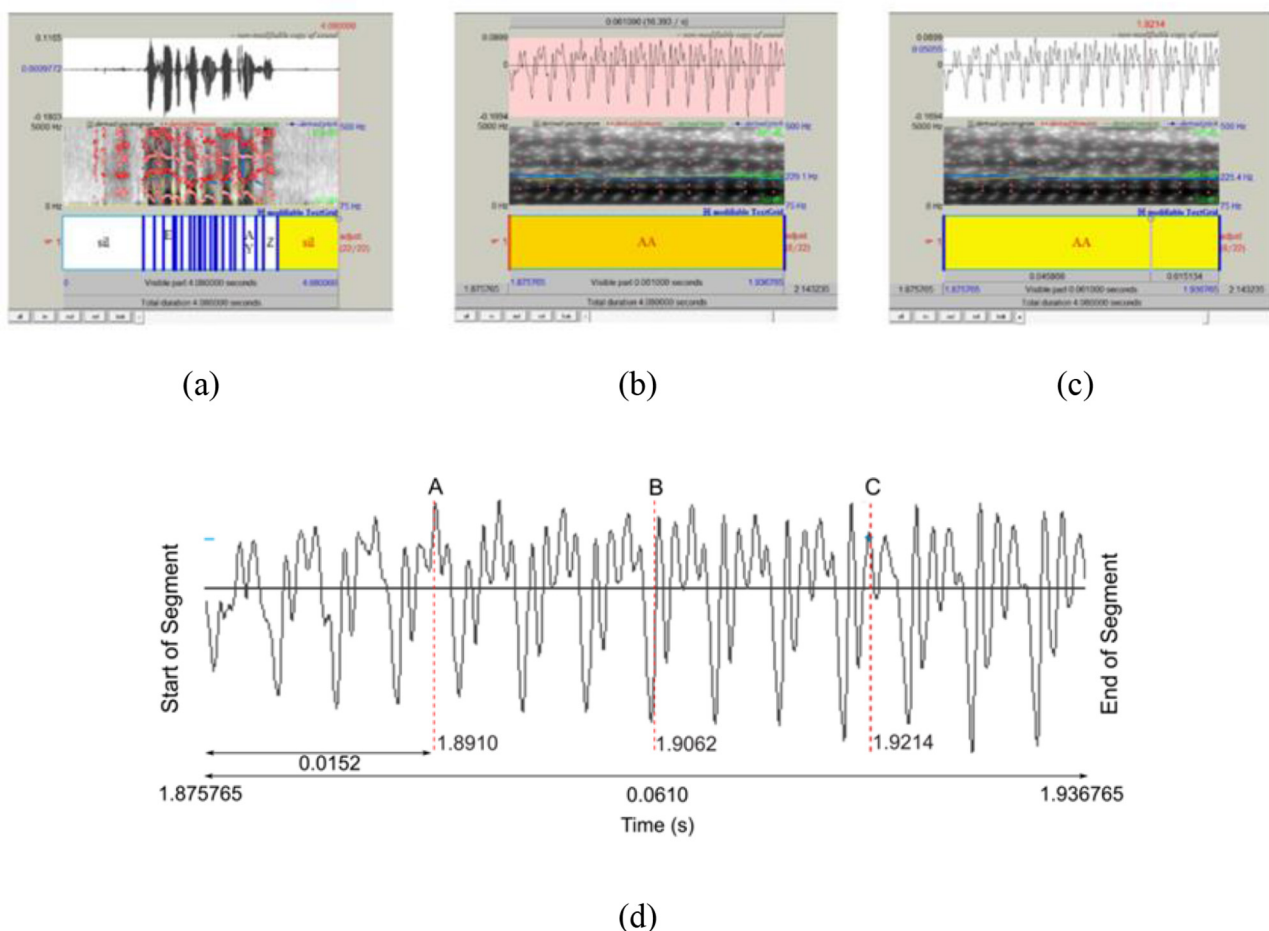
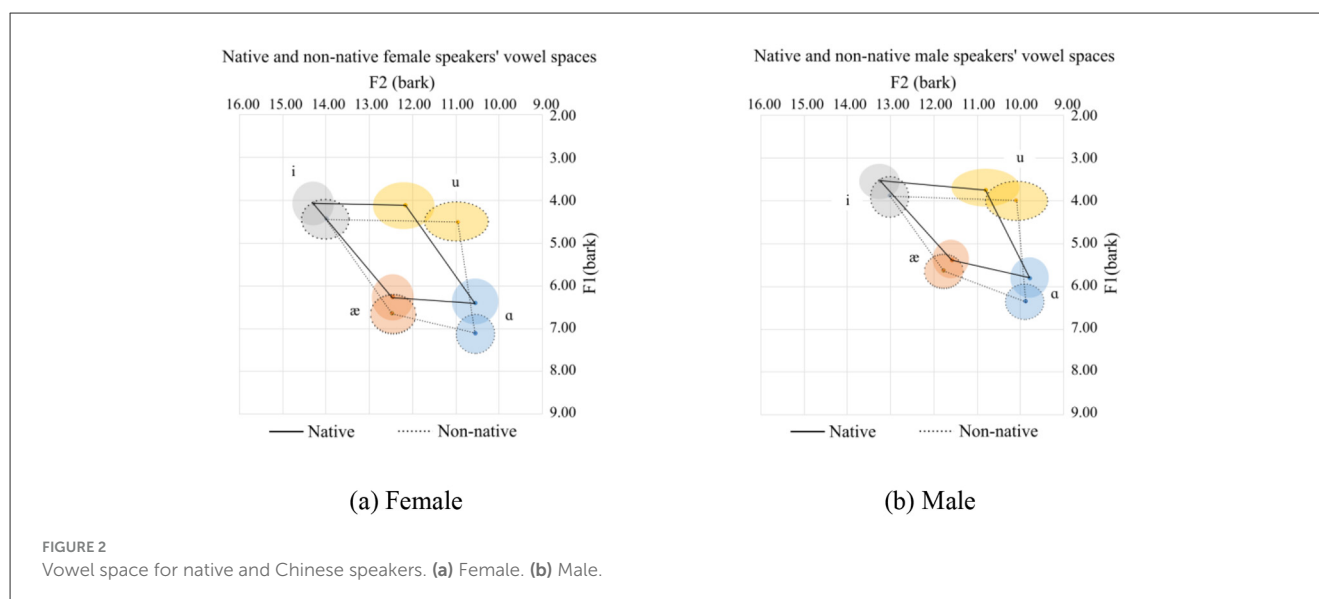


FIGURE 1

A visualization of capturing a vowel. (a) Visual representation of the sound making up the speech; (b) selection of the desired vowel, /a/, sound; (c) extraction of the formants F1 and F2 at a particular point in time; (d) extraction of Formant Frequencies.



audio segment, as shown in Figure 1, where the top window shows the waveform, the middle window shows the spectrogram, and the lower window shows the time intervals of the occurrence of the phonemes. Once such a segment is identified, in this case, /a/, from the sentence “I said apartment five times,” which begins at 1.875765 second (s) and ends at 1.936765 s, the time duration (0.0610 s) is sliced into four equal intervals of 0.0152 s ($0.0610/4$) to obtain three sample points (A at 1.8910 s, B at 1.9062 s, and C at 1.9214 s) as shown in Figure 2. Using linear predictive coding (LPC), the formant (F1 and F2) values at these three locations are computed by the software based on the spectra. These three pairs of formant values are then averaged to produce a single pair of formant (F1 and F2) values for that particular corner vowel. This average value gives a more representative value as it takes into consideration the transition (high to low) at the start of the vowel, the steady state, and the final transition (high to low) at the end of the phoneme. By so doing, any hesitation or on-the-fly correction in the pronunciation of the corner vowel can be detected. This procedure is repeated for any further occurrences of that particular vowel in the speech segment.

The same set of procedures is carried out again for the other corner vowels of interest and is repeated for all the participants. With this, the core data set of this study is obtained and it contains 31,344 distinct data points.

3 Results and discussion

3.1 General view of the corner vowels production for native and Chinese speakers

The vowel space for native and Chinese females is shown in Figure 2a in the bark scale. Since the tongue height is inversely related to the F1, the lower the value, the higher the tongue position. Regarding the F2, the higher the value, the more fronted the tongue position. In this regard, as compared with their native counterparts, Chinese female speakers tend to use a lower and more pulled-back

tongue position relative to the palate. This difference in the height of the tongue is nearly similar for the high vowels (/i/ and /u/) and slightly smaller for the lower front vowels (/æ/). As for the lower back vowel, /a/, however, is pronounced at a much lower tongue position. As for the advancement of the tongue, which can be interpreted from the F2 value, the high vowels are pronounced with a more pulled-back tongue position. The pull-back is more for /u/ than for the high vowels. As for the low vowels (/æ/ and /a/), they are pronounced with the same advancement of the tongue.

The following observations can be made about the Chinese speakers:

1. For /i/, Chinese speakers tend to pronounce it with a lower, more open lip and a more pulled-back tongue position.
2. For /u/, Chinese speakers tend to pronounce it with a lower and more pulled-back tongue position than their native counterparts. Additionally, there are more variations in the F2 for /u/ in native and Chinese speakers than for the other corner vowels.
3. For /æ/, the tongue position is lower with a more open mouth position.
4. For /a/, the tongue is in a lower position with more rounded lips.

The mean and standard deviations of the formants F1 and F2 in hertz, together with the number of speakers and tokens collected for the females and males, are summarized in Table 1. For the males, a similar observation to the females can be observed.

Out of the 849 tokens with the /u/ sound by the native speakers, 287 tokens begin with the /t/ sound, and 334 tokens begin with the /j/ sound. Among the 849 tokens for the /u/ sound in the coda position, only 24 tokens are followed by a silent sound, and an /l/ or /s/ sound follows the rest. Vowels like /i/, /u/, and /a/ are mostly used without a coda, and usually, the consonant sounds /l/ and /s/ are on the onset position of the following word. For example, the /l/ sound comes from the word “learn” in the sentence: “if you want to learn Vietnamese, I think it will be easier than Japanese” (Appendix C Record 2013). Among the 287 tokens that

TABLE 1 F1 and F2 for native and Chinese speakers.

		Native								Chinese							
Female	C. Vowel	ɑ		æ		i		u		ɑ		æ		i		u	
	Formant	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2
	Mean	690	1414	672	1,888	405	2,478	412	1,832	787	1,407	720	1,896	444	2,385	453	1,520
	SD	153	264	154	282	1,235	311	146	395	137	217	125	291	119	360	118	349
	#Speakers	6		6		6		6		44		44		44		44	
	#Tokens	310		486		753		423		2,288		3,545		5,380		3,125	
P.value	ɑ	F1: p.value: 0.000 Cohen's d: −0.695; F2: p.value: 0.000 Cohen's d: 0.032															
	æ	F1: p.value: 0.000 Cohen's d: −0.037; F2: p.value: 0.000 Cohen's d: −0.027															
	i	F1: p.value: 0.000 Cohen's d: −0.087; F2: p.value: 0.000 Cohen's d: 0.26															
	u	F1: p.value: 0.000 Cohen's d: −0.336; F2: p.value: 0.000 Cohen's d: 0.879															
Male	C. Vowel	ɑ		æ		i		u		ɑ		æ		i		u	
	Formant	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2
	Mean	609	1,244	555	1,646	346	2,121	369	1,484	678	1,264	585	1,698	383	2,051	396	1,327
	SD	118	179	110	190	101	262	110	349	107	188	97	213	121	244	119	308
	#Speakers	6		6		6		6		40		40		40		40	
	#Tokens	313		486		754		426		2,071		3,236		4,916		2,834	
P.value	ɑ	F1: p.value: 0.000 Cohen's d: −0.63; F2: p.value: 0.066 Cohen's d: −0.10															
	æ	F1: p.value: 0.000 Cohen's d: −0.30; F2: p.value: 0.000 Cohen's d: −0.24															
	i	F1: p.value: 0.000 Cohen's d: −0.31; F2: p.value: 0.000 Cohen's d: 0.28															
	u	F1: p.value: 0.000 Cohen's d: 0.10; F2: p.value: 0.000 Cohen's d: 0.50															

begin with the T sound, it is quite possible that the carrier word is the preposition word “to.” Similarly, among the 334 tokens that begin with a /j/ sound, the majority are for the word “you,” and few are for “January” (36 tokens), “computer” (11 tokens), “usually” (12 tokens), and “disputing” (12 tokens). The native speakers pronounce the vowel correctly by not stressing the preposition “to”; consequently, they pronounce it as a reduced vowel like a schwa. This effect can be observed in the following sentences: Did he go to the hospital? (Appendix C Record 2001), He had no trouble learning how to make a video (Appendix C Record 2011), and If you want to learn Vietnamese (Appendix C Record 2013). This is also done for words in the infinitive form. As for the 263 tokens with the pronunciation of “you,” out of the 344 tokens with the Y sound, only 12 “you’s” should be emphasized according to the text: I can run faster than you can. Expect for this sentence, all the 249 other “you’s” should not be emphasized. The native speakers know that when “you” is used as a pronoun or as a functional word, it must not be emphasized.

When used as a function word, both “to” and “you” are pronounced with reduced strength by the native speakers, resulting in the /u/ sounding like schwa, which will put the tongue in a much more fronted position. In contrast, the Chinese speakers put the same stress on the syllables when pronouncing “to” and “you,” which pulls back and lowers the tongue to pronounce the/u/sound.

The pronunciation of the /u/ sound in the following subsections will likely follow the same general pattern.

3.2 Effect of the consonant context on the corner vowel

3.2.1 Sonorants

Sonorants are a group of speech sounds produced with a relatively open vocal tract, allowing air to flow freely. These sounds include glide, nasals, and liquids. Since most vowels are produced with the consonants in the onset or coda positions, the consonantal effects on the vowels will be examined. The sonorants have been split into the liquids and the nasals. The nasals are treated separately because the vowels are significantly affected by nasalization. These create anti-formants, which reduce the intensity of the formants, blurring the peak of formant frequencies, which results in an erroneous calculation in formant tracking.

3.2.1.1 Sonorants in the onset position

In this case, the liquid consonants /r/ and /l/ and the glide consonants /j/ and /w/ were included while the nasal ones were excluded. It can be seen from the Figures 3a, b that besides the /ɑ/ sound, the other three corner vowels /i/, /æ/, and /u/, almost follow the general trend. The statistics of the formants are given in Table 2. One possible explanation for the deviation in the pronunciation of /ɑ/ is as follows:

Among the four sonorants (/r/, /l/, /j/, /w/) on the onset position of the /ɑ/ sound, only the /w/ sound appears in this case. The /w/ sound in the onset position can affect the

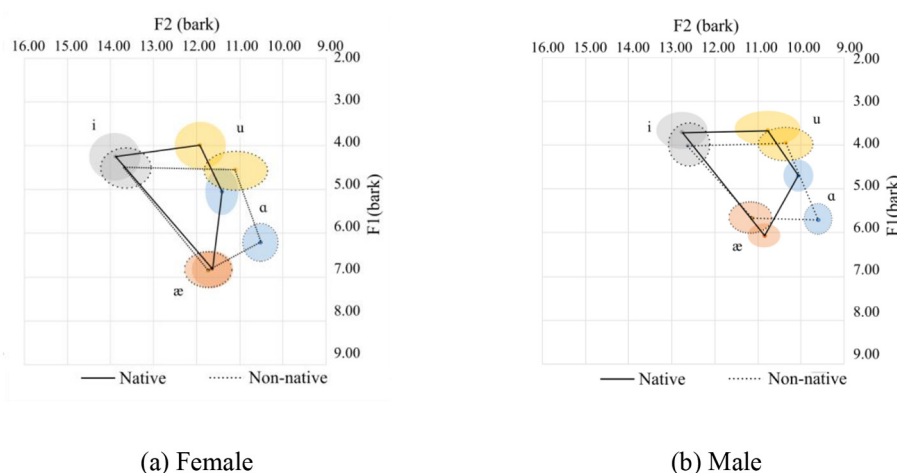


FIGURE 3
Vowel spaces for onset sonorants. (a) Female. (b) Male.

production of the / α / sound as /w/ is a labial-velar sound, which involves rounding the lips and raising the tongue, similar to the pronunciation of velars. When /w/ is followed by an / α /, which is a low vowel, the lips tend to remain rounded, meanwhile, the tongue's elevation is still at a velar sound position, resulting in a more rounded or protruded lip position for the / α / sound. This tongue position results in lower F1 and F2 values. We can also see from the statistics that the pronunciation of the corner variables / α / and /i/ are relatively similar among the native and non-native females. As for the males, only the F2 for /i/ is relatively similar.

3.2.1.2 Sonorant in the coda position

The liquid consonants /r/ (12 tokens) and /l/ (86 tokens) and the glide consonants /j/ (48 tokens) and /w/ (87 tokens) were considered while the nasal sounds were excluded. The formant F1 and F2 values in Hertz are given in Table 3.

In general, Chinese female speakers tend to pronounce the corner vowels with a more frontal tongue position at a lower tongue height, except for / α /, which is pronounced at a slightly higher position. Additionally, the vowel space for the Chinese speakers is larger than that of the native speakers, and the distance between /i/ and /u/ for the native speakers is shorter than their Chinese counterparts. This is primarily due to the pronunciation of /i/. Of the 126 tokens from native speakers for male and female, the majority (102 tokens) end with sonorants /w/, and /l/ and /r/ sounds are 12 tokens for each. The /w/ sound in the coda position can affect the production of the /i/ sound as /w/ is a labial-velar sound, which involves rounding the lips and raising the tongue, similar to the pronunciation of velars. When /i/ is followed by a /w/ sound, which is a low vowel, the lips tend to be prepared for a rounded position. Meanwhile, the tongue's position is relatively posterior, resulting in a lower F1 and F2 value. The tongue position can be visualized in Figures 4a, b for the females and males respectively. The statistics are provided in Table 4, where it can be seen that there is a small Cohen's d value

for F2 for / α / and for /i/ for the females and F1 for /i/ for the males.

3.2.2 Nasals

Nasal sounds are speech sounds produced by allowing air to flow through the nose as well as the mouth. The pronunciation is achieved by lowering the soft palate (or velum) at the back of the mouth, allowing air to pass through the nasal cavity while also passing through the mouth. The resulting sound is often described as having a nasal quality. The nasalization of the sound produced can significantly affect the formant F1 and F2 values due to the formation of anti-formants. Hence the nasal sonorants are considered here.

3.2.2.1 Nasals in the onset position

In this case, the consonants /n/, /m/, and /ŋ/ were included. In general, Chinese speakers tend to pronounce the upper corner vowels /i/ and /u/ with a more pulled-back and at a lower tongue height position. Additionally, the tongue height for /i/ and /u/ is similar. As for the low corner vowels / α / and / α /, / α / is similar to the native's pronunciation, with a slightly more frontal pronunciation in the females' pronunciation, while it is different for the males' pronunciation, which is much more frontal. On the other hand, / α / is quite different as it is pronounced with a more frontal and lower tongue position than the female native speakers. The vowel space for the Chinese speakers is slightly smaller than their native counterparts for the sonorant nasal coda.

Although the two graphs, Figures 5a, b, show an inconsistent pattern between the female and the male speakers, they follow the general pattern, as discussed in Section 3.1. The statistics are summarized in Table 4.

3.2.2.2 Nasals in the coda position

It can be observed from Figures 6a, b that the vowel space for the Chinese speakers is much larger than that of the native-speakers. Except for /i/, the Chinese female speakers pronounce all the other corner vowels with a more posterior tongue. It is

TABLE 2 The F1 and F2 by native and Chinese speakers for onset sonorants.

		Native								Chinese							
Female	C. Vowel	ɑ		æ		i		u		ɑ		æ		i		u	
	Formant	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2
	Mean	518	1,606	741	1,668	426	2,346	394	1,773	661	1,399	746	1,700	452	2,275	455	1,547
	SD	135	191	117	261	137	363	102	419	119	186	118	296	121	372	100	318
	#Speakers	6		6		6		6		44		44		44		44	
	#Tokens	53		66		180		184		401		472		1,315		1,355	
P.value	ɑ	F1: p.value: 0.000 Cohen's d: −1.18; F2: p.value: 0.000 Cohen's d: 1.10															
	æ	F1: p.value: 0.751 Cohen's d: −0.04; F2: p.value: 0.401 Cohen's d: −0.10															
	i	F1: p.value: 0.001 Cohen's d: −0.21; F2: p.value: 0.018 Cohen's d: 0.19															
	u	F1: p.value: 0.000 Cohen's d: −0.60; F2: p.value: 0.000 Cohen's d: 0.68															
Male	C. Vowel	ɑ		æ		i		u		ɑ		æ		i		u	
	Formant	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2
	Mean	473	1,296	639	1,469	363	1,984	359	1,480	594	1,205	590	1,544	399	1,939	390	1,376
	SD	88	139	85	180	104	311	106	354	90	132	92	235	127	270	101	280
	#Speakers	6		6		6		6		40		40		40		40	
	#Tokens	56		66		180		186		360		437		1,192		1,231	
P.value	ɑ	F1: p.value: 0.000 Cohen's d: −1.34; F2: p.value: 0.000 Cohen's d: 0.68															
	æ	F1: p.value: 0.000 Cohen's d: 0.53; F2: p.value: 0.000 Cohen's d: −0.32															
	i	F1: p.value: 0.000 Cohen's d: −0.28; F2: p.value: 0.049 Cohen's d: 0.16															
	u	F1: p.value: 0.000 Cohen's d: −0.30; F2: p.value: 0.000 Cohen's d: 0.35															

also observed that the tongue' elevation is lower for the Chinese speakers. /u/ is also pronounced with a much more posterior tongue position by the Chinese speakers.

The three corner vowels /i/, /u/, and /ɑ/ follow a similar pattern except for /æ/, for which the native speakers pronounce with a conspicuous higher tongue position with more rounded lips than the Chinese speakers. A possible reason is as follows:

Of the 972 tokens for the native speakers' /æ/ sound, 300 tokens end with the sonorant nasals (/n/, /m/, and /ŋ/). Among these, 288 tokens belong to the alveolar sound /n/, and the 12 remaining tokens belong to the /ŋ/ sound. There are no tokens for the /m/ sound. So, generally speaking, after pronouncing the /æ/ sound, the tongue will be raised in anticipation of the nasal /n/ sound, which is an alveolar nasal sound. Naturally, the tongue will move closer to the alveolar ridge area, lowering the F1 value. The statistics of the data are summarized in Table 5 where we can see the strong similarity in the pronunciation of /ɑ/ for the males.

Generally, an anti-formant effect is likely to occur whenever there is a nasal sound after a vowel. The nasal cavity is open when the vowel sound is pronounced, and the velum will go down earlier than expected to prepare the following nasal sound. However, the current corpus does not show that case.

3.2.3 Obstruents

The obstruents are speech sounds that involve a more obstructed vocal tract than the sonorants, resulting in a burst

of sound or friction when produced, such as stops, fricatives, and affricates.

3.2.3.1 Obstruent in the onset position

In this case, the consonants /t/ /s/ /d/ /z/ /p/ /b/ /v/ /f/ /θ/ /ð*/ /ʃ/ /ʒ/ and /k/ were included. As can be seen from Figure 7a, the vowel space for the Chinese female speakers is significantly larger than that produced by the native female speakers. The Chinese female speakers pronounce the upper corner vowels, /i/ and /u/, with a slightly more posterior tongue position for /i/ and a much more posterior position for /u/. As for the lower corner vowels, these are pronounced with a slightly more frontally and in a lower tongue position than the natives. There is a partial overlap in the pronunciation of all three corner vowels, /ɑ/, /æ/, and /i/, by the native and Chinese speakers and no overlap in the pronunciation of /u/. A similar trend can be observed for the male speakers.

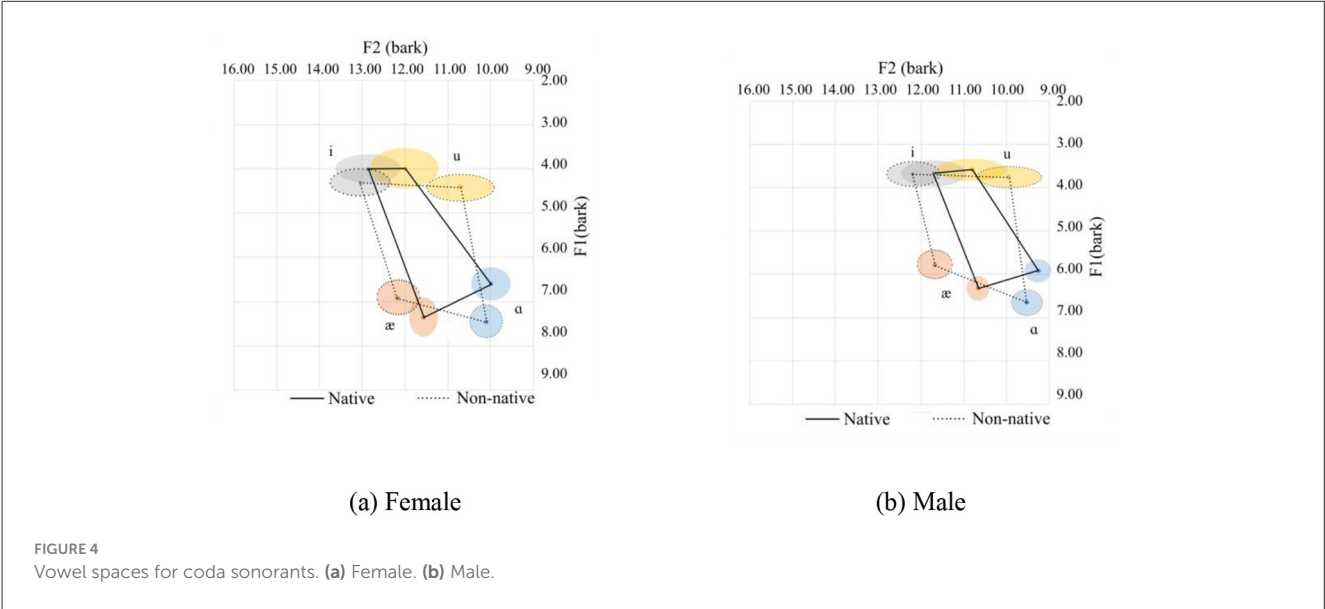
Although the difference between the female and the male speakers can be seen easily in Figures 7a, b, the observed pattern for each is consistent with the general pattern shown in Figure 5. This consonant environment (obstruent in the onset position) does not affect the production of the corner vowels beyond that observed in the general trend in Section 3.1. The statistics are summarized in Table 6 below.

3.2.3.2 Obstruent in the coda position

In this case, the consonants /t/ /s/ /d/ /z/ /p/ /b/ /v/ /f/ /θ/ /ð*/ /ʃ/ /ʒ/ and /k/ were included. The vowel space for the Chinese female speakers is significantly larger than that produced by the

TABLE 3 The F1 and F2 by native and Chinese speakers for coda sonorants.

		Native								Chinese							
Female	C. Vowel	ɑ		æ		i		u		ɑ		æ		i		u	
	Formant	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2
	Mean	712	1,288	822	1,640	396	2,033	396	1,792	836	1,307	756	1,812	429	2,084	440	1,466
	SD	109	206	136	169	72	429	112	437	111	164	110	260	76	424	75	370
	#Speakers	6		6		6		6		44		44		44		44	
	#Tokens	132		30		64		115		969		220		475		827	
P.value	ɑ	F1: p.value: 0.000 Cohen's d: −1.11; F2: p.value: 0.231 Cohen's d: −0.11															
	æ	F1: p.value: 0.000 Cohen's d: 0.58; F2: p.value: 0.000 Cohen's d: −0.68															
	i	F1: p.value: 0.000 Cohen's d: −0.43; F2: p.value: 0.363 Cohen's d: −0.12															
	u	F1: p.value: 0.000 Cohen's d: −0.54; F2: p.value: 0.000 Cohen's d: 0.86															
Male	C. Vowel	ɑ		æ		i		u		ɑ		æ		i		u	
	Formant	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2
	Mean	620	1,136	672	1,423	358	1,701	350	1,488	717	1,191	603	1,671	360	1,827	366	1,297
	SD	69	121	76	119	73	360	56	381	83	143	85	203	68	329	57	313
	#Speakers	6		6		6		6		40		40		40		40	
	#Tokens	132		30		62		118		881		200		424		747	
P.value	ɑ	F1: p.value: 0.000 Cohen's d: −1.19; F2: p.value: 0.000 Cohen's d: −0.39															
	æ	F1: p.value: 0.000 Cohen's d: 0.82; F2: p.value: 0.000 Cohen's d: −1.27															
	i	F1: p.value: 0.788 Cohen's d: −0.02; F2: p.value: 0.000 Cohen's d: −0.37															
	u	F1: p.value: 0.000 Cohen's d: −0.28; F2: p.value: 0.000 Cohen's d: 0.59															



native female speakers due mainly to the pronunciation of /u/. The female Chinese speakers pronounce the upper corner vowels with a lower tongue position. The tongue's elevation is similar for both corner vowels, /i/ and /u/, by the Chinese speakers as opposed to a slightly more elevated tongue position for /u/ by the native female speakers. The tongue position is slightly more posterior for /i/ but is significantly more pulled back for /u/ for pronunciations by the Chinese female speakers. For the lower corner vowels, the native

females' tongue position tends to be more elevated but at a similar tongue advancement. In general, the pronunciation of /u/ in the obstruent coda position is more different for the Chinese speakers. These are shown in Figures 8a, b and the statistics are summarized in Table 7.

Based on the dataset, the obstruent in the coda position does not affect the production of the four corner vowels beyond that observed in Section 3.1 is still applicable.

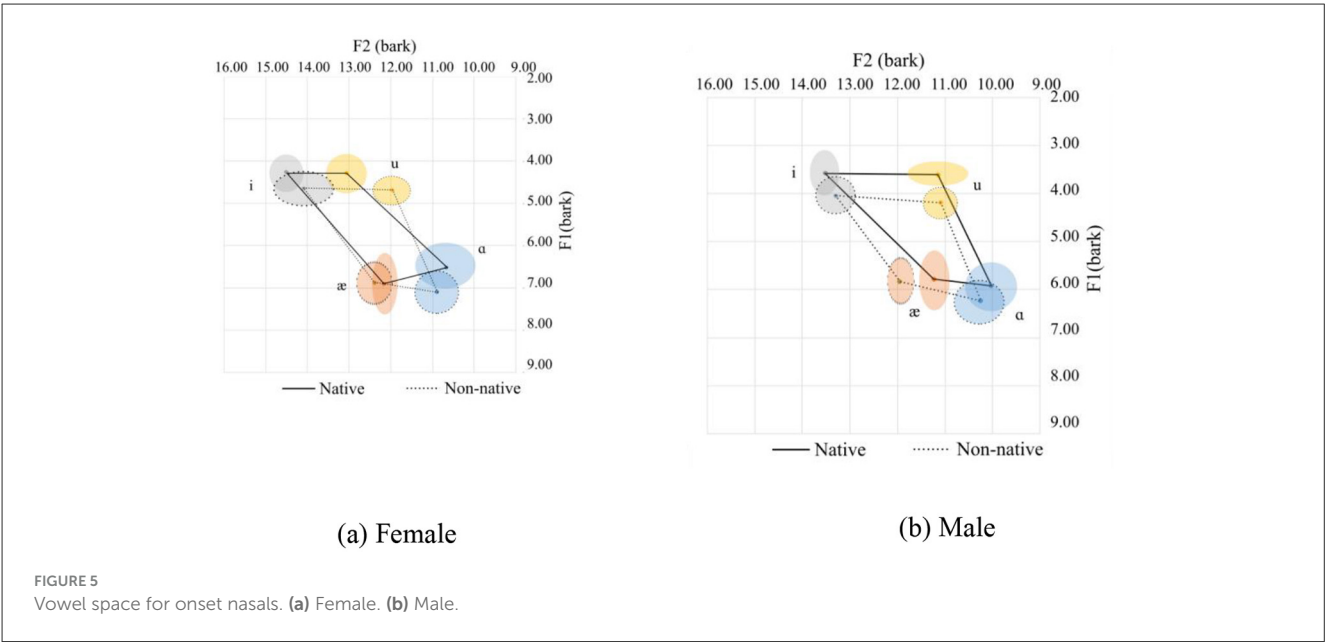


TABLE 4 The F1 and F2 by native and Chinese speakers for onset nasals.

		Native								Chinese							
Female	C. Vowel	ɑ		æ		i		u		ɑ		æ		i		u	
	Formant	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2
	Mean	705	1,453	766	1,800	424	2,559	427	2,071	785	1,490	755	1,866	468	2,431	471	1,753
	SD	147	346	212	167	106	287	108	305	145	251	145	233	104	422	82	243
	#Speakers	6		6		6		6		44		44		44		44	
	#Tokens	95		18		125		30		669		107		923		218	
P.value	ɑ	F1: p.value: 0.000 Cohen's d: −0.55; F2: p.value: 0.000 Cohen's d: −0.13															
	æ	F1: p.value: 0.772 Cohen's d: 0.07; F2: p.value: 0.000 Cohen's d: −0.29															
	i	F1: p.value: 0.000 Cohen's d: −0.42; F2: p.value: 0.000 Cohen's d: 0.31															
	u	F1: p.value: 0.000 Cohen's d: −0.51; F2: p.value: 0.000 Cohen's d: 1.26															
Male	C. Vowel	ɑ		æ		i		u		ɑ		æ		i		u	
	Formant	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2
	Mean	624	1,302	609	1,560	350	2,205	352	1,559	663	1,348	611	1,742	401	2,139	415	1,528
	SD	131	238	155	160	121	202	53	320	120	241	123	175	94	247	75	179
	#Speakers	6		6		6		6		40		40		40		40	
	#Tokens	98		18		126		30		611		114		837		198	
P.value	ɑ	F1: p.value: 0.003 Cohen's d: −0.32; F2: p.value: 0.082 Cohen's d: −0.19															
	æ	F1: p.value: 0.941 Cohen's d: −0.01; F2: p.value: 0.000 Cohen's d: −1.05															
	i	F1: p.value: 0.000 Cohen's d: −0.52; F2: p.value: 0.004 Cohen's d: 0.27															
	u	F1: p.value: 0.000 Cohen's d: −0.86; F2: p.value: 0.436 Cohen's d: 0.15															

3.2.4 Different places of articulations

The place of articulation refers to the specific location where speech sounds are produced in the vocal tract. These locations can be divided into two main categories: oral and nasal. The orals include the alveolar, dental, glottal, labials, palatal, post alveolar and velar. Only the alveolar, labials, post alveolar, and velar are considered.

3.2.4.1 Alveolar in the onset position

The consonants /t/, /s/, /d/ and /z/ were included in this case. It can be observed from Figures 9a, b that the vowel space for the native speakers is smaller than that of the Chinese speakers. Except for /u/, the sounds produced by the native and Chinese speakers overlap. In general, for the upper corner vowels, the tongue position is more pulled-back and is slightly lower for the Chinese speakers.

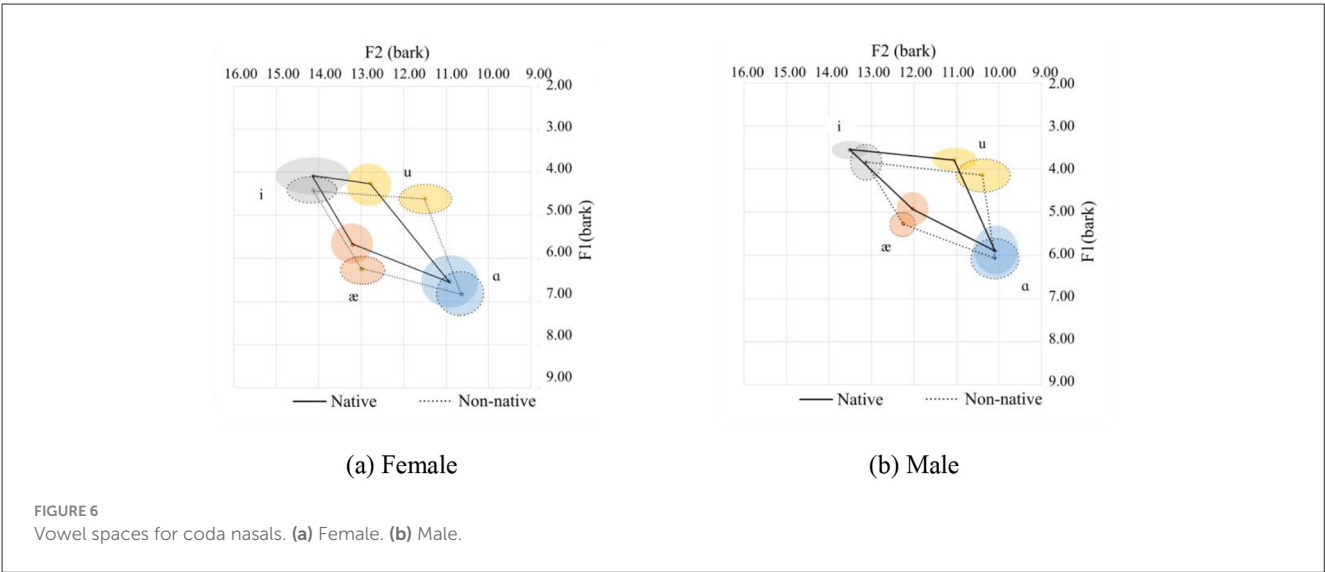


TABLE 5 The F1 and F2 by native and Chinese speakers for coda nasals.

		Native								Chinese							
Female	C. Vowel	ɑ		æ		i		u		ɑ		æ		i		u	
	Formant	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2
	Mean	709	1,507	594	2,110	405	2457	425	1,988	748	1,433	663	2,049	440	2,435	462	1,641
	SD	158	334	118	307	102	470	113	304	145	270	93	296	73	373	81	306
	#Speakers	6		6		6		6		44		44		44		44	
	#Tokens	78		150		46		42		559		1,092		354		302	
P.value	ɑ	F1: p.value: 0.027 Cohen's d: −0.26; F2: p.value: 0.000 Cohen's d: 0.26															
	æ	F1: p.value: 0.000 Cohen's d: −0.71; F2: p.value: 0.017 Cohen's d: 0.20															
	i	F1: p.value: 0.000 Cohen's d: −0.45; F2: p.value: 0.706 Cohen's d: 0.05															
	u	F1: p.value: 0.000 Cohen's d: −0.43; F2: p.value: 0.000 Cohen's d: 1.13															
Male	C. Vowel	ɑ		æ		i		u		ɑ		æ		i		u	
	Formant	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2
	Mean	623	1,314	499	1,765	345	2,206	370	1,522	641	1,318	539	1,819	379	2,082	410	1,385
	SD	148	230	92	205	45	227	57	258	123	260	72	170	121	235	98	277
	#Speakers	6		6		6		6		40		40		40		40	
	#Tokens	78		150		52		42		506		996		329		271	
P.value	ɑ	F1: p.value: 0.234 Cohen's d: −0.14; F2: p.value: 0.890 Cohen's d: −0.01															
	æ	F1: p.value: 0.000 Cohen's d: −0.53; F2: p.value: 0.000 Cohen's d: −0.30															
	i	F1: p.value: 0.000 Cohen's d: −0.29; F2: p.value: 0.000 Cohen's d: 0.53															
	u	F1: p.value: 0.000 Cohen's d: −0.42; F2: p.value: 0.000 Cohen's d: 0.49															

As for the lower corner vowels, Chinese female speakers pronounce them slightly more frontally and in a lower tongue position. The general trend in Section 3.1 can still be applied to the alveolar sounds in the onset position in this dataset. The formant values are summarized in Table 8 together with their statistics.

3.2.4.2 Alveolar in the coda position

The consonants /t/, /s/, /d/ and /z/ were included. It can be seen from Figure 10 that there is a partial overlap in the

sound produced by both groups except for /u/. Besides, the vowel space for the Chinese female speakers is larger. The tongue's elevation is lower for the Chinese females for all corner vowels, and the tongue is more frontal for the lower corner vowels.

Although the difference between females and males can be easily seen in Figures 10a, b, the general pattern of each is consistent with the general pattern shown in Figure 5. Apparently, this consonant environment (Alveolar in the onset position) does not

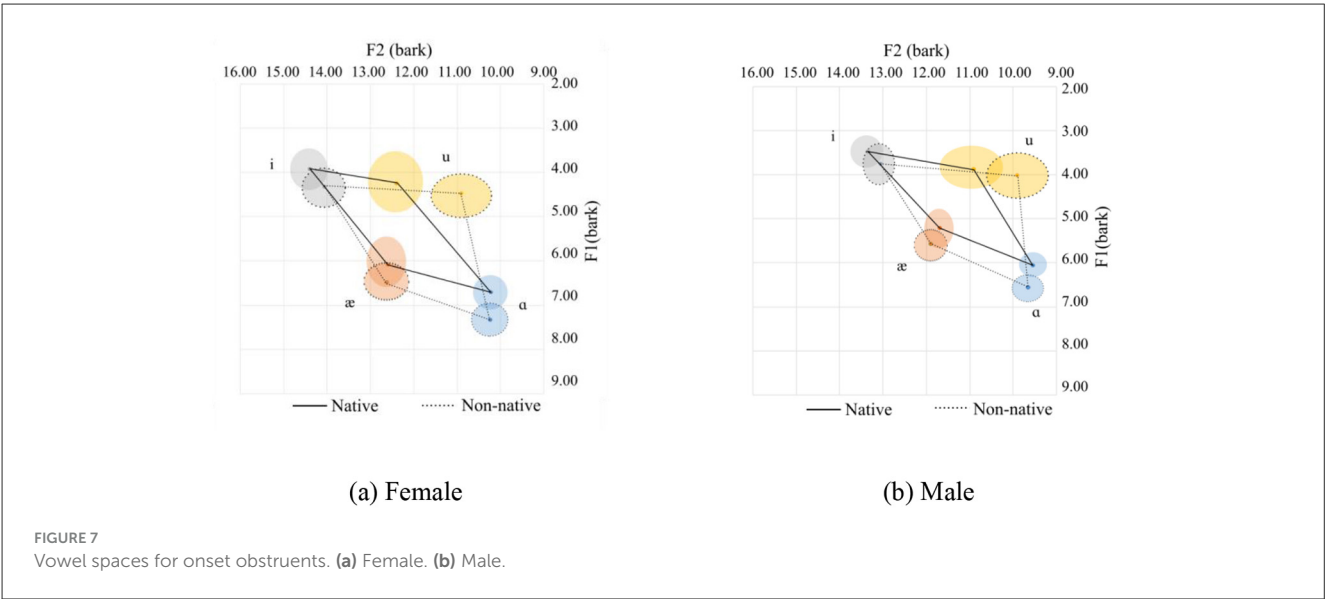


TABLE 6 The F1 and F2 by native and Chinese speakers for onset obstruents.

		Native								Chinese							
Female	C. Vowel	ɑ		æ		i		u		ɑ		æ		i		u	
	Formant	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2
	Mean	727	1,332	648	1,927	389	2,514	431	1,886	817	1,340	698	1,940	431	2,401	450	1,508
	SD	114	173	155	271	117	277	185	357	113	189	122	289	114	329	139	347
	#Speakers	6		6		6		6		44		44		44		44	
	#Tokens	119		298		373		191		879		2,020		2,583		1,417	
P.value	ɑ	F1: p.value: 0.000 Cohen's d: −0.79; F2: p.value: 0.672 Cohen's d: −0.04															
	æ	F1: p.value: 0.000 Cohen's d: −0.39; F2: p.value: 0.481 Cohen's d: −0.04															
	i	F1: p.value: 0.000 Cohen's d: −0.36; F2: p.value: 0.000 Cohen's d: 0.34															
	u	F1: p.value: 0.087 Cohen's d: −0.13; F2: p.value: 0.000 Cohen's d: 1.08															
Male	C. Vowel	ɑ		æ		i		u		ɑ		æ		i		u	
	Formant	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2
	Mean	637	1,192	534	1,675	339	2,155	384	1,510	705	1,218	577	1,729	369	2,059	400	1,287
	SD	75	127	109	176	88	227	121	342	86	148	98	200	122	220	141	313
	#Speakers	6		6		6		6		40		40		40		40	
	#Tokens	120		302		373		192		795		1,884		2,385		1,283	
P.value	ɑ	F1: p.value: 0.000 Cohen's d: −0.80; F2: p.value: 0.078 Cohen's d: −0.17															
	æ	F1: p.value: 0.000 Cohen's d: −0.43; F2: p.value: 0.000 Cohen's d: −0.27															
	i	F1: p.value: 0.000 Cohen's d: −0.25; F2: p.value: 0.000 Cohen's d: 0.43															
	u	F1: p.value: 0.128 Cohen's d: −0.11; F2: p.value: 0.000 Cohen's d: 0.70															

have a significant effect on the production of the corner vowels. The statistics are provided in Table 9 below.

3.2.4.3 Labials in the onset position

The consonants /p/, /b/, /v/, and /f/ were included. In general, the vowel space for Chinese female speakers is slightly larger than that of native female speakers as can be seen from Figure 11a. In addition, the sound produced for the two front corner vowels, /i/

and /æ/, is partially overlapped, while the sound produced for the rear corner vowels, /u/ and /ɑ/, is not overlapped. Moreover, it can be readily observed that the difference between the /u/ sound is relatively small for Chinese speakers. Out of the 849 tokens for the vowel /u/ sound, only 24 tokens for the native speakers belong to the labials in the onset position. These tokens belong to the /f/ sound. And all of the 24 /f/ sounds end with the final /d/ sound, which can be traced to the word “food” in the sentence “I like

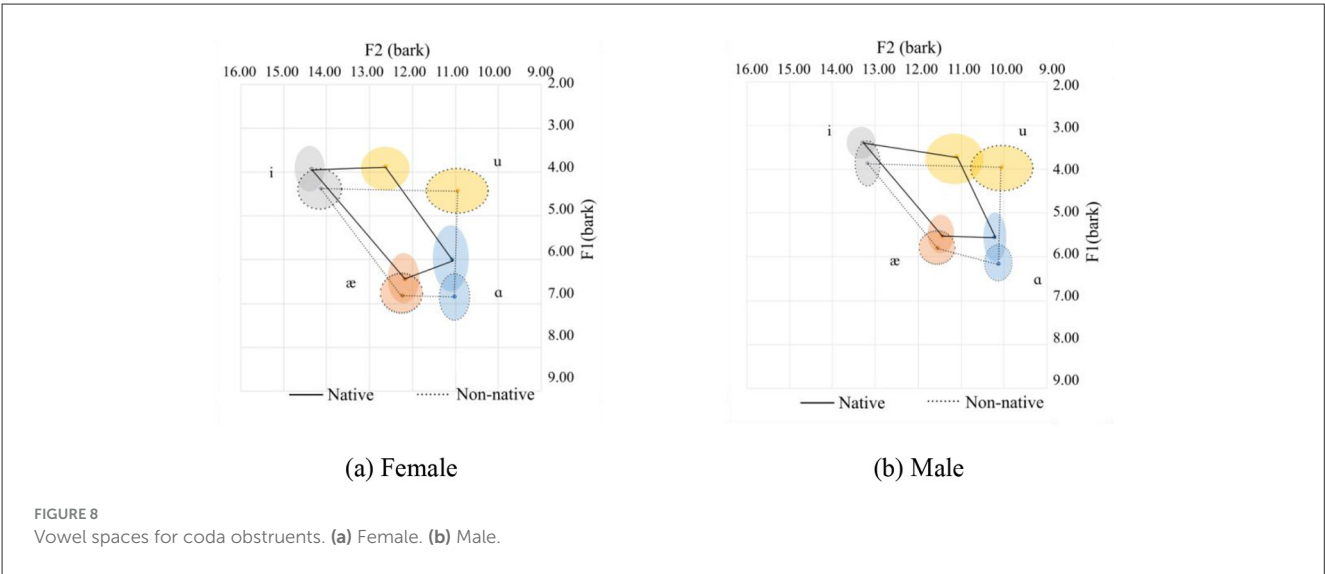


TABLE 7 The F1 and F2 by native and Chinese speakers for coda obstruents.

		Native								Chinese							
Female	C. Vowel	ɑ		æ		i		u		ɑ		æ		i		u	
	Formant	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2
	Mean	642	1,526	695	1,803	389	2,497	384	1,945	751	1,509	745	1,829	438	2,428	445	1,518
	SD	194	197	153	200	123	248	120	314	148	169	130	262	115	315	141	345
	#Speakers	6		6		6		6		44		44		44		44	
	#Tokens	88		306		320		163		667		2,228		2,276		1,202	
P.value	ɑ	F1: p.value: 0.000 Cohen's d: −0.70; F2: p.value: 0.386 Cohen's d: 0.09															
	æ	F1: p.value: 0.000 Cohen's d: −0.37; F2: p.value: 0.093 Cohen's d: −0.10															
	i	F1: p.value: 0.000 Cohen's d: −0.42; F2: p.value: 0.000 Cohen's d: 0.22															
	u	F1: p.value: 0.000 Cohen's d: −0.43; F2: p.value: 0.000 Cohen's d: 1.25															
Male	C. Vowel	ɑ		æ		i		u		ɑ		æ		i		u	
	Formant	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2
	Mean	579	1,331	571	1,610	328	2,130	368	1,549	652	1,311	605	1,642	382	2,093	392	1,319
	SD	146	114	106	150	84	208	149	319	108	138	101	209	127	183	140	314
	#Speakers	6		6		6		6		40		40		40		40	
	#Tokens	91		306		337		165		600		2,038		2,075		1,074	
P.value	ɑ	F1: p.value: 0.000 Cohen's d: −0.64; F2: p.value: 0.185 Cohen's d: 0.14															
	æ	F1: p.value: 0.000 Cohen's d: −0.33; F2: p.value: 0.010 Cohen's d: −0.15															
	i	F1: p.value: 0.000 Cohen's d: −0.44; F2: p.value: 0.001 Cohen's d: 0.19															
	u	F1: p.value: 0.041 Cohen's d: −0.16; F2: p.value: 0.000 Cohen's d: 0.73															

Japanese food, but Korean food is too spicy for me.” Compared with the function words “to” and “you,” “food” is a content word that should be stressed. Consequently, for the female and male native speakers, the /u/ sound has a lower and posterior tongue position which the Chinese speakers also have. The vowel space for the males are given in Figure 11b and the statistics are provided in Table 10.

3.2.4.4 Labials in the coda position

This section contains only three corner vowels including /i/, /u/, and /æ/, as this dataset does not contain the /ɑ/ sound. This explains why the shape of Figures 12a, b is triangular instead of quadrilateral. In general, the vowel space for the native female speakers is significantly smaller than that of the Chinese female speakers. There is a relatively good

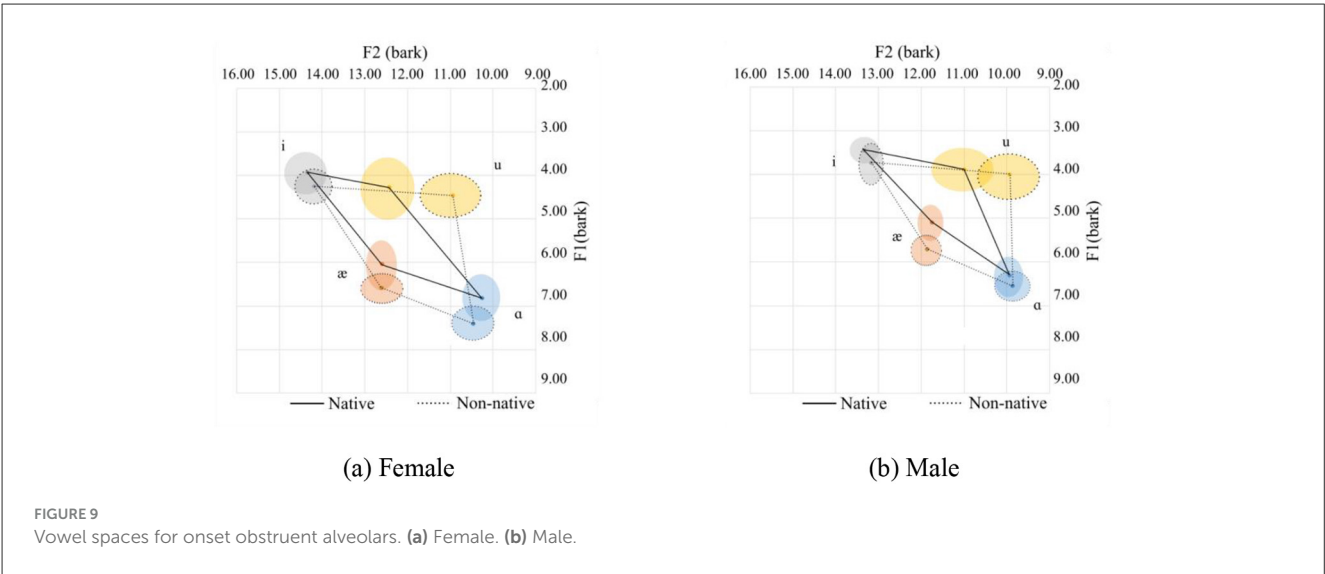


TABLE 8 The F1 and F2 by native and Chinese speakers for onset obstruent alveolars.

		Native								Chinese							
Female	C. Vowel	ɑ		æ		i		u		ɑ		æ		i		u	
	Formant	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2
	Mean	748	1,342	641	1,920	389	2,500	434	1,895	829	1,391	710	1,935	423	2,433	449	1,516
	SD	156	199	157	215	119	294	191	358	119	243	102	274	102	288	142	353
	#Speakers	6		6		6		6		44		44		44		44	
	#Tokens	17		101		174		179		125		608		1,275		1,325	
P.value	ɑ	F1: p.value: 0.000 Cohen's d: −0.65; F2: p.value: 0.418 Cohen's d: −0.20															
	æ	F1: p.value: 0.000 Cohen's d: −0.61; F2: p.value: 0.610 Cohen's d: −0.05															
	i	F1: p.value: 0.000 Cohen's d: −0.32; F2: p.value: 0.016 Cohen's d: 0.23															
	u	F1: p.value: 0.215 Cohen's d: −0.10; F2: p.value: 0.000 Cohen's d: 1.07															
Male	C. Vowel	ɑ		æ		i		u		ɑ		æ		i		u	
	Formant	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2
	Mean	673	1,271	520	1,688	335	2,148	385	1,528	705	1,258	594	1,719	369	2,092	399	1,295
	SD	125	143	101	161	69	215	124	342	96	174	100	189	127	182	144	318
	#Speakers	6		6		6		6		40		40		40		40	
	#Tokens	18		104		181		180		115		589		1,193		1,197	
P.value	ɑ	F1: p.value: 0.000 Cohen's d: −0.31; F2: p.value: 0.776 Cohen's d: 0.07															
	æ	F1: p.value: 0.000 Cohen's d: −0.73; F2: p.value: 0.123 Cohen's d: −0.16															
	i	F1: p.value: 0.000 Cohen's d: −0.28; F2: p.value: 0.000 Cohen's d: 0.30															
	u	F1: p.value: 0.232 Cohen's d: −0.09; F2: p.value: 0.000 Cohen's d: 0.72															

overlap for the corner vowels /i/ and /æ/ and no overlap for /u/. In general, even without the /ɑ/ sound, it can be seen that the pattern of the remaining three corner vowel sounds follows the general trend as in Figure 5. It is worth mentioning that, with obstruent labials in the coda position, it is evident that, for Chinese speakers, the pronunciation of the vowel /æ/ is quite accurate as the centroid of this particular sound is very close to the one for

the native speakers regardless of the gender. The summary of the dataset for this section is provided in Table 11.

3.2.4.5 Post-alveolar in the onset position

The consonants containing /ʒ/ and /ʃ/ are considered for the post-alveolar in the onset position. Only the three corner vowels /ɑ/, /æ/, and /i/ are present.

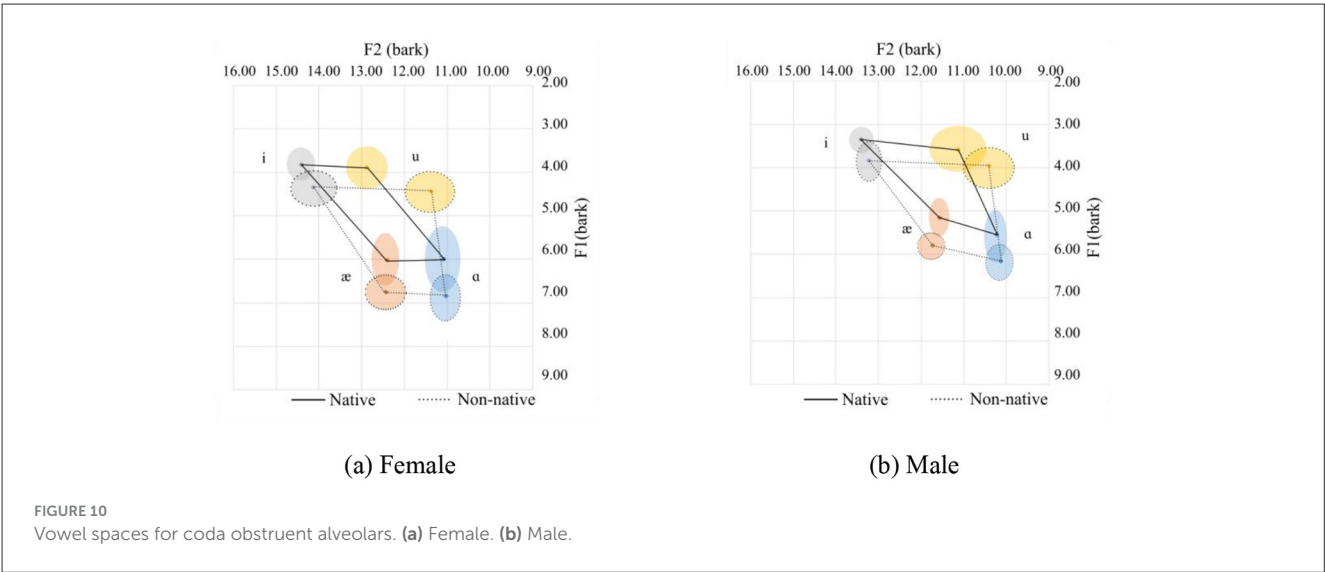


TABLE 9 The F1 and F2 by native and Chinese speakers for coda obstruent alveolars.

		Native								Chinese							
Female	C. Vowel	ɑ		æ		i		u		ɑ		æ		i		u	
	Formant	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2
	Mean	642	1,526	640	1,864	375	2,516	385	2,010	751	1,509	734	1,879	434	2,430	446	1,607
	SD	194	197	158	183	85	237	121	275	148	169	116	256	100	329	126	296
	#Speakers	6		6		6		6		44		44		44		44	
	#Tokens	88		162		192		59		666		1,176		1,387		431	
P.value	ɑ	F1: p.value: 0.000 Cohen's d: −0.70; F2: p.value: 0.385 Cohen's d: 0.09															
	æ	F1: p.value: 0.000 Cohen's d: −0.77; F2: p.value: 0.456 Cohen's d: −0.06															
	i	F1: p.value: 0.000 Cohen's d: −0.60; F2: p.value: 0.000 Cohen's d: 0.26															
	u	F1: p.value: 0.000 Cohen's d: −0.48; F2: p.value: 0.000 Cohen's d: 1.37															
Male	C. Vowel	ɑ		æ		i		u		ɑ		æ		i		u	
	Formant	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2
	Mean	579	1,331	528	1,639	324	2,167	354	1,555	652	1,311	605	1,687	378	2,107	392	1,385
	SD	146	114	106	122	70	183	143	316	108	138	98	192	112	188	122	272
	#Speakers	6		6		6		6		40		40		40		40	
	#Tokens	91		162		198		59		600		1,081		1,266		393	
P.value	ɑ	F1: p.value: 0.000 Cohen's d: −0.64; F2: p.value: 0.185 Cohen's d: 0.14															
	æ	F1: p.value: 0.000 Cohen's d: −0.77; F2: p.value: 0.000 Cohen's d: −0.26															
	i	F1: p.value: 0.000 Cohen's d: −0.50; F2: p.value: 0.000 Cohen's d: 0.32															
	u	F1: p.value: 0.000 Cohen's d: −0.30; F2: p.value: 0.000 Cohen's d: 0.61															

In general, Chinese speakers pronounce the corner vowels for the post-alveolar in a more pulled-back tongue position. From Figures 13a, b, the height of the tongue is similar for /i/ and higher for both /æ/ and /ɑ/, and the vowel space is smaller for the Chinese speakers. In addition, there is a partial overlap in the pronunciation of all the corner vowels between the two groups. This means that the Chinese speakers' overall pronunciation is quite similar to the native speakers except for the /æ/ sound.

It can be seen that the /æ/ sound for the native speakers is a little lower and more fronted than that of the Chinese speakers. A possible explanation is as follows. Out of the 972 tokens for the /æ/ sound for native speakers, only 72 belong to the /ɜ/ sound in the onset post-alveolar position. Among these 72 tokens, 36 end with the /n/ sound, and the remaining 36 end with the /p/ sound. By looking through the corpus, the /p/ sound belongs to the word "Japanese," and the /n/ sound belongs to the word "January." The

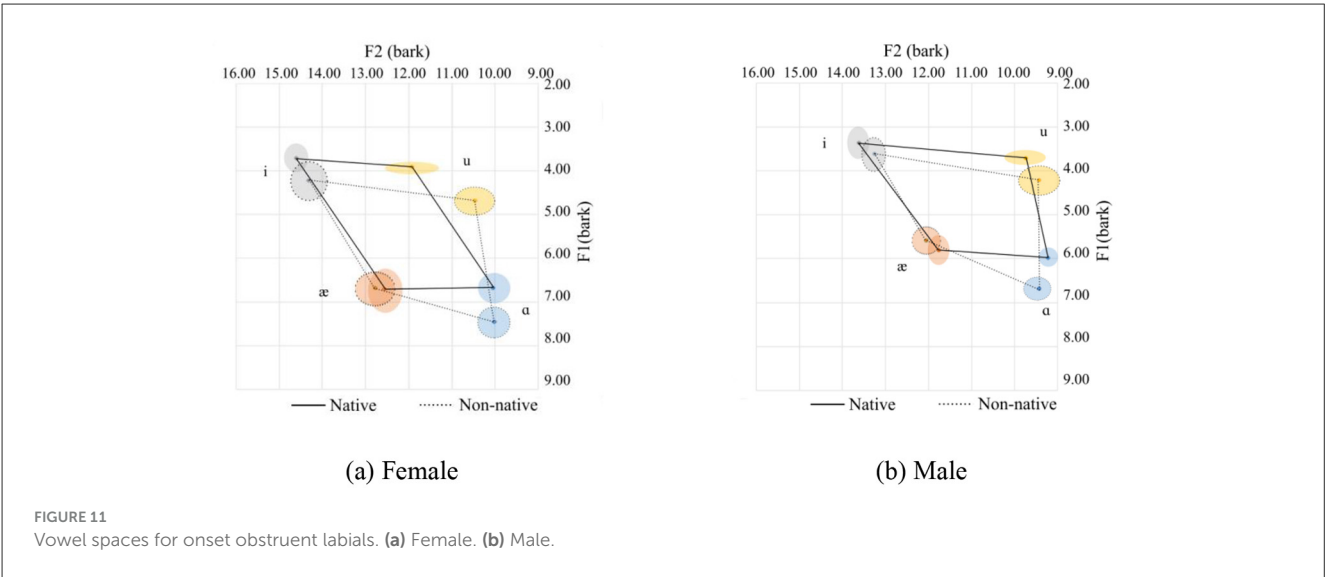


TABLE 10 The F1 and F2 by native and Chinese speakers for onset obstruent labials.

		Native								Chinese							
Female	C. Vowel	ɑ		æ		i		u		ɑ		æ		i		u	
	Formant	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2
	Mean	723	1,296	730	1,916	363	2,587	384	1,751	836	1,289	725	1,980	421	2,493	470	1,390
	SD	102	177	146	239	78	217	31	316	107	162	113	254	113	290	83	211
	#Speakers	6		6		6		6		44		44		44		44	
	#Tokens	72		30		109		12		532		204		785		92	
P.value	ɑ	F1: p.value: 0.000 Cohen's d: −1.06; F2: p.value: 0.737 Cohen's d: 0.04															
	æ	F1: p.value: 0.814 Cohen's d: 0.04; F2: p.value: 0.000 Cohen's d: −0.25															
	i	F1: p.value: 0.000 Cohen's d: −0.53; F2: p.value: 0.000 Cohen's d: 0.33															
	u	F1: p.value: 0.000 Cohen's d: −1.08; F2: p.value: 0.000 Cohen's d: 1.60															
Male	C. Vowel	ɑ		æ		i		u		ɑ		æ		i		u	
	Formant	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2
	Mean	628	1,131	606	1,686	328	2,236	362	1,233	721	1,172	577	1,766	354	2,115	418	1,178
	SD	59	82	89	131	91	172	38	193	76	125	85	171	95	175	90	206
	#Speakers	6		6		6		6		40		40		40		40	
	#Tokens	72		30		108		12		480		193		720		85	
P.value	ɑ	F1: p.value: 0.000 Cohen's d: −1.25; F2: p.value: 0.000 Cohen's d: −0.34															
	æ	F1: p.value: 0.000 Cohen's d: 0.33; F2: p.value: 0.000 Cohen's d: −0.48															
	i	F1: p.value: 0.000 Cohen's d: −0.27; F2: p.value: 0.000 Cohen's d: 0.69															
	u	F1: p.value: 0.000 Cohen's d: −0.65; F2: p.value: 0.000 Cohen's d: 0.26															

native speakers acknowledged that these should be pronounced with strength. The first syllable for January or the second syllable for Japanese is consequently stressed. On the contrary, the Chinese speakers did not pronounce it with stress. A summary of the statistics of the formant values are provided in Table 12.

3.2.4.6 Velars in the coda position

For the velar in the coda position, there is only the consonant /k/, and the vowel space is composed of only three corner vowels

including /æ/, /i/, and /u/. Figures 14a, b together with Table 13 summarizes the data.

In general, the vowel space for Chinese female speakers is significantly larger than that of the native females owing to the pronunciation of /u/. Of the 849 tokens produced by the native speakers for the /u/ sound, 36 tokens end with the velar /k/ sound. Out of these, 12 tokens are for the word “you,” 12 tokens are for the word “to,” and the remaining 12 tokens are for the word “who.” The words “you,” “to,” and “who” server as the functional words

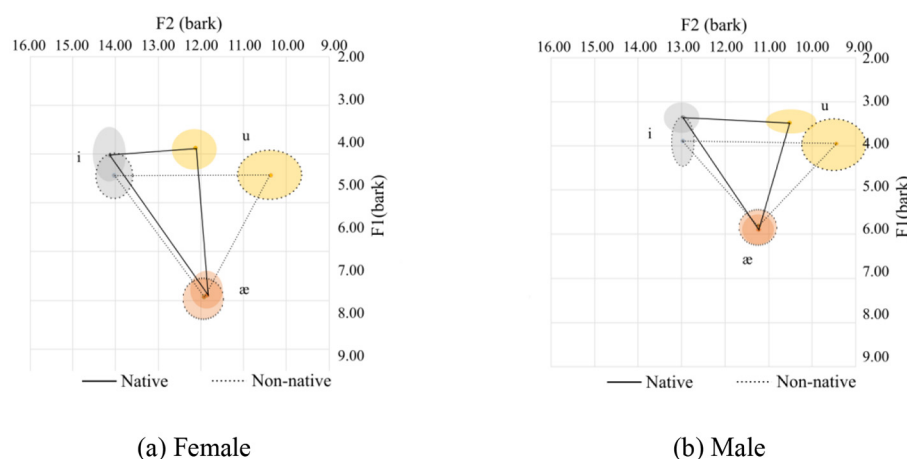


FIGURE 12
Vowel spaces for coda obstruent labials. (a) Female. (b) Male.

in the sentence, which explains why the /u/ sound is pronounced with reduced emphasis like the schwa. In contrast, Chinese speakers do not consider these as functional words and pronounce them with emphasis.

3.3 Discussion

3.3.1 Vowel space

It's not difficult to see that compared to Chen's data, the vowel space derived from this corpus is much smaller, regardless of whether the speakers are native or Chinese, male or female. One possible explanation for this is that the speech samples for the present study are mostly from a sentence instead of an isolated word, that is to say, vowels in this study are from "continuous speech" with numerous segments that exercise the full articulation of the tongue in the oral cavity (Lindblom, 1983). In this case, it is highly probable that vowels are produced without being fully articulated, and the overall speakers' vowel space should shrink accordingly. In contrast, the vowels from Chen's study were read in the same carrier sentence: "Say _____ again five times." In this case, the subjects put some stress on the target word, resulting in the vowels included in the target words being produced in a fully articulated form. The vowel space from the two studies are provided in Figures 15a, b.

3.3.2 Effect of the consonant context on the vowel production

The vowels in this study are not from a uniform carrier sentence as carried out in Chen (1999), Peterson and Barney (1952), and Hillenbrand et al. (1995). Although the same vowels are pooled to calculate the mean and standard deviation of the vowel, words, including the target vowel, are different. This means that the same vowels are not from a homogenous phonological environment. Segmental variations may cause changes in the formant values because of coarticulation. The general trend shown in Figure 5 may

have different facets. Instead of dividing segmental environments according to the identity of onset and coda segments, we grouped some consonants according to their phonological features, such as sonorant, obstruent, labial, etc. Compared to the general trend in Figure 5, not all phonological features have a conspicuous effect, as the dataset did not contain enough tokens. The groups are summarized in Table 14.

These observations and their explanations are as follows:

- The labial-velar sound /w/ in the coda position mainly affects the /i/ corner vowel. This /w/ sound requires the rounding of the lips and raising of the tongue, resulting in /i/ being pronounced similarly to a velar. When /w/ precedes /i/, the lips anticipate a rounded position, while the tongue is relatively pulled back, resulting in a lower F1 and F2 value.
- The /u/ corner vowel is affected by an /f/ consonant in the labial coda position. The dataset contains only 24 tokens for the /u/ vowel for the native speakers belonging to the /f/ sound, which is a labial in the onset position. And all of the 24 F sounds end with the final /d/ sound, which can be traced to the word "food" in the sentence "I like Japanese food, but Korean food is too spicy for me." Compared with the function words "to" and "you," "food" is a content word that should be stressed. Consequently, the native speakers produce the /u/ sound with a lower and more pulled-back tongue position than the Chinese speakers.

/u/ is also affected by the consonant velar /k/ in the coda position due to the words "you," "to," and "who," which serve as the function word in the sentences. This explains why the /u/ sound is pronounced with reduced emphasis, like the schwa by the native speakers. In contrast, Chinese speakers do not consider these as functional words and pronounce them with emphasis.

- The /æ/ corner vowel was mostly affected by the anticipation of the nasal /n/ sound, which requires the tongue to be raised to produce the alveolar nasal sound. Consequently, this

TABLE 11 The F1 and F2 by native and Chinese speakers for coda obstruent labials.

		Native								Chinese							
Female	C. Vowel	ɑ		æ		i		u		ɑ		æ		i		u	
	Formant	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2
	Mean	–	–	755	1,715	402	2,420	383	1,796	–	–	762	1,741	446	2,388	448	1,386
	SD	–	–	119	190	147	260	109	265	–	–	147	240	131	274	148	331
	#Speakers	–		6		6		6		–		44		44		44	
	#Tokens	–		126		74		56		–		921		508		396	
P.value	ɑ	F1: p.value: – Cohen’s d: –; F2: p.value: – Cohen’s d: –															
	æ	F1: p.value: 0.653 Cohen’s d: –0.04; F2: p.value: 0.256 Cohen’s d: –0.11															
	i	F1: p.value: 0.000 Cohen’s d: –0.33; F2: p.value: 0.346 Cohen’s d: 0.11															
	u	F1: p.value: 0.000 Cohen’s d: –0.45; F2: p.value: 0.000 Cohen’s d: 1.26															
Male	C. Vowel	ɑ		æ		i		u		ɑ		æ		i		u	
	Formant	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2
	Mean	–	–	616	1,560	325	2,053	338	1,409	–	–	608	1,563	386	2,029	394	1,196
	SD	–	–	84	172	77	236	66	247	–	–	106	208	144	163	164	313
	#Speakers	–		6		6		6		–		40		40		40	
	#Tokens	–		126		79		54		–		838		465		346	
P.value	ɑ	F1: p.value: –Cohen’s d: –; F2: p.value: 0.000 Cohen’s d:															
	æ	F1: p.value: 0.435 Cohen’s d: 0.07; F2: p.value: 0.852 Cohen’s d: –0.01															
	i	F1: p.value: 0.000 Cohen’s d: –0.44; F2: p.value: 0.775 Cohen’s d: 0.13															
	u	F1: p.value: 0.000 Cohen’s d: –0.36; F2: p.value: 0.000 Cohen’s d: 0.69															
Female	C. Vowel	ɑ		æ		i		u		ɑ		æ		i		u	
	Formant	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2
	Mean	–	–	755	1,715	402	2,420	383	1,796	–	–	762	1,741	446	2,388	448	1,386
	SD	–	–	119	190	147	260	109	265	–	–	147	240	131	274	148	331
	#Speakers	–		6		6		6		–		44		44		44	
	#Tokens	–		126		74		56		–		921		508		396	
Male	C. Vowel	ɑ		æ		i		u		ɑ		æ		i		u	
	Formant	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2
	Mean	–	–	616	1,560	325	2,053	338	1,409	–	–	608	1,563	386	2,029	394	1,196
	SD	–	–	84	172	77	236	66	247	–	–	106	208	144	163	164	313
	#Speakers	–		6		6		6		–		40		40		40	
	#Tokens	–		126		79		54		–		838		465		346	

anticipation moves the tongue closer to the alveolar ridge area, which naturally lowers the F1 value.

It can be observed that the /æ/ sound for the native speakers is slightly lower and more fronted than that of the Chinese speakers. A possible explanation is that the influence is due to the /ɜ/ sound in the onset post-alveolar position followed by either an /n/ or a /p/ sound. By looking through the corpus, the one with the /p/ sound belongs to the word “Japanese,” and the one with the /n/ sound belongs to the word “January.” The native speakers know that these should be pronounced with a strong tone, regardless of whether the first syllable is for January or the second for Japanese. On the contrary, the Chinese speakers did not pronounce it with a relatively strong tone.

- The /a/ corner vowel can be affected by the /w/ sound in the onset position. /w/ is a labial-velar sound, which requires the rounding of the lips and raising the tongue, similar to the pronunciation of the velars. When /w/ precedes /a/, which is a low vowel, the lips tend to remain rounded with the tongue still in the velar position resulting in a more rounded or protruded lips position for the /a/ sound. This lowers the F1 and F2 values.

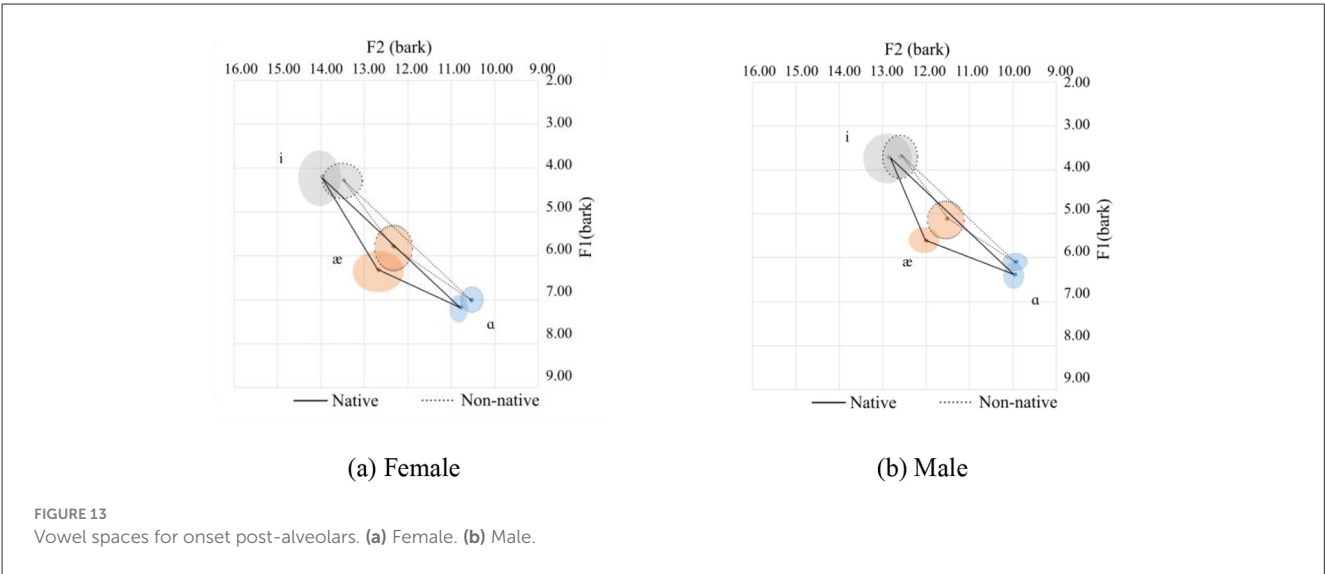


TABLE 12 The F1 and F2 by native and Chinese speakers for onset post-alveolars.

		Native								Chinese							
Female	C. Vowel	ɑ		æ		i		u		ɑ		æ		i		u	
	Formant	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2
	Mean	788	1,448	678	1,960	423	2,369	–	–	768	1,396	608	1,851	426	2,200	–	–
	SD	95	96	130	346	169	326	–	–	85	118	143	247	106	300	–	–
	#Speakers	6		6		6		–		44		44		44		–	
	#Tokens	6		36		48		–		44		264		353		–	
P.value	ɑ	F1: p.value: 0.000 Cohen's d: 0.23; F2: p.value: 0.000 Cohen's d: 0.44															
	æ	F1: p.value: 0.000 Cohen's d: 0.49; F2: p.value: 0.000 Cohen's d: 0.41															
	i	F1: p.value: 0.847 Cohen's d: –0.02; F2: p.value: 0.000 Cohen's d: 0.55															
	u	F1: p.value: –Cohen's d: –; F2: p.value: –Cohen's d: –															
Male	C. Vowel	ɑ		æ		i		u		ɑ		æ		i		u	
	Formant	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2
	Mean	687	1,448	581	1,753	335	2,148	–	–	767	1,396	594	1,719	369	2,092	–	–
	SD	125	143	101	161	69	215	–	–	96	174	100	189	127	182	–	–
	#Speakers	6		6		6		–		40		40		40		–	
	#Tokens	6		36		181		–		44		476		1,193		–	
P.value	ɑ	F1: p.value: 0.000 Cohen's d: –0.31; F2: p.value: 0.312 Cohen's d: 0.07															
	æ	F1: p.value: 0.000 Cohen's d: –0.73; F2: p.value: 0.002 Cohen's d: –0.16															
	i	F1: p.value: 0.000 Cohen's d: –0.28; F2: p.value: 0.000 Cohen's d: 0.30															
	u	F1: p.value: –Cohen's d: –; F2: p.value: –Cohen's d: –															

4 Conclusions

In this contribution, the pronunciation of the four corner vowels /i/, /u/, /æ/, and /ɑ/ from the AESOP-ILAS dataset source has been studied to understand better the factors that can help Chinese speakers achieve more native like pronunciation. Unlike previous studies, which measured the formants at the middle of the target vowel, the formants in this study were measured at the start, middle, and end portions of the vowel pronunciation,

and then their averages were taken. A significant reduction in the vowel space was observed in this study compared to previous studies, irrespective of the speaker's linguistic background or gender. This observation could be attributed to the methodology used to collect the speech samples as the vowel sounds were derived from a predominantly sentence-based corpus, meaning they were extracted from continuous speech where the tongue fully exercises its articulatory potential. Consequently, the vowels produced were not fully articulated, resulting in a reduced vowel space.

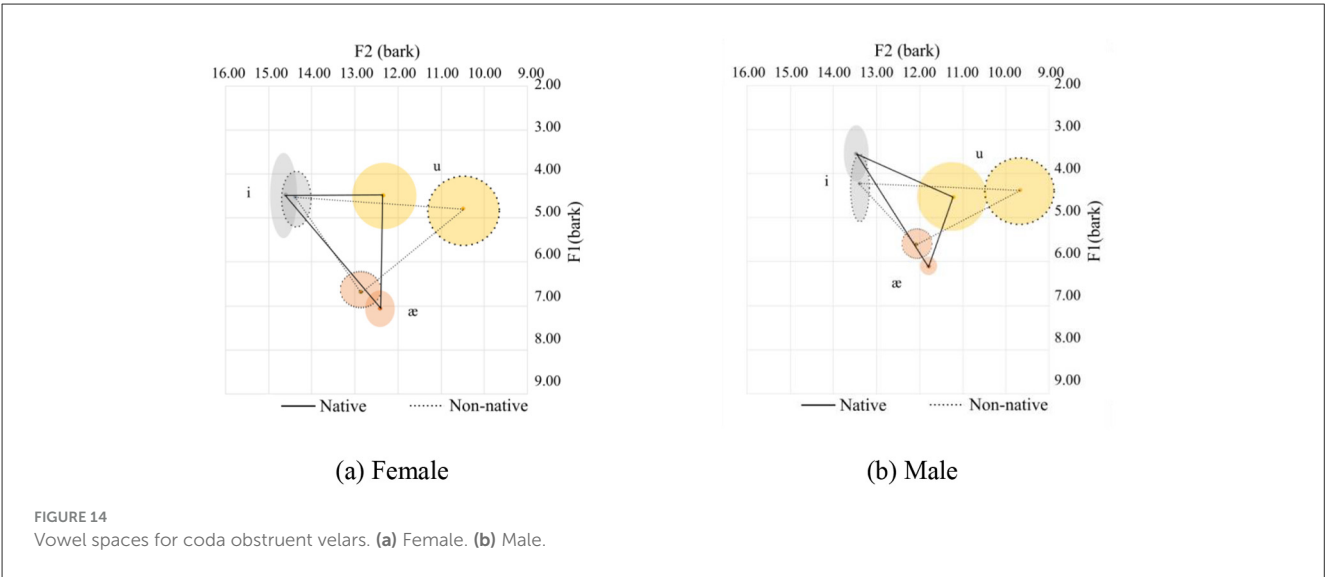


TABLE 13 The F1 and F2 by native and Chinese speakers for coda obstruent velars.

		Native								Chinese							
Female	C. Vowel	ɑ		æ		i		u		ɑ		æ		i		u	
	Formant	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2
	Mean	–	–	778	1,870	375	2,587	385	1,874	–	–	742	2,004	433	2,504	445	1,422
	SD	–	–	121	189	235	235	180	388	–	–	116	257	155	234	242	391
	#Speakers	–		6		6		6		–		44		44		44	
	#Tokens	–		18		193		18		–		130		210		123	
P.value	ɑ	F1: p.value: –Cohen’s d: –; F2: p.value: – Cohen’s d: –															
	æ	F1: p.value: 0.000 Cohen’s d: 0.30; F2: p.value: 0.000 Cohen’s d: –0.53															
	i	F1: p.value: 0.000 Cohen’s d: 0.02; F2: p.value: 0.000 Cohen’s d: 0.35															
	u	F1: p.value: 0.001 Cohen’s d: –0.17; F2: p.value: 0.000 Cohen’s d: 1.15															
Male	C. Vowel	ɑ		æ		i		u		ɑ		æ		i		u	
	Formant	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2
	Mean	–	–	643	1,697	350	1,697	353	1,588	–	–	582	1,775	428	2,155	392	1,245
	SD	–	–	55	103	144	181	182	383	–	–	89	181	188	146	203	335
	#Speakers	–		6		6		6		–		40		40		40	
	#Tokens	–		18		30		18		–		119		188		155	
P.value	ɑ	F1: p.value: – Cohen’s d: –; F2: p.value: – Cohen’s d: –															
	æ	F1: p.value: 0.000 Cohen’s d: 0.71; F2: p.value: 0.000 Cohen’s d: –0.45															
	i	F1: p.value: 0.000 Cohen’s d: –0.42; F2: p.value: 0.000 Cohen’s d: –3.02															
	u	F1: p.value: 0.028 Cohen’s d: 0.07; F2: p.value: 0.000 Cohen’s d: 1.00															

Unexpectedly, the vowel space for the Chinese speakers was larger than that of the native speakers compared with previous studies due to the speech samples coming from a sentence rather than an isolated word, which painted a more reliable picture of the vowel space. In addition, Chinese speakers pronounced the corner vowels with lower F1 and F2 values or with a lower and pulled-back tongue position, and the most striking difference was with the pronunciation of /a/ and /u/.

The differences in the /u/ corner vowel were traced back to the pronunciation of “to,” “you,” and “who,” which significantly raised F2 for /u/ in the velar coda position with consonant /k/, and to the word “food” in the labial onset position, which significantly lowered F2. As for /a/, it was affected by a preceding /w/ sound in the sonorant onset position, which lowered F1 and raised F2. The difference in /i/ was mostly observed in the sonorant coda position with a following /w/ sound, which lowered F1 and F2. /æ/ was mostly affected in the nasal coda position by a preceding /n/

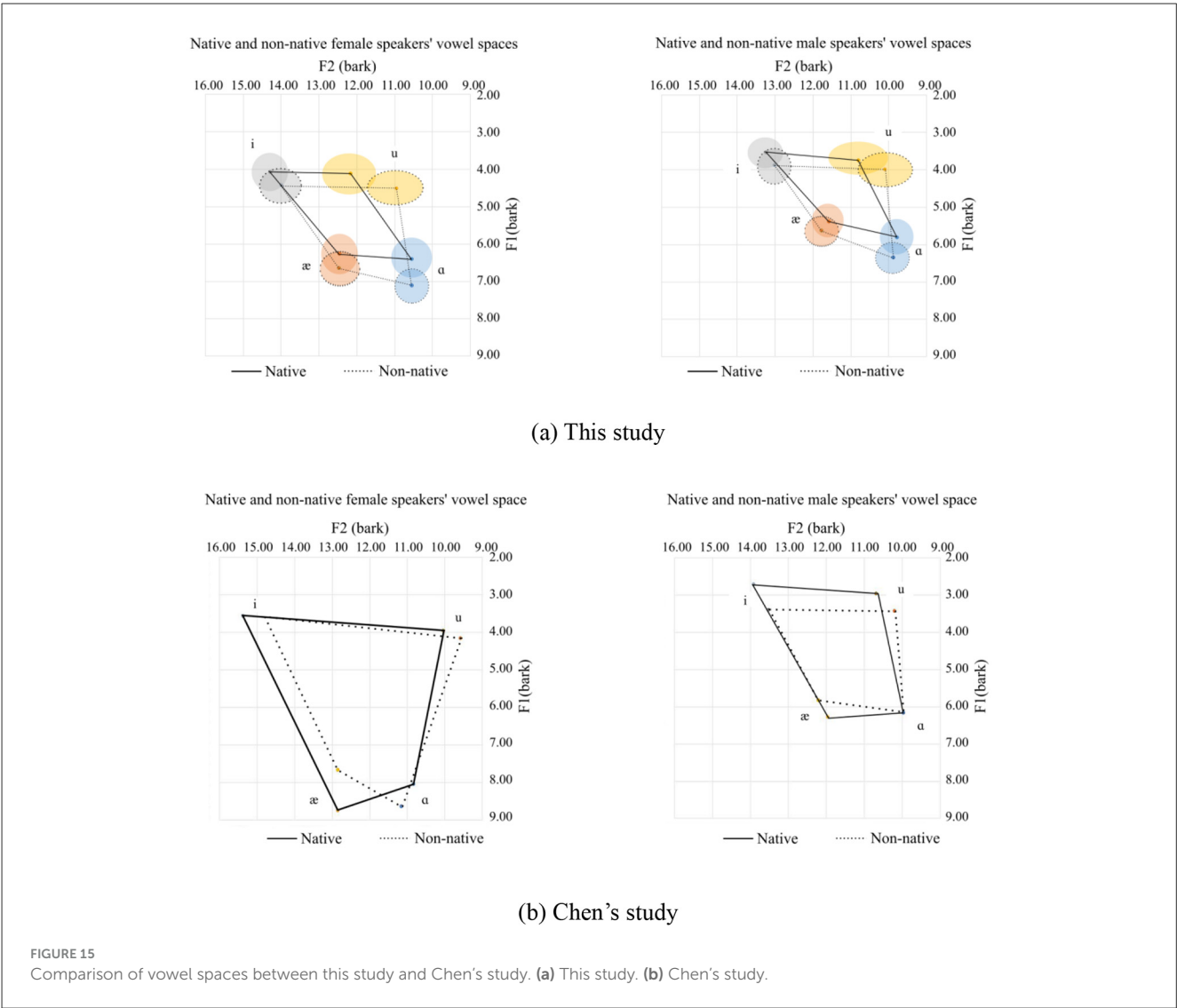


TABLE 14 Summary of the contextual consonantal effects on the vowels.

Corner vowel	Context in a syllable	Affected by	Effect
/i/	Sonorant coda (/r/*, /l/*, /j/*, /w/)	With the /w/ sound after /i/ sound	F1 is slightly lower, F2 is significantly lower
/u/	Labials Onset (/p/*, /b/*, /v/*, /f/)	Mainly from the stressed content word “food”	F2 is significantly lower
	Velar Coda (/k/)	Mainly from the unstressed content word “you,” “to,” and “who”	F2 is significantly higher compared with the general trend
/æ/	Nasal Coda (/n/*, /m/*, /ŋ/*)	With the /n/ sound after /æ/ sound	F1 is significantly lower
	Post-alveolar onset (/ʒ/, /ʃ/)	Words are “Japanese”/“January”	F1 is significantly higher observed particularly among the Chinese speakers
/ɑ/	Sonorant onset (/r/*, /l/*, /j/*, /w/)	With the /w/ sound before /ɑ/ sound	F1 is significantly lower/F2 is slightly higher

sound which lowered the F1 formant, and in the post-alveolar in the onset position due to the pronunciation of the words “January” and “Japanese,” which raised F1. It is believed that by incorporating when to stress a certain word in a sentence, the pronunciation of the Chinese speakers will be greatly improved.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

HZ: Visualization, Resources, Writing – original draft, Validation, Formal analysis, Writing – review & editing, Funding acquisition, Data curation, Supervision, Methodology, Software.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported by the 2024 Young Doctor Scientific Research Start-up Fund Project of Harbin University (HUDF2024206), and the 2025 Heilongjiang Provincial Key Research Project on Economic and Social Development (Foreign Language Specialization) Project Number: WY2025072.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Best, C. T., and Tyler, M. D. (2007). “Nonnative and second-language speech perception: commonalities and complementarities” in *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*, eds. O.-S. Bohn and M. J. Munro (Amsterdam: John Benjamins), 13–34. doi: 10.1075/llt.17.07bes
- Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., and Tohkura, Y. (1997). Training Japanese listeners to identify English/r/and/l/: IV. Some effects of perceptual learning on speech production. *J. Acoustical Soc. Am.* 101, 2299–2310. doi: 10.1121/1.418276
- Celce-Murcia, M., Brinton, D. M., and Goodwin, J. M. (1996). *Teaching Pronunciation: A Reference for Teachers of English to Speakers of Other Languages*. Cambridge: Cambridge University Press.
- Chen, Y. (1999). *The acoustic realization of English vowels by Mandarin speakers* (Unpublished doctoral dissertation). University of Illinois at Urbana-Champaign, Champaign, IL, United States.
- Chen, Y., Robb, M., Gilbert, H., and Lerman, J. (2001). Vowel production by Mandarin speakers of English. *Clin. Linguist. Phonetics* 15, 427–440. doi: 10.1080/02699200110044804
- Escudero, P. (2005). *Linguistic perception and second language acquisition: Explaining the attainment of optimal phonological categorization* (Doctoral dissertation). Utrecht University, Utrecht, Netherlands. LOT Dissertation Series.
- Flege, J. E. (1995). “Second language speech learning: theory, findings, and problems,” in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, ed. W. Strange (New York, NY: York Press), 233–277.
- Flege, J. E., Bohn, O.-S., and Jang, S. (1997). Effects of experience on non-native speakers’ production and perception of English vowels. *J. Phon.* 25, 437–470. doi: 10.1006/jpho.1997.0052
- Flege, J. E., Munro, M. J., and MacKay, I. R. A. (1999). Factors affecting strength of perceived foreign accent in a second language. *J. Acoust. Soc. Am.* 97, 3125–3134. doi: 10.1121/1.413041
- Hillenbrand, J., Clark, M. J., and Houde, R. A. (2001). Some effects of duration on vowel recognition. *J. Acoust. Soc. Am.* 110, 3013–3022. doi: 10.1121/1.1323463
- Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (1995). Acoustic characteristics of American English vowels. *J. Acoust. Soc. Am.* 97, 3099–3111. doi: 10.1121/1.411872
- Jones, D. (1960). *An Outline of English Phonetics*, 9th Edn. Cambridge University Press.
- Ladefoged, P., and Johnson, K. (2015). *A Course in Phonetics*, 7th Edn. Cengage Learning.
- Lindblom, B. (1983). “Economy of speech gestures,” in *The Production of Speech*, ed. P. F. MacNeilage (Cham: Springer), 217–245. doi: 10.1007/978-1-4613-8202-7_10
- Munro, M. J., and Derwing, T. M. (2008). Segmental acquisition in adult ESL learners: a longitudinal study of vowel production. *Lang. Learn.* 58, 479–502. doi: 10.1111/j.1467-9922.2008.00448.x
- Peterson, G. E., and Barney, H. L. (1952). Control methods used in a study of the vowels. *J. Acoust. Soc. Am.* 24, 175–184. doi: 10.1121/1.1906875
- Stevens, K. N., and House, A. S. (1963). Perturbation of vowel articulations by consonantal context: an acoustical study. *J. Speech Hear. Res.* 6, 111–128. doi: 10.1044/jshr.0602.111

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2025.1598904/full#supplementary-material>



OPEN ACCESS

EDITED BY

Antonio Benítez-Burraco,
University of Seville, Spain

REVIEWED BY

Amy Scott,
University of Canterbury, New Zealand
Victoria Sánchez-Gómez,
University of Salamanca, Spain
Ava Karusoo-Musumeci,
La Trobe University, Australia

*CORRESPONDENCE

Suziyani Mohamed
✉ suziyani@ukm.edu.my

RECEIVED 16 May 2025

REVISED 22 November 2025

ACCEPTED 09 December 2025

PUBLISHED 12 January 2026

CITATION

Zhang ZH, Mohamed S and Bakar KA (2026)
Parental emotional expression and children's
oral narrative ability: a systematic review.
Front. Psychol. 16:1629744.
doi: 10.3389/fpsyg.2025.1629744

COPYRIGHT

© 2026 Zhang, Mohamed and Bakar. This is
an open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](#). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with these
terms.

Parental emotional expression and children's oral narrative ability: a systematic review

Zhi Hong Zhang, Suziyani Mohamed* and Kamariah Abu Bakar

Faculty of Education, Universiti Kebangsaan Malaysia, Selangor, Malaysia

This study aims to systematically examine the relationship between parental emotional expression and children's oral narrative abilities, as well as the developmental trends in this research field. Based on the Scopus database and adhering to PRISMA principles, 81 articles published between 2010 and 2025 were selected. A combined approach of bibliometric analysis and thematic analysis was employed for the investigation. The bibliometric analysis revealed an overall upward trend in research activity within this field, with peaks in 2015 and 2021. Recent years have witnessed shifts in research focus and methodologies. Themes have gradually expanded from general language development to emotional socialization and family interactions, demonstrating interdisciplinary convergence. Thematic analysis indicates that parental emotional expression influences children's oral narrative development through multiple pathways, including the family literacy environment, digital devices, parent-child interaction, and cultural background. Positive emotional expression and high-quality parent-child interactions promote language richness and emotional depth in children's narratives, whereas negative or absent emotional expression may inhibit their language organization abilities. Research indicates that parental emotional expression exerts both direct and indirect effects on children's narrative abilities. Future studies should strengthen interdisciplinary collaboration by integrating perspectives from psychology, education, and digital technology to deepen understanding of the mechanisms through which the family emotional environment influences children's language development.

KEYWORDS

parents, emotional expression, children's oral narrative, bibliometric analysis, systematic review

1 Introduction

Preschool is a critical period for the development of children's oral narrative ability. Oral narrative ability is the ability to express language in an organized manner out of context (Bailey et al., 2020). Miles and Chapman (2002) define oral narrative ability as the ability to understand, interpret, and retell stories accurately and meaningfully. Proficient language and communication skills are fundamental for individuals to adapt successfully to society. Narrative competence, as a crucial aspect of children's language development, significantly influences their social, cognitive, and psychological development (Gardner-Neblett, 2022; Hidayat and Fitria, 2021; van der Wilt et al., 2022). In recent years, more and more researchers have begun to pay attention to the development of children's oral narrative abilities. For example, Scott et al. (2022) and Johnson et al. (2023) successively designed a new systemic framework for evaluating children's oral narrative abilities. By

extracting features from recordings of children retelling stories, this framework enables fair assessments of children's language skills to support and enhance teaching practices. Additionally, Lucero (2018) focused on analyzing the impact of bilingual education on children's oral narrative abilities. Research findings indicate a significant correlation between children's oral narrative performance across languages, with bilingual children demonstrating distinct advantages over monolingual peers. However, Clark-Whitney and Melzi (2023) shifted focus to examining the influence of executive function on oral narrative abilities in bilingual children. Their study revealed that executive function more effectively promotes the development of both oral narrative skills and early reading abilities in children.

Oral narrative ability typically encompasses two categories: personal narratives and fictional narratives (Allen et al., 1994; Kim, 2018). The former is based on children's autobiographical memories, primarily involving the organization and expression of real-life events; the latter relies on imaginative construction, emphasizing creativity and plot structure. However, this study focuses on children's personal narrative ability, as personal narratives more directly reflect children's socialization experiences and emotional processing within the family's emotional environment. Parents' emotional expressions typically occur within authentic interactive contexts and exert a critical influence on how children narrate their own experiences.

The family is an important environment for children's growth and emotional enlightenment, providing emotional support and security, and is an initiating place for children's language development, even though a systematic school curriculum cannot be separated from the support and cooperation of family emotional education (Buschmann, 2020; Göksun et al., 2025; Roostin, 2018). It has been shown that a family's socioeconomic status, the availability of learning resources, parenting styles, and level of education affect the development of children's language communication abilities (Fang et al., 2018; Qi and Wu, 2020; Schwab and Lew-Williams, 2016). Family support for children's language and learning and the resources provided are beneficial in promoting children's language communication abilities (Cohen et al., 2020; Zheng et al., 2021). Family support and resources refer to parental involvement in children's learning, access to educational materials at home, and emotional encouragement provided by caregivers (Fuller et al., 2019; Otero-Mayer et al., 2025). These factors are considered key environmental influences on children's language and narrative development.

Parental emotional expression refers to the primary mode or style of verbal and non-verbal expression exhibited by parents in the home, i.e., how emotions are expressed and conveyed through words, facial expressions, body language, etc., (Tang et al., 2023). The core function of emotional expression is to communicate needs and wants, initially realized within the family and then extended to the wider social environment, where its form, intensity, and inhibition change depending on the context (Keltner et al., 2019). Parents are crucial to the development of children's oral narrative abilities, in which emotional expression can intuitively serve as a stimulus to the development of children's oral narrative abilities (Leyva et al., 2014; Masters et al., 2022). Parental emotional expression influences language development through the family

literacy environment, thus affecting children's narrative ability. Researchers usually consider the family literacy environment as factors in the home that influence literacy development and define it in different dimensions (Dong et al., 2020; Torppa et al., 2022).

Romero-González et al. (2023) argues that the family literacy environment affects children's reading ability through interpersonal interactions, the physical environment of the home, and motivational atmosphere. Fish and Pinkerman (2003) summarized the family literacy environment in terms of three dimensions: contextual variables, child characteristics, and parent-child interactions. Some scholars have proposed specific components for the home reading and writing environment. Pietropoli and Gracia (2025) argue that it should encompass the quality of family activities, material resources, and the social environment. Fekonja-Pekljaj et al. (2010) share a similar perspective, suggesting that the family literacy environment should include the quality of family activities, physical resources, and the social context. In addition, Sénéchal and LeFevre (2002), as well as Puglisi et al. (2017) categorized the family literacy environment into two formal literacy activities and informal literacy activities dimensions, where formal literacy activities are those in which adults directly teach reading or improve print-related skills at home, and informal literacy activities are those in which children engage in reading or are directed to pay attention to print in the environment. As digital technology has long been immersed in family life, it has led to an increase in children's use of smartphones and tablets (Holloway et al., 2013). As a result, family literacy environments are no longer limited to traditional print activities. Based on the above points, this study defines the family literacy environment as a multidimensional concept encompassing interpersonal interactions within the family, the material and physical resources of the family environment, the sociocultural context, formal and informal literacy activities, and the influence of digital technology integration into family life.

Previous review studies have primarily focused on multifaceted aspects within the domain of factors influencing child development. Research by Liu and Zhao (2024) underscored the significance of these factors in child development, yet did not specifically examine the impact of parental emotional expression on children's oral narrative abilities. Empirical research on the effects of parental emotional expression on children's oral narrative abilities remains extremely limited, with only a handful of studies such as Sher-Censor and Yates (2015) venturing into this specific domain. In Western countries, particularly the United States, Canada, and some European nations, a substantial body of empirical research exists on the effects of family environments on children's language and cognitive development. In contrast, empirical research in China has primarily focused on the effects of parenting styles on children's language development (Yanxue and Bhaumik, 2024; Yu et al., 2024), the effects of family socioeconomic status on children's language and cognitive development (González et al., 2024; Madya et al., 2019), and the impact of bilingual environments on language development (Zhang et al., 2023).

Although an increasing number of studies have examined the relationship between parental emotional expression and children's oral narrative ability, existing empirical research is mostly based on Western studies, neglecting the cultural heterogeneity of

Asian societies in parenting practices and child development; In addition, the research results on the relationship between parental emotional expression and children's oral narrative ability are still remain fragmented, and there is no comprehensive review has systematically synthesized this relationship. Based on this, the purpose of this research systematic review is (i) to systematically organize and synthesize existing literature on the relationship between parental emotional expression and children's oral narrative abilities in the Asian countries, thereby clarifying the current state of research and its primary conclusions of Asian culture; (ii) to identify research hotspots and developmental trends in this field in the Asian countries through bibliometric analysis; and (iii) to summarize theoretical and methodological shortcomings in existing studies in the Asian countries, thereby providing direction for future regional targeted empirical research. The operational framework of this study is shown below in [Figure 1](#).

2 Data and methods

This article follows the PRISMA guidelines for conducting and reporting literature reviews. It mainly includes identification the articles, screening the articles, and eligibility and included the articles.

2.1 Identification

This study utilized the Scopus database, which is recognized as one of the most authoritative and comprehensive academic databases globally. It encompasses over 294 research areas ([Yoon et al., 2023](#)). According to the Scopus database, a total of 26 journals related to parental emotional expression and children's oral narrative ability are indexed in it, indicating its strong representation and academic depth.

Although existing studies have suggested combining multiple databases to expand the coverage of literature ([Younger, 2010](#)), in this study, Scopus was finally selected as the only core data source due to the considerations of research focus, data processing efficiency, and resource integration. To ensure the systematic and comprehensive nature of the search, the search terms used were: TITLE-ABS-KEY (((("parents" OR "caregivers" OR "guardians") AND ("emotional expression" OR "affect" OR "emotion" OR "feelings") AND ("children" OR "youth" OR "kids" OR "offspring") AND ("oral narrative" OR "storytelling" OR "narrative ability" OR "verbal skills") AND ("development" OR "skills" OR "competence" OR "literacy")))) AND PUBYEAR > 2009 AND PUBYEAR < 2026 AND (LIMIT-TO (LANGUAGE, "English")) AND (LIMIT-TO (DOCTYPE, "ar") OR LIMIT-TO (DOCTYPE, "cp") OR LIMIT-TO (DOCTYPE, "re")). A total of 81 potentially relevant documents were retrieved in Scopus.

In addition, in order to try to compensate for the possible coverage limitations of a single database, this study synchronously referred to important research findings mentioned in existing reviews and highly cited literature in the field, to provide more in-depth insights into the research on parental emotional expression and children's oral narrative ability. After this supplementary search, no further relevant articles were identified beyond those

already retrieved through the initial database search. Therefore, a total of 81 articles were selected. These selected articles will be utilized in bibliometric analysis, providing comprehensive and reliable data support for in-depth examination of the field's document profiles, publication trends, publications by authors, publications by countries, and top keywords. See [Table 1](#) for details.

2.2 Screening

In order to maintain the high quality of the content and to ensure a manageable number of articles were selected, the search for articles was limited to "papers," "published articles" and "refereed articles," "conferences," and "scholarly papers" published in peer-reviewed journals, as this is the main source of empirical data. Only English-language resources were included in the review. All articles published between 2010 and 2025. Most importantly, articles published in the field of "Developmental and Educational Psychology" and "Education" were selected in order to increase the likelihood of retrieving relevant articles. In addition to this, after careful consideration of existing literature and research objectives, we found that studies focused on the Asian region can explore research objectives in a more in-depth and targeted manner, offering valuable insights that research in other regions may not fully capture. Therefore, all studies from Asian countries were included in the screening of the article, while articles from non-Asian countries were excluded. A total of 14 articles were retained at this stage (see [Table 2](#)).

2.3 Eligibility and included

The third stage is eligibility stage. Based on the articles retrieved from the database in Method 2.2, all abstracts and the main content of the articles of the 14 original samples were read to ensure that they met the inclusion criteria. Any articles relating to the relationship between parents and children's oral language were retained. The tendency was to retain articles that contained descriptive ambiguity in their abstracts because the literature on parental emotions and children's oral narrative ability is extensive and covers many different dimensions, such as parental emotional expression, parental emotional transmission, parental emotional response, family interactions, children's creativity, children's emotional development, and so on. Therefore, no inclusion/exclusion criteria were used other than whether the papers included both parental emotion and children's narrative ability topics in their abstracts. Similarly, articles that did not focus on Asian countries and regions were similarly excluded. After screening, the number of relevant articles has been reduced to six. Please refer to [Table 3](#) for details. These selected studies will serve as the foundation for thematic analysis, providing rich data and case support for subsequent theme extraction. The methodological framework for the review is shown in [Figure 2](#).

Given the limited number of studies ultimately included, this research employed a narrative thematic synthesis approach rather than inductive thematic analysis involving line-by-line coding. The researchers conducted multiple rounds of repeated

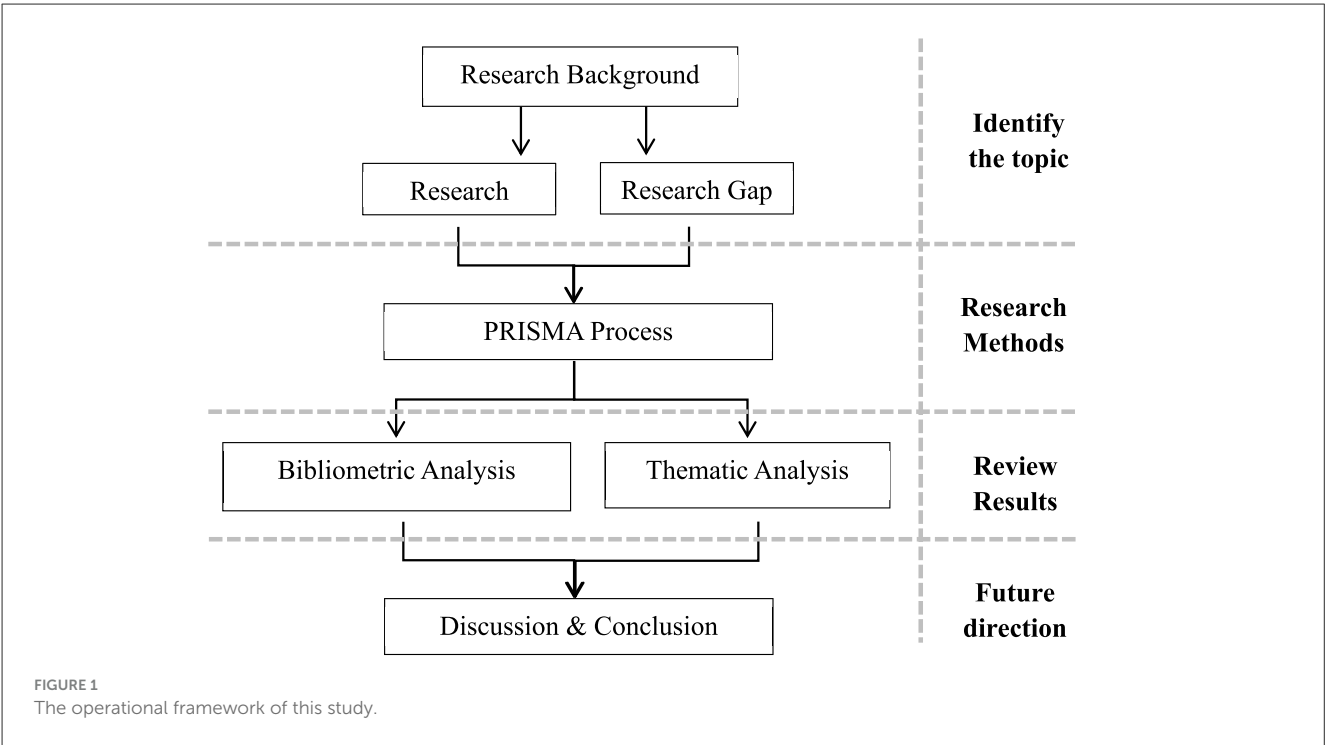


TABLE 1 Keywords used for search and number of papers retrieved.

1st step	
Item	Details
Keywords	"parents," "caregivers," "guardians"
	"emotional expression," "affect," "emotion," "feelings"
	"children," "youth," "kids," "offspring"
	"oral narrative," "storytelling," "narrative ability," "verbal skills"
	"development," "skills," "competence," "literacy"
Operators	"OR," "AND"
Period	2010-2025
Language	English
Data source	Scopus
Search date	March 2025
Total number	81

readings of all included studies to familiarize themselves with their research focus, core conclusions, and conceptual frameworks. Through repeated comparisons across the literature, they identified recurring viewpoints and patterns. Subsequently, these shared elements were integrated and elevated to form higher-level descriptive themes.

3 Review results

The analysis focused on two core areas: an in-depth bibliometric analysis of 81 documents indexed in the Scopus database, and a systematic thematic analysis of these documents

TABLE 2 The inclusion and exclusion criteria.

Criteria	Inclusion	Exclusion
Article type	Research article, review article, conference paper	Book, book chapter, erratum, editorial
Language	English	Non-English
Year	2010-2025 year	Less than 2010 year
Country/region	Asian countries	Non-Asian countries
Subject areas	Developmental and educational psychology, education	Non-developmental and educational psychology, education

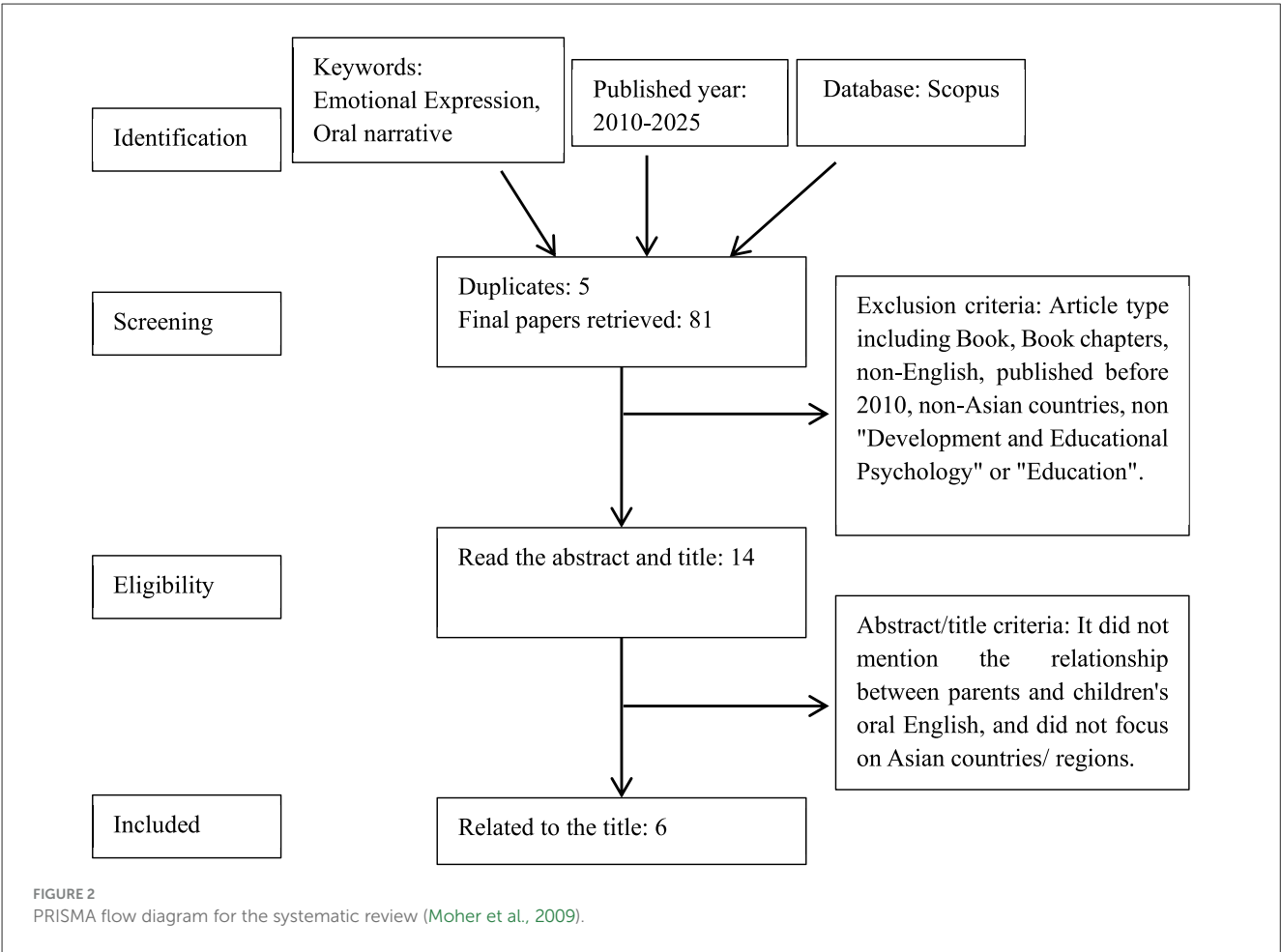
on the topic of parental emotional expression and children's oral narrative ability. Firstly, at the level of bibliometric analysis, 81 articles were selected from Section 2.1 as the dataset to analyze their literature, publication trends, co authors, and countries. Our aim is to reveal the research hotspots and future research trends in this field. Secondly, in terms of thematic analysis, the six articles selected in Section 2.3 were treated as a dataset to explore the intrinsic relationship between parental emotional expression and children's oral narrative ability in published articles. Through careful combing and summarization of the literature content, it was classified as the influence of home literacy environment, digital devices, parent-child interaction, and cultural background factors on children's oral narrative ability.

3.1 Bibliometric analysis

This study conducted a bibliometric analysis based on 81 articles retrieved from the Scopus database, primarily utilizing

TABLE 3 List of studies included in the systematic review.

No.	Authors	Published year	Title	Study design	Country of focus	Key findings	Database type
1	Debarati M. et al.	2024	Digital tools for direct assessment of autism risk during early childhood: a systematic review	Review	Israel, United States, Australia, etc.	Evaluating the risk of autism in children using digital tools can be achieved by recording and analyzing their interactions and verbal responses with digital interfaces. Indirectly reflecting the role of family interaction patterns in children's oral development	Scopus
2	Anita W. et al.	2011	Development of the language subtest in a developmental assessment scale to identify Chinese preschool children with special needs	Questionnaire	Hong Kong, China	Develop and validate the language subtest of the Preschool Developmental Assessment Scale (PDAS), and compare and analyze the differences in language abilities between children with Special Educational Needs (SEN) and ordinary children through questionnaire research on children	Scopus
3	Sebastian et al.	2019	Parents, Television, and Children's Emotional Expressions: A Cross-Cultural Multilevel Model	Questionnaire, interview	China, Malaysia, Thailand, etc.	Using a cross-cultural multilevel model, we analyzed self-reported data from 3,570 children aged 6–15 from 13 different countries (Asia, Europe, Middle East, and South America) to explore the impact of parents and television on children's emotional expression	Scopus
4	Hong et al.	2021	Meta-Analysis of RCTs of Technology-Assisted Parent-Mediated Interventions for Children with ASD	Review	Israel, United States, Australia, etc.	Technology assisted parent mediated interventions may be effective in improving emotional recognition in children with autism, but no significant differences were observed in social communication, social functioning, and language outcomes. Currently, there is no single technology assisted parent mediated intervention that fully meets evidence-based intervention standards	Scopus
5	Wen et al.	2024	Association between home literacy environment and eye movements during storybook reading: implications for vocabulary and word reading in Chinese preschoolers	Questionnaire	China	Interactive reading is the only significant component of the family reading and writing environment that affects children's attention to text and word reading. Children's attention to text is significantly correlated with their vocabulary and word reading ability, indicating that more attention to text can help improve reading ability. Although other components of the family's reading and writing environment (such as language stimulation, reading books, collaborative activities and conversations, and zone of proximal development stimulation) are associated with certain aspects of reading ability, interactive reading is most prominent in promoting children's attention to text and word reading	Scopus
6	Hsieh, M.-F.	2025	This is My Favorite Book': Exploring Children's Perspectives on Books and Reading	Interview	Taiwan, China	Children have a wide range of preferences for book types, including storybooks, information books, and animated adaptations. They actively participate in reading, understand and share their interests, and express their opinions. Children cherish the time spent reading with adults, believing that it can enhance reading comprehension and bring joy. At the same time, they hope to have more opportunities to choose books and participate in purchasing decisions. These findings emphasize the initiative of children in reading and the need for adult support	Scopus



Excel and VOS viewer software. Following the Scopus search, a Worldwide Scientific Production Map was automatically generated to analyze publication distribution by country. When employing VOS viewer, the analysis type was set to co-occurrence analysis of author keywords to identify research hotspots and keyword associations, with a threshold of 3 applied to ensure keyword representativeness. Other analyses—including publication distribution by author, publication trends, and literature overview—were processed using Excel.

3.1.1 Documents profiles

Among the documents screened under the theme of this study, Article, as the most dominant form of publication in the field, accounted for 83.95% of the total number of publications, which was significantly higher than Conference Paper (12.35%) and Review (3.70%; see Table 4). This result suggests that Article, as the core carrier of original research results, dominates in scholarly communication and knowledge accumulation. Articles usually adopt a strict peer-review mechanism and follow a systematic research methodology and theoretical framework (Lim, 2021), so their high percentage reflects the systematic and scientific nature of research in the field. In contrast, Conference Paper is mainly used for communication of cutting-edge research in academic conferences (Lee, 2019), and its relatively small number may

indicate the relatively limited conference communication of this research topic. Meanwhile, Reviews accounted for only 3.70%, indicating that the systematic organization and integration of existing research results in this field is limited.

Table 5 reveals the distribution of different publication sources. The study shows that Journal is the most important source of publication in this field, accounting for 86.42%, far exceeding Conference Proceeding (11.11%) and Book Series (2.47%). This trend further validates the centrality of journals in the dissemination and accumulation of scholarly results. Articles have become the main publication channel for researchers in the field due to their long review cycle, high academic rigor and wide academic influence. In contrast, Conference Proceedings still play an important role in promoting academic communication and presenting the latest research results, but their influence is usually not as high as that of high-quality journals, and Book Series accounts for the lowest proportion of 2.47%, which indicates that the results of this research topic are rarely published in the form of scholarly monographs or book series. However, as an important carrier of systematic reviews, theoretical innovations and interdisciplinary research, book series still have a certain academic value for the accumulation and dissemination of knowledge in this field.

Table 6 shows the distribution of publications in different subject areas, revealing the research hotspots and disciplinary

TABLE 4 Document type.

Document type	Total publications (TP)	Percentage (%)
Article	68	83.95
Conference paper	10	12.35
Review	3	3.70

TABLE 5 Source type.

Source type	Total publications (TP)	Percentage (%)
Journal	70	86.42
Conference proceeding	9	11.11
Book series	2	2.47

TABLE 6 Subject area.

Subject area	Total publications (TP)	Percentage (%)
Medicine	32	39.51
Psychology	30	37.04
Social sciences	15	18.52
Computer science	11	13.58
Nursing	7	8.64
Health professions	6	7.41
Neuroscience	6	7.41
Biochemistry, genetics and molecular biology	4	4.94
Arts and humanities	3	3.70
Environmental science	3	3.70
Mathematics	3	3.70
Decision sciences	2	2.47
Engineering	2	2.47
Pharmacology, toxicology and pharmaceutics	2	2.47
Physics and astronomy	2	2.47
Dentistry	1	1.23
Energy	1	1.23
Materials science	1	1.23

diversity of the topic of parental emotional expression and children’s oral narrative abilities. Medicine (39.51%) and psychology (37.04%) together constitute the core areas of this research topic with a relatively high percentage. This indicates that the research on this topic focuses mainly on the medical and psychological perspectives of children’s language development and parents’ emotional expression. In addition, social sciences (18.52%) served as the third most important area of study, reflecting the importance of this theme in educational psychology, family relationships, and social development. In addition to the core disciplines, the higher share of computer

science (13.58%) indicates that more and more studies try to utilize technological tools such as Artificial Intelligence and Natural Language Processing to analyze children’s narrative abilities or the characteristics of parents’ emotional expressions. For example, computer-assisted analysis tools may be used to automatically assess children’s oral expression (Hu et al., 2022; Sinclair et al., 2021) or to identify and quantify patterns of parental emotional expression (Narimisaei et al., 2024; Tuyen et al., 2017). Meanwhile, nursing (8.64%), health sciences (7.41%), and neuroscience (7.41%) also suggest that the topic is expanding into the broader field of health and cognitive sciences. This may involve early intervention (Custode et al., 2024; Hou, 2022; van der Wilt et al., 2019), child mental health (Conway et al., 2021; Osborne and Ahinkorah, 2024; Wang, 2023), and how neural mechanisms (Benítez-Barrera et al., 2024; Nguyen, 2023; Romeo et al., 2018; Tulling, 2022; Zhai et al., 2023) affect children’s language abilities. A small number of studies involved biochemistry, genetics and molecular biology (4.94%) exploring the influence of genetic factors. Arts and humanities (3.70%), on the other hand, focused on the cultural dimension of narrative expression. The multidisciplinary nature of the studies not only contributes unique perspectives to the field but also increases the diversity of research.

3.1.2 Publication trends

From the data in Figure 3 and Table 7, the number of publications on this topic shows a fluctuating growth trend in general. 2010–2014 has a low publication volume, with an average of 2.75 articles, which is the initial stage. 2015 has become a turning point for the increase of the topic’s attention. 2015–2019 has an average of 4.8 articles, which is a smooth development stage. Among them, the TC (total citations) in 2015 reached 411, and the C/P (average citations per publication) was as high as 82.20, indicating that the studies published in that year had high academic impact. 2020 present, the average number of articles is 7.67, which is a high growth stage. The TP (11) and TC (220) in 2021 show that some of the research results in that year had far-reaching impacts in the academic community, thus enhancing the overall impact of the topic. After 2022, although the publication volume is still high, its academic impact declines significantly, with the TC of only 18 in 2023 and the C/P dropping to 2.00, indicating that the citation rate of the research in this field and its academic impact are declining. 2022–2024, the average number of publications is 8.33, which is a high growth phase. The data for 2024–2025 suggests that the field may be in a phase of adjustment, especially as 2025 is only incomplete data as of March, and future trends still need to be further observed, and in the future the topic may further consolidate its position in the academic community as more research is conducted and new publications appear.

3.1.3 Publications by authors

From the data in Tables 8, 9 the academic output of this research field has a certain degree of authorship concentration, and at the same time shows the trend of multi-authorship cooperation. These authors come from different countries and regions, showing that this research field is characterized by internationalization. Among the prolific authors, Pace, C.S. (Italy), Muzi, S. (Italy), and Tenenbaum, H.R. (UK) are the most influential researchers

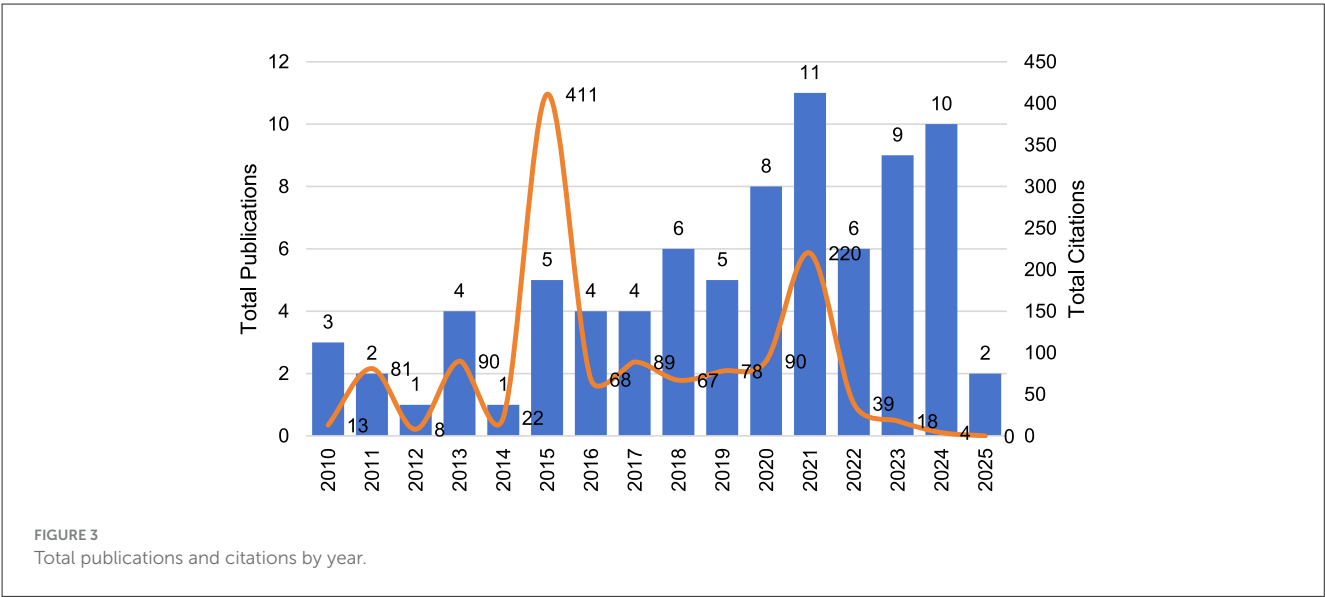


TABLE 7 Year of publication.

Year	TP	NCP	TC	C/P	C/CP	h	g
2010	3	2	13	4.33	6.50	2	3
2011	2	2	81	40.50	40.50	2	2
2012	1	1	8	8.00	8.00	1	1
2013	4	4	90	22.50	22.50	3	4
2014	1	1	22	22.00	22.00	1	1
2015	5	5	411	82.20	82.20	4	5
2016	4	3	68	17.00	22.67	3	4
2017	4	3	89	22.25	29.67	3	4
2018	6	5	67	11.17	13.40	3	6
2019	5	5	78	15.60	15.60	5	5
2020	8	7	90	11.25	12.86	5	8
2021	11	10	220	20.00	22.00	7	11
2022	6	6	39	6.50	6.50	3	6
2023	9	6	18	2.00	3.00	3	3
2024	10	3	4	0.40	1.33	1	1
2025	2	0	0	0.00	0.00	0	0
Total	81	63	1,298	16.02	20.6		

TP, total number of publications; NCP, number of cited publications; TC, total citations; C/P, average citations per publication; C/CP, average citations per cited publication; h, h-index; g, g-index.

with two publications in the field and high citation counts (TC all more than 50), respectively. In addition, a number of authors are from the UK, USA, Italy, India, Iran, Japan and other countries, showing that the research in this field has some internationalization characteristics, but is still dominated by scholars from Europe and the USA. Some authors (e.g., Aier, A. and Akard, T.F.) have higher average citation rates (C/P), indicating that their research may have greater academic influence.

Table 9 reflects the co-authorship of papers in this field, with relatively few single-authored papers (8.64%) and the

highest percentage of co-authored papers with 3–5 authors (14.81%–17.28%), suggesting that this research topic favors the collaboration of small research teams. The percentage of papers with more than 10 authors is extremely low (1.23%–1.23%), indicating that large-scale collaborative research is less likely to occur in this area. As the number of authors increases, the number of documents increases and then decreases, showing an inverted U-shaped distribution trend. These data reflect that moderate-sized teams (e.g., 3–6 people) in academic collaborations may be more conducive to efficiently

TABLE 8 Top 25 most productive authors.

Author's name	Affiliation	Country	TP	NCP	TC	C/P	C/CP	<i>h</i>	<i>g</i>
Pace, C.S.	University of Genoa	Italy	3	3	51	17	17	3	0
Aznar, A.	Kingston University	United Kingdom	2	2	69	34.5	34.5	2	0
Muzi, S.	University of Genoa	Italy	2	2	14	7	7	2	0
Tenenbaum, H.R.	University of Surrey	United Kingdom	2	2	69	34.5	34.5	2	0
Abeyweera, R.	University of Colombo	Sri Lanka	1	1	5	5	5	1	0
Abid, M.	Duke University	United States	1	0	0	0	0	0	1
Acierno, S.P.	Seattle Children's Hospital	United States	1	1	8	8	8	1	0
Aier, A.	St. John's Medical College	India	1	1	13	13	13	1	0
Akard, T.F.	Vanderbilt University	United States	1	1	4	4	4	1	0
Al-Ammouri, I.	The University of Jordan	Jordan	1	0	0	0	0	0	1
Ambad, R.	Shalinitai Meghe Institute of Medical Sciences	India	1	0	0	0	0	0	1
Aminabadi, N.A.	Tabriz University of Medical Sciences	Iran	1	1	20	20	20	1	0
An, P.	Southern University of Science and Technology	China	1	0	0	0	0	0	1
Anagnostou, E.	University of Toronto	Canada	1	1	41	41	41	1	0
Ananthanarayan, S.	University of Oldenburg, Oldenburg	Germany	1	1	34	34	34	1	0
Anderson, E.	StoryAidEU	United Kingdom	1	1	3	3	3	1	0
Angelini, J.	University of Modena and Reggio Emilia	Italy	1	1	1	1	1	1	1
Annesi-Maesano, I.	/	France	1	1	35	35	35	1	0
Anyango, H.O.	African Population and Health Research Center (APHRC)	Kenya	1	1	12	12	12	1	0
Anziom, B.	L'Association pour la Traduction, l'Alphabétisation, et le Développement Holistique de l'Etre Humain (ASTRADHE)	Cameroon	1	1	8	8	8	1	0
Asano, K.	Tohoku University	Japan	1	1	31	31	31	1	0
Asano, M.	Tohoku University	Japan	1	1	31	31	31	1	0
Asiri, M.S.	Sulu State College	Philippines	1	0	0	0	0	0	1
Asta, L.	University of Modena and Reggio Emilia	Italy	1	1	1	1	1	1	1
Avello, E.	Pontificia Universidad Católica de Chile	Chile	1	1	13	13	13	1	0

TP, total number of publications; NCP, number of cited publications; TC, total citations; C/P, average citations per publication; C/CP, average citations per cited publication; h, h-index; g, g-index.

producing research results. The high number of citations not only represents the influence of the authors' research results in the academic community, but may also mean that these studies have made important contributions to the development of the field.

3.1.4 Publications by countries

From the data in Table 10 and Figure 4 of the World Research Output Map, the academic contributions in this research field are mainly concentrated in the United States (88 articles), Italy (33 articles), and Canada (30 articles), which shows that North

TABLE 9 Frequency of author collaboration.

Author count	Frequency	Percentage (%)
1	7	8.64
2	10	12.35
3	12	14.81
4	14	17.28
5	9	11.11
6	13	16.05
7	6	7.41
8	4	4.94
9	3	3.70
10	1	1.23
14	1	1.23
31	1	1.23
Total	81	100

No author is listed.

American and European countries dominate the research in this field. The United Kingdom (22 articles) and China (17 articles) also have high academic outputs, indicating that the topic has received more attention in English-speaking countries and some Asian countries. The relatively high research outputs from Asian countries such as Japan (16 articles), India (14 articles), and the Philippines (13 articles) indicate that the theme has a certain degree of academic impact on a global scale. In addition, emerging countries such as Chile (11 articles), and Sri Lanka (10 articles) have also contributed to this research, reflecting the gradual participation of scholars from developing countries in this field. The research map further validates this distribution trend, with the United States having the most significant research output globally, and growing research contributions from Europe, Asia, and some Latin American and African countries. Overall, the research field is dominated by Europe and the United States, while showing a trend of multinational participation and global cooperation, which may further expand to developing countries in the future. Gray-marked countries have relatively low or zero output in terms of scientific publications, and they need more support and attention to promote their scientific research endeavors.

3.1.5 Top keywords

This study focuses on the relationship of parental emotional expression on children’s oral narrative ability, and explores related research directions through the keyword co-occurrence analysis map in Figure 5. The map presents a multifaceted and interrelated conceptual network, with lines indicating the co-occurrence of keywords, and the thickness of the lines may indicate the strength of the co-occurrence. In exploring the relationship between “parents’ emotional expression” and “children’s oral narrative ability,” scholars need to pay attention to a series of closely related potential variables, such as children’s emotional state,

TABLE 10 Top 20 Countries contributed to the publications.

Country	Freq
USA	88
Italy	33
Canada	30
UK	22
China	17
Japan	16
India	14
Denmark	13
Philippines	13
Chile	11
Sri Lanka	10
Romania	9
France	8
Germany	8
Iran	7
Netherlands	7
Greece	6
Kenya	6
Australia	5
Finland	5

cognitive ability, family environment, parent-child relationship, etc. (Figure 5). The results of the analysis show that “Storytelling” is a core concept in this field, and is closely related to “Attachment,” “Family” and “Children,” indicating the importance of the family context in the development of children’s narrative abilities. Attachment theory suggests that parental emotional support not only affects children’s sense of security, but may also enhance their narrative expression through emotional regulation and language input (Bang, 2009; Fitton, 2012; Polovina, 2005). In addition, the strong association between “Language” and “Preschoolers” highlights the importance of oral narrative abilities in early language development, suggesting that positive feedback and emotional value from parents in everyday interactions may promote children’s narrative integrity, emotional expression, and story organization. The study also revealed the potential role of “Digital Storytelling” and “Educational Technology,” suggesting that future research could explore how parents can use digital media to enhance children’s narrative abilities and optimize their learning experiences with emotional support. The relationship between parental emotional expression in children’s narrative development is an emerging topic in this topic, and it is also suggested that future research could explore diverse intervention approaches in conjunction with digital technology to enhance children’s narrative expression, especially in culturally diverse and special populations of children, to further validate the mechanism of its influence.

Country Scientific Production

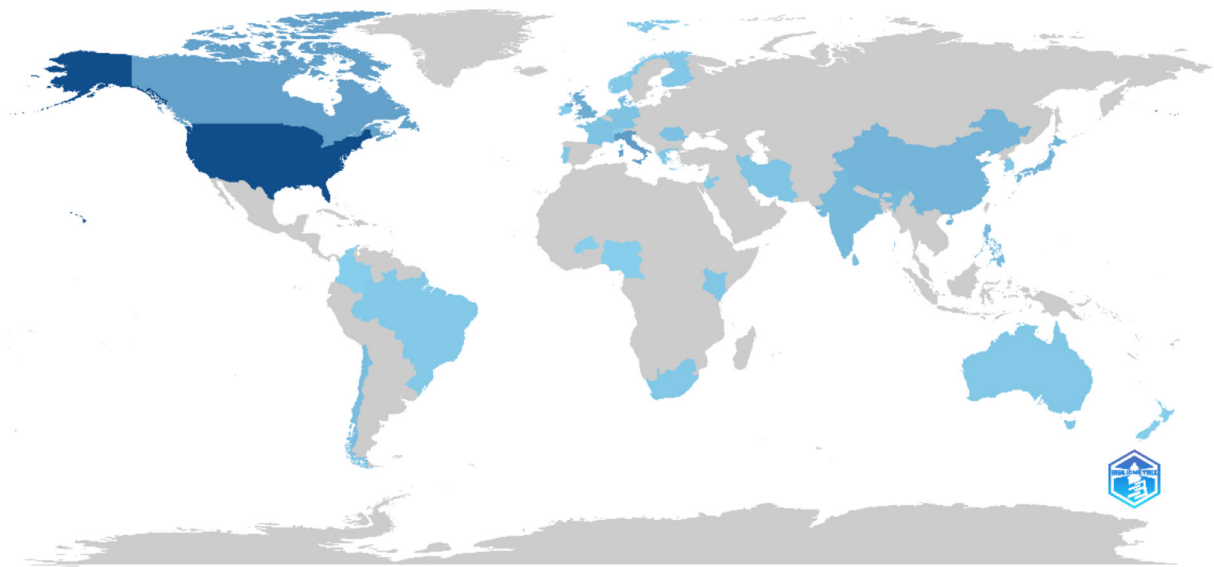


FIGURE 4
Worldwide scientific production map indexed by Scopus.

3.2 Thematic analysis

Thematic analysis was used to analyze the topics in this study. In the thematic analysis process, we used the six selected articles in **Section 2.3** as our analytical units. These six articles are all published in the Asian region, and focusing on research in the Asian region can explore research goals more deeply and targetedly. First, each study was read in detail to extract specific descriptions and data regarding parental emotional expression and children's oral narrative abilities. Similar descriptions and data were then grouped under the same latent theme. Based on the above narrative theme integration process, this study ultimately extracts the following four core themes: family literacy environment, digital devices, parent-child interaction, and cultural background. The following subsections describe each dimension in detail.

3.2.1 The impact of family literacy environment

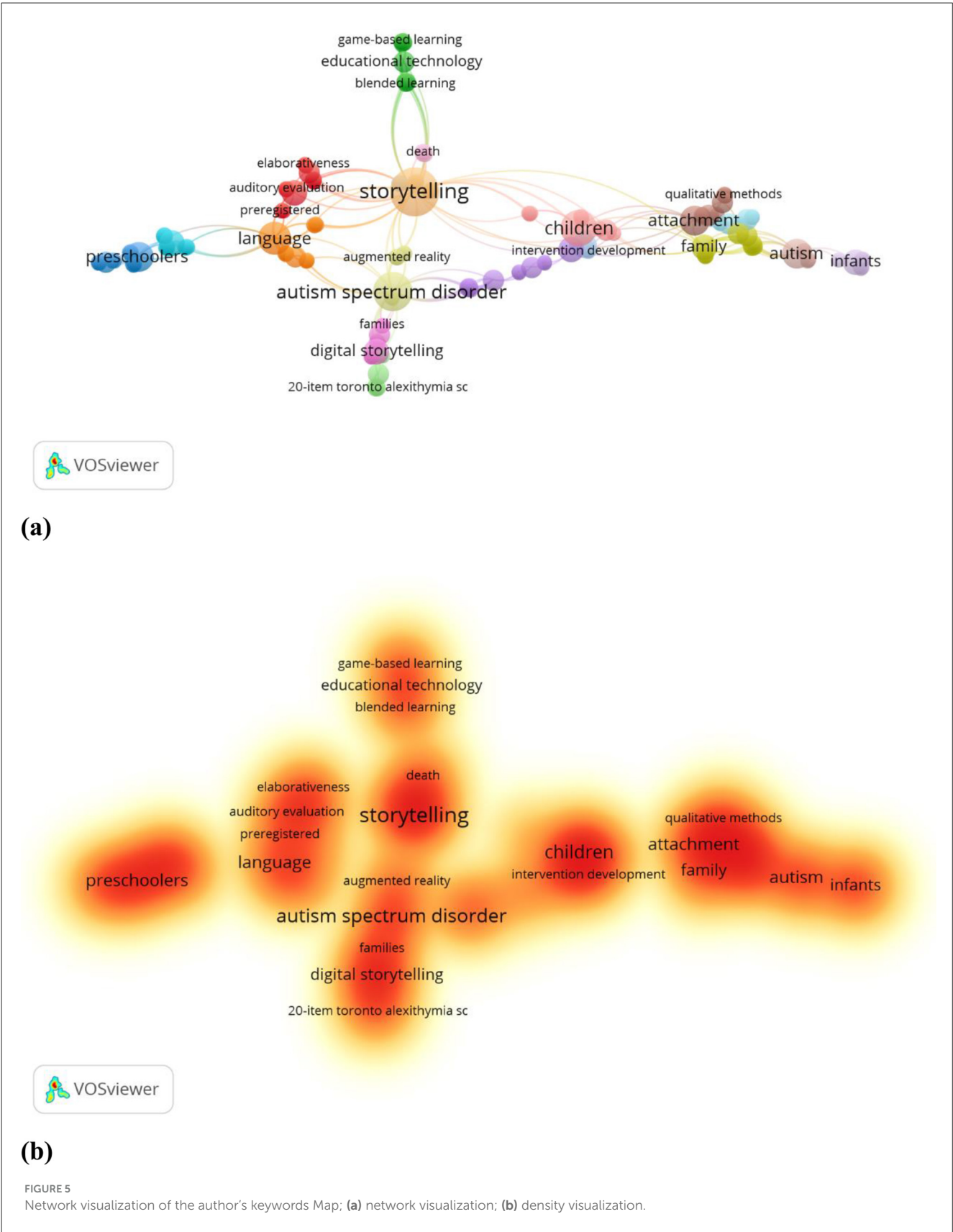
In the selected articles, the article "Association between home literacy environment and eye movements during storybook reading: implications for vocabulary and word reading in Chinese preschoolers," [Wen et al. \(2024\)](#) employed advanced eye-tracking technology combined with quantitative analysis to systematically examine how multiple key dimensions of the family literacy environment (such as language stimulation and interactive reading) influence children's eye movement patterns during storybook reading. This impact was particularly evident in their attention to text regions and their ability to recognize vocabulary and words. The findings revealed that interactive reading, as a core element of the family literacy environment, significantly enhanced children's attention to text and word reading proficiency ($\beta = 0.37, p < 0.02$).

These conclusions resonate with multiple studies on family literacy environments in non-Asian regions, further substantiating the profound influence of the home environment on the development of early reading abilities in children.

Family literacy resources can be divided into traditional literacy resources and electronic product resources. [Noguerón-Liu and Driscoll \(2021\)](#) collected data through interviews, home visits, and children's reading assessments and used a case study approach for their analysis, the results suggest that mothers in bilingual families access literacy resources from schools (e.g., summer programs), community agencies (e.g., libraries), and family networks to improve children's expressive abilities and adaptation to transnational life.

Formal literacy activities refer to structured instructional methods designed to develop children's reading and writing abilities, such as word recognition, phonics, and reading comprehension ([Spencer et al., 2013](#)). Formal literacy activities can help children acquire literacy abilities and promote their language development, which indirectly improves progress in narrative expression. [Brown et al. \(2014\)](#) study showed that narrative interventions improved story retelling in at-risk children, which was associated with literacy activities. [Spencer et al. \(2020a,b\)](#) showed that bilingual instruction enhances preschoolers' oral language abilities, which may include narrative abilities. In addition, a study by [Kirby et al. \(2021\)](#) demonstrated that oral narrative instruction enhanced kindergarten writing, which was closely related to literacy activities.

Informal literacy activities are unstructured daily activities at home, such as reading storybooks together, telling stories, playing word games, or engaging in narrative-related conversations. Informal literacy activities positively influence children's narrative



abilities by enhancing language development, comprehension, and interest in reading. [Silinskas et al. \(2020\)](#) found that parents' involvement in family literacy activities, such as storytelling, was

effective in improving children's literacy abilities and increasing independent reading abilities resulting in stronger narrative abilities. However, storytelling activities may have different effects

depending on the frequency. Bitetti and Hammer (2016) found in a study of Spanish-English bilingual children that mothers' frequent reading of storybooks to their children significantly improved the quality of the children's narratives, and that co-reading activities not only provided linguistic input, but also helped the children to understand the narrative sequence and character development through pictures and storylines. Therefore, the quality of interaction during storytelling was also emphasized as being extremely important in this study as well. Miller-Goldwater et al. (2024) noted that co-reading books can support memory integration in children, especially when the book engages through embedded questions for two individuals, and when both individuals integrate facts while reading. Additionally, in that study it was noted that mothers' storytelling and children's less frequent storytelling (once a month or once a week on average) did not have a significant effect on narrative abilities.

3.2.2 The impact of digital devices

In the selected articles, the article "Digital tools for direct assessment of autism risk during early childhood: a systematic review" by Debarati et al. (2024) systematically reviews various tools and methods for assessing early childhood autism risk using digital technology. The application of digital technology in the field of children's health assessment is becoming increasingly widespread. With the development of new technologies, children's use of digital devices is increasing, which brings new changes to the home learning environment. The intrusion of digital devices, as opposed to traditional literacy activities, has given children flexibility of place, time, and element in the language learning process. In the critical field of children's oral narrative development, numerous studies non-Asian regions have focused on the impact of digital devices.

An analysis by Verhoeven et al. (2020) showed that computer-assisted literacy interventions in preschool and kindergarten environments could improve phonological awareness and reading-related abilities. Additionally, Teepe et al. (2016) showed that technology-enhanced storytelling using digital devices had a positive impact on parent-child interactions and children's productive vocabulary knowledge. A study by Neumann (2016) of 2–4 year olds showed that the frequency of writing with tablets was associated with print awareness, print knowledge, and sound knowledge positively, and non-digital writing activities were positively associated with emergent skills in the home. However, e-reading was not correlated with children's literacy abilities. However, excessive or inappropriate use of digital media can also have negative effects. Passive consumption of digital media (e.g., prolonged television viewing) may be detrimental to narrative development, especially in the absence of interaction. Research has shown that background television noise distracts children and reduces quality interactions with caregivers, which affects language development (Pedrotti et al., 2024). Schwarzer et al. (2022) in the LIFE Child study analyzed 296 children aged 2–5 years and found that more than 1 h of daily screen time was significantly associated with lower percentile rankings in cognitive, language, and socio-emotional skills, with a particularly pronounced decline in language abilities ($b = -12.88$, $p < 0.01$). The impact of digital

media activities on children's narrative abilities is not monolithic, but dependent on the mode and context of use. Highly interactive and educational media use may promote narrative abilities, whereas excessive passive consumption may have a negative impact. Parents play a key role in selecting content and engaging in interactions to maximize benefits and minimize risks.

Digital devices, such as tablets, smartphones, and multimedia storytelling apps, have become important mediums in modern homes and educational settings and are widely used in children's language development process. Digital devices often contain rich multimedia content such as video stories, animated storybooks, and interactive language games. These multisensory stimuli not only help children better understand story content and language structures (Takacs et al., 2015), but also provide vivid and interactive learning experiences that help children expand their vocabulary and improve their syntactic understanding, enriching their oral narrative abilities. Meanwhile, many digital storytelling apps help children clarify the beginning, development and ending of a story through structured templates and appropriate prompts, acting as Digital scaffolding (DSC), which effectively improves children's mastery of narrative structure, and thus enhances the coherence and logic of their narratives (Spencer et al., 2020a,b).

The widespread use of digital devices may also have some negative effects on children's narrative abilities. First, overreliance on digital devices may reduce the frequency of direct language interactions between children and adults, undermining the positive effects of shared reading or face-to-face narrative activities on children's language development (Strouse and Ganea, 2017). Second, while structured narrative templates in digital devices can help children construct stories, their long-term use may limit children's ability to create and express themselves freely, making them more inclined to rely on established templates and lacking spontaneous creative narratives (Kucirkova et al., 2015). In addition, rich interactive elements (e.g., animation and sound effects) in digital devices may distract children's attention from deeper understanding of narrative content and linguistic details by focusing more on the surface features of the story (Bus et al., 2015), which reduces the quality of narrative expression.

3.2.3 The impact of parent-child interaction

In the selected articles, Hsieh (2023) in the article "This is My Favorite Book": Exploring Children's Perspectives on Books and Reading" utilized both qualitative and quantitative research to reveal children's desire for shared reading with adults and their wish for greater participation and choice. This highlights the crucial role of parent-child interaction in reading and reminds parents and educators to foster children's reading initiative. Another study, Hong et al. (2022) in "Meta-Analysis of RCTs of Technology-Assisted Parent-Mediated Interventions for Children with ASD" evaluated the effectiveness of technology-assisted parent-mediated interventions for children with autism. It found these interventions had some impact on emotion recognition but showed limited effectiveness in social communication and language development, suggesting interventions need to be more targeted and personalized. Also, numerous studies non-Asian regions have further confirmed the profound impact of parent-child interaction

patterns and targeted interventions on children's reading and social-language development.

Children's cognitive and language development relies on the process of interaction with adults or more experienced peers. High-quality parent-child language interactions can provide children with rich and varied language input and promote the development of vocabulary building and language comprehension (Hart and Risley, 1995). Rowe (2012) study showed that interactions in which parents use a more varied vocabulary with appropriate language extensions can lead to richer narratives and more precise vocabulary use by children. In addition, positive emotional support and encouragement not only create a favorable emotional climate for children, but also significantly increase their initiative and self-confidence in participating in oral narrative activities, thus effectively enhancing narrative motivation and narrative quality (Fivush and Nelson, 2006). In the specific narrative process, parents can build up "narrative scaffolding" through open-ended questions, detailed descriptions, etc., to help children gradually grasp the structure of the story (beginning, development, climax, and end), which effectively enhances their narrative logic and coherence (Haden et al., 1997).

However, the low quality of parent-child interactions (e.g., single language input, lack of or oversimplified responses) may result in children's expressive language abilities being limited by a monotonous vocabulary, poor content, and a lack of logic in their narratives (Hoff, 2006). In addition, low quality emotional interactions (e.g., negative parental feedback or neglectful behaviors) may result in low narrative motivation or even narrative anxiety, leading to avoidance of or passive participation in oral narrative activities and weakening the development of narrative ability (Laible, 2011). At the same time, if parents are too dominant in their interactions (e.g., frequent instructions or correction of errors), children may have difficulty organizing stories on their own, limiting their narrative creativity and spontaneity (Tamis-LeMonda et al., 2001).

3.2.4 The impact of cultural background

In addition to the hot topics of research mentioned above, there are significant differences in the oral narrative styles of children from different cultural backgrounds. Among the selected articles, Sebastian et al. (2019) in "Parents, Television, and Children's Emotional Expressions: A Cross-Cultural Multilevel Model" conducted cross-cultural comparisons to examine the effects of parenting styles and television media on children's emotional expression, highlighting cultural differences and underscoring the role of cultural context in children's emotional socialization. This study suggests that research interventions should account for cultural factors. Anita et al. (2011) in "Development of the language subtest in a developmental assessment scale to identify Chinese preschool children with special needs" focused on developing language assessment tools for children with special needs within the Chinese cultural context. Through rigorous procedures, they ensured the tool's cultural adaptability and measurement validity, providing a reference for assessment tool development in China and across cultures. Furthermore, research non-Asian regions has also examined the

relationship between cultural background and children's oral narrative abilities.

Children in European and American cultural settings tend to have an individualistic orientation, with more emphasis on personal emotions and individual experiences in their narratives. Children from East Asian cultures, on the other hand, showed a stronger collectivist orientation, focusing more on interpersonal interactions and social norms in their narrative content (Minami, 2008). This difference further influences children's choices of story structure, character relationships, and plot expression. In addition, parental interaction styles also show differences depending on cultural background and significantly affect children's narrative expression. For example, North American and European families are more likely to adopt an elaborative style, which provides rich emotional experiences and detailed descriptions in their interactions, thus effectively enhancing children's narrative structure and language complexity (Fivush and Wang, 2005). In contrast, Asian families have a more conservative style of interaction, with parents focusing more on emphasizing social values and tending to adopt a concise narrative language style (Wang and Leichtman, 2000).

4 Discussion

This study combines bibliometric and thematic analyses to systematically explore the influence of parental emotional expression on children's oral narrative ability, and reveals the future development trend in this field based on existing research. The findings not only verified the key role of parental emotional expression in children's language development, but also revealed the research hotspots, international cooperation patterns and interdisciplinary integration trends in this field.

Bibliometric analyses show that the field is moving toward interdisciplinary convergence, and the combination of artificial intelligence, psychology, and neuroscience is driving more accurate assessment of children's oral narrative ability and intervention strategies (Tables 4–6). Publication Trends show that the field experienced high-impact bursts in 2015 and 2021, but the declining citation rates in recent years may reflect changes in research hotspots, uneven research quality, or fluctuations in the field's attention (Table 7, Figure 3). In terms of academic output, the internationalization of this research topic has gradually increased, with European and American scholars still dominating, but research contributions from Asia and emerging countries are increasing, and the trend of global cooperation is becoming more and more significant (Tables 8, 9). It is noteworthy that the field is still mainly dominated by Europe and the United States, but there has been a tendency to expand to developing countries (Table 10, Figure 4). In addition, the present study demonstrated that parental emotional expression, parent-child interaction, cultural background, and technological interventions collectively influence children's oral narrative ability, but existing studies have paid less attention to the interaction of these factors (Figure 5).

Thematic analysis further validated the relationship between parental emotional expression on children's narrative development. Parental emotional expression not only directly affects children's oral narrative ability, but also displays a multilevel mechanism

of action moderated by the family literacy environment, the use of digital devices, parent-child interactions, and the cultural background. The family literacy environment provides children with initial language stimulation and narrative experiences, digital devices enrich children's language input and output channels with emerging media, the quality of parent-child interactions directly affects children's emotional experiences and narrative motivation, and the cultural background shapes children's narrative styles and thematic preferences. Together, these factors determine the breadth and depth of the influence of parental emotional expression on children's oral narrative ability.

Numerous studies have shown that a rich family literacy environment positively contributes to children's language development and narrative ability. Specifically, the number of accessible reading materials in the home, the frequency of shared reading activities, and positive parental guidance of text in daily life significantly enhance children's vocabulary and syntactic complexity (Dahl Leonard et al., 2025; Galea, 2024). On top of this, positive emotional expressions, such as enthusiasm and empathy, displayed by parents during the reading process further stimulated children's interest and willingness to participate in narrative activities. However, poor literacy resources and a negative emotional climate may inhibit children's self-expression during the narrative process, limiting the development of their narrative ability.

At the same time, the widespread use of digital media has brought about complex and contradictory influences on the construction of children's narrative ability. On the one hand, digital storytelling apps with high interactivity and rich content can facilitate children's comprehension and retelling of storylines through multichannel sensory stimulation (Choi et al., 2025; Ye et al., 2024). In particular, children's narrative motivation and confidence in expression may be significantly enhanced when parents accompany children with positive emotions in the use of digital devices. On the other hand, over-reliance on or inappropriate use of digital devices may diminish direct oral communication and emotional interactions between children and their parents, thereby reducing high-quality oral input and ultimately having an inhibitory effect on narrative ability (Sundqvist et al., 2024; Mills, 2020). These results suggest that the way digital technology is used and its mode of integration with parent-child interactions in family contexts has a key moderating role in the development of children's narrative ability.

In addition, parent-child interactions also play a non-negligible role in the formation of children's narrative ability. High-quality interactions, such as empathic listening, extended questioning, and positive feedback, not only facilitate children's use of richer linguistic resources in their narratives, but also contribute to the construction of more coherent and logical story structures (Rowe, 2022; Gardner-Neblett and Iruka, 2024). Positive emotional expression further enhances complexity in children's linguistic output and increases their sense of self-efficacy in narrative activities. In contrast, the absence of emotional support or excessive parental dominance in narratives may lead to a decline in children's

narrative motivation, which in turn leads to homogenization of narrative content and a lack of emotional depth (Mills, 2020; Gonzalez, 2025). Thus, parent-child interactions not only affect the quantitative level of narrative ability, but also have a profound impact on the level of narrative quality and emotional expression.

Further, the cultural background, as an important part of the macro-socialization mechanism, also plays a profound role in shaping children's narrative styles. In European and American cultures, parents often encourage children to narrate their personal emotional experiences in a self-centered manner, an orientation that enriches narrative content and enhances the depth of emotional expression (Gardner-Neblett and Iruka, 2024). In East Asian cultures, parents emphasize social responsibility and collective identity, and children's narratives tend to focus on social relationships and norm-following (Gonzalez, 2025). Cultural background not only shapes the form and frequency of parental emotional expression, but also indirectly influences the developmental path of children's narrative ability through the socialization process. This cultural difference suggests that cross-cultural variability and its underlying mechanisms need to be fully considered when understanding the influence of family factors on children's narrative ability.

In summary, the family literacy environment, digital devices, parent-child interactions, and cultural background play multilayered and intertwined roles in the development of children's narrative ability. Although existing studies have provided rich empirical evidence, there is still a lack of theoretical integration, limited exploration of mechanisms, and cross-cultural comparisons in this area. In order to further deepen the understanding of the developmental mechanisms of children's narrative ability and to enhance the application value of the research, future studies can explore the following directions. First, interdisciplinary cooperation should be strengthened to promote the integration and application of neuroscience and psychology in order to optimize the assessment tools and intervention methods for children's language development. Second, it is necessary to improve the rigor of research methods, enhance the reproducibility of data, and summarize and integrate existing research results through systematic reviews, so as to provide a more solid foundation for theoretical construction. Third, promoting international cooperation is also an important direction for the future, especially in expanding the coverage of research on developing countries and culturally diverse populations, so as to enrich the global perspectives of the field and enhance the generalizability of the research results. In addition, combining innovative interventions with digital technology and exploring intelligent and personalized programs to promote children's language development are expected to open up new paths of practice. Finally, research on the integrated mechanism of emotional expression, parent-child interaction, and cultural factors needs to be further deepened to optimize family education strategies to more effectively support the development of children's oral narrative abilities. Through the above multi-dimensional expansion, future research is expected to make more far-reaching progress in theory deepening and practical application, contributing to the construction of a scientific guidance system for global children's language development.

5 Recommendations

Based on the bibliometric analysis, this study explores the research trends, internationalization, interdisciplinary integration, and future directions of the field. However, there are still some limitations, such as limited sample coverage, small sample size, and a single research method. Future research should expand the sample coverage, adopt mixed research methods, use dynamic analysis techniques and increase practical application studies to improve the scientific, systematic and practical impact of the research, so as to promote the development of the field toward a more precise, globalized and practice-oriented development.

6 Conclusion

This study screened the literature through PRISMA principles, and then explored the research trends, academic influence, international cooperation patterns, dynamic evolution of interdisciplinary development, and analysis of research hot topics in the field of the relationship between parental emotional expression and children's oral narrative ability based on bibliometric analysis and thematic analysis. The results indicated that the field experienced research peaks in 2015 and 2021, reflecting the stage-by-stage academic interest in the topic. However, the decline in citations in recent years may indicate the change of research hotspots and the adjustment of research methods, and the theoretical framework and research design need to be further optimized in the future to maintain the academic influence of the field. Further, the research on this topic is developing toward interdisciplinary integration, with the combination of psychology, neuroscience, artificial intelligence technology and other fields providing new perspectives on the assessment and intervention of children's oral narrative ability. Meanwhile, there is a growing trend of international cooperation. Although European and American scholars still dominate, research contributions from Asia, Latin America, and other developing countries have gradually risen, indicating that the field is evolving toward a globalized and diversified research pattern. In addition, this study suggests that parental emotional expression, parent-child interaction, cultural background and technological intervention jointly affect children's language development, but the interaction of these factors still needs to be further explored in order to optimize family education and intervention strategies.

Parents' emotional expression patterns directly affect children's oral narrative abilities; positive emotional expression promotes language richness, whereas negative emotional expression may inhibit language development. High-quality parent-child interactions effectively moderated the effects of parental emotional expression on children's narrative ability, while cultural context and technological interventions played an indirect role. Future research should continue to focus on children's language development from a cross-cultural perspective and explore the application of cutting-edge technologies, such as artificial intelligence and natural language processing, in children's language assessment and intervention, in order to optimize family education strategies and promote the continued development of the field.

Overall, this study reveals the current state of research in this field and the thematic categorization of this research topic, and provides a reference for future academic exploration. Future research should further expand in the areas of theoretical deepening, methodological innovation, cross-cultural research, and digital interventions, in order to promote the continued development of the field and provide more accurate and scientific guidance for children's language development.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

ZZ: Writing – original draft, Conceptualization, Methodology, Data curation, Investigation, Visualization, Software, Formal analysis, Writing – review & editing. SM: Supervision, Writing – review & editing. KB: Supervision, Writing – review & editing.

Funding

The author(s) declared that financial support was not received for this work and/or its publication.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

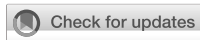
All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Allen, M. S., Kertoy, M. K., Sherblom, J. C., and Pettit, J. M. (1994). Children's narrative productions: a comparison of personal event and fictional stories. *Appl. Psychol.* 15, 149–176. doi: 10.1017/S0142716400005300
- Anita, W. M.-Y., Cynthia, L., Elaine, S. K.-L., Catherine, L. C.-C., and Grace, C. P.-S. (2011). Development of the language subtest in a developmental assessment scale to identify Chinese preschool children with special needs. *Res. Dev. Disabil.* 32, 297–305. doi: 10.1016/j.ridd.2010.10.004
- Bailey, A. L., Moughamian, A. C., Kelly, K. R., McCabe, A., and Huang, B. H. (2020). Leap-frog to literacy: maternal narrative supports differentially relate to child oral language and later reading outcomes. *Early Child Dev. Care* 190, 1136–1149. doi: 10.1080/03004430.2018.1521807
- Bang, H. J. (2009). Review of attachment theory and parenting. *Korean J. Woman Psychol.* 14, 67–91. doi: 10.18205/kpa.2009.14.1.004
- Benítez-Barrera, C., Behboudi, M., and Maguire, M. (2024). Neural oscillations during predictive sentence processing in young children. *Brain Lang.* 254:105437. doi: 10.1016/j.bandl.2024.105437
- Bitetti, D., and Hammer, C. S. (2016). The home literacy environment and the english narrative development of Spanish-English Bilingual children. *J. Speech Lang. Hear. Res.* 59, 1159–1171. doi: 10.1044/2016_JSLHR-L-15-0064
- Brown, J. A., Garzarek, J. E., and Donegan, K. L. (2014). Effects of a narrative intervention on story retelling in at-risk young children. *Topics Early Child. Spec. Educ.* 34, 154–164. doi: 10.1177/0271121414536447
- Bus, A. G., Takacs, Z. K., and Kegel, C. A. (2015). Affordances and limitations of electronic storybooks for young children's emergent literacy. *Dev. Rev.* 35, 79–97. doi: 10.1016/j.dr.2014.12.004
- Buschmann, A. (2020). "Einbezug der Eltern in die Sprachförderung," in *Sprachentwicklung – Diagnostik – Förderung im Kleinkind- und Vorschulalter*, eds S. Sachse, A.-K. Bockmann, and A. Buschmann (Heidelberg, Germany: Springer), 283–307. doi: 10.1007/978-3-662-60498-4_13
- Choi, N., Cyebukayire, P., and Choi, J. D. (2025). Tinker tales: interactive storytelling framework for early childhood narrative development and AI literacy. *arXiv [preprint] arXiv:2504.13969*. Available online at: <https://arxiv.org/abs/2504.13969>
- Clark-Whitney, E., and Melzi, G. (2023). Executive function and narrative language abilities in emergent bilingual preschoolers: an exploratory study. *Lang. Speech Hear. Serv. Sch.* 54, 584–599. doi: 10.1044/2022_LSHSS-21-00190
- Cohen, F., Schünke, J., Vogel, E., and 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. doi: 10.3389/fpsyg.2020.01282
- Conway, L., Cook, F., Cahir, P., Brown, S., Reilly, S., Gartland, D., et al. (2021). Children's language abilities at age 10 and exposure to intimate partner violence in early childhood: results of an Australian prospective pregnancy cohort study. *Child Abuse Negl.* 111:104794. doi: 10.1016/j.chiabu.2020.104794
- Custode, S., Bailey, J., Sun, L., Katz, L., Ullery, M., Messinger, D., et al. (2024). Preschool language environments and social interactions in an early intervention classroom: a pilot study. *J. Early Interv.* 46, 338–355. doi: 10.1177/10538151231176176
- Dahl Leonard, K., Hall, C., Cho, E., Capin, P., Roberts, G. J., Kehoe, K. F., et al. (2025). Examining the effects of family implemented literacy interventions for school aged children: A meta analysis. *Educ. Psychol. Rev.* 37:10. doi: 10.1007/s10648-025-09985-3
- Debarati, M., Supriya, B., Georgia, L. E., Vaisnavi, R., Jayashree, D., Hiba, I., et al. (2024). Digital tools for direct assessment of autism risk during early childhood: a systematic review. *Autism* 26, 6–31. doi: 10.1177/13623613221133176
- Dong, Y., Dong, W.-Y., Wu, S. X.-Y., and Tang, Y. (2020). The effects of home literacy environment on children's reading comprehension development: a meta-analysis. *Kuram Uygul. Egit. Bilim.* 20, 63–82. doi: 10.12738/jestp.2020.2.005
- Fang, S., Huang, J., Curley, J., and Birkenmaier, J. (2018). Family assets, parental expectations, and children educational performance: an empirical examination from China. *Child. Youth Serv. Rev.* 87, 60–68. doi: 10.1016/j.childyouth.2018.02.018
- Fekonja-Pekljaj, U., Marjanović-Umek, L., and Kranjc, S. (2010). Children's storytelling: the effect of preschool and family environment. *Eur. Early Child. Educ. Res. J.* 18, 55–73. doi: 10.1080/13502930903520058
- Fish, M., and Pinkerman, B. (2003). Language skills in low-SES rural Appalachian children: normative development and individual differences, infancy to preschool. *J. Appl. Dev. Psychol.* 23, 539–565. doi: 10.1016/S0193-3973(02)00141-7
- Fitton, V. (2012). Attachment theory: history, research, and practice. *Psychoanal. Soc. Work* 19, 121–143. doi: 10.1080/15228878.2012.666491
- Fivush, R., and Nelson, K. (2006). Parent-child reminiscing locates the self in the past. *Br. J. Dev. Psychol.* 24, 235–251. doi: 10.1348/026151005X57747
- Fivush, R., and Wang, Q. (2005). Mother-child conversations of emotionally salient events: exploring the functions of emotional reminiscing in European-American and Chinese families. *Soc. Dev.* 14, 473–495. doi: 10.1111/j.1467-9507.2005.00312.x
- Fuller, M. G., Vaucher, Y. E., Bann, C. M., Das, A., and Vohr, B. R. (2019). Lack of social support as measured by the Family Resource Scale Screening Tool is associated with early adverse cognitive outcome in extremely low birth weight children. *J. Perinatol.* 39, 1546–1554. doi: 10.1038/s41372-019-0462-2
- Galea, C. (2024). *Bookworm Babies 'Do Better in Kindy'*. The Australian. Available online at: <https://www.theaustralian.com.au/education/global-study-shows-benefits-of-reading-to-babies-from-birth/news-story/ad017632964eef4573682fd2a9bc8f8e> (Accessed November 17, 2024).
- Gardner-Neblett, N. (2022). What predicts oral narrative competence among African American children? Exploring the role of linguistic and cognitive skills. *J. Speech Lang. Hear. Res.* 65, 2931–2947. doi: 10.1044/2022_JSLHR-22-00002
- Gardner-Neblett, N., and Iruka, I. U. (2024). Becoming fictional storytellers: African American children's oral narratives. *Child Dev.* 94, 45–60. doi: 10.1111/cdev.14075
- Göksun, T., Aktan-Erciyes, A., Karadöller, D. Z., and Demir-Lira, O. E. (2025). The multifaceted nature of early vocabulary development: connecting children's characteristics with parental input types. *Child Dev. Perspect.* 19, 30–37. doi: 10.1111/cdep.12524
- Gonzalez, J. E. (2025). *Cultural Influences in Children's Narratives*. Bilingualistics. Available online at: <https://bilingualistics.com/cultural-influences-in-childrens-narratives/> (Accessed August 20, 2025).
- González, L., Popovic, M., Rebagliato, M., Estarlich, M., Moirano, G., Barreto-Zarza, F., et al. (2024). Socioeconomic position, family context, and child cognitive development. *Eur. J. Pediatr.* 183, 2571–2585. doi: 10.1007/s00431-024-05482-x
- Haden, C. A., Haine, R. A., and Fivush, R. (1997). Developing narrative structure in parent-child reminiscing across the preschool years. *Dev. Psychol.* 33, 295–307. doi: 10.1037/0012-1649.33.2.295
- Hart, B., and Risley, T. R. (1995). *Meaningful Differences in the Everyday Experience of Young American Children*. Baltimore, MD: Paul H. Brookes Publishing Co.
- Hidayat, N., and Fitri, N. (2021). Literacy theories of children's language development. *5th International Conference on Early Childhood Education (ICECE 2020)* (Atlantis Press). doi: 10.2991/assehr.k.210322.031
- Hoff, E. (2006). How social contexts support and shape language development. *Dev. Rev.* 26, 55–88. doi: 10.1016/j.dr.2005.11.002
- Holloway, D., Green, L., and Livingstone, S. (2013). *Zero to Eight: Young Children and their Internet Use*. LSE Media & Communications.
- Hong, J. P., Kannan, K., Shashidhara, M., Preeti, K., Richard, K., Paul, R., et al. (2022). Meta-analysis of RCTs of technology-assisted parent-mediated interventions for children with ASD. *J. Autism Dev. Disord.* 52:3325–3343. doi: 10.1007/s10803-021-05206-2
- Hou, Y. (2022). Research on children's cognitive education based on pathological linguistics. *Occup. Ther. Int.* 2022:8692238. doi: 10.1155/2022/8692238
- Hsieh, M.-F. (2023). 'This is My Favorite Book': exploring children's perspectives on books and reading. *Early Child. Educ. J.* 53, 489–505. doi: 10.1007/s10643-023-01585-6
- Hu, N., Li, S., Li, L., and Xu, H. (2022). The educational function of English children's movies from the perspective of multiculturalism under deep learning and artificial intelligence. *Front. Psychol.* 12:759094. doi: 10.3389/fpsyg.2021.759094
- Johnson, A., Veeramani, H., Shankar, N. B., Alwan, A. (2023). An equitable framework for automatically assessing children's oral narrative language abilities. *Interspeech* 2023, 4608–4612. doi: 10.21437/Interspeech.2023-1257
- Keltner, D., Sauter, D., Tracy, J., and Cowen, A. (2019). Emotional expression: advances in basic emotion theory. *J. Nonverbal Behav.* 43, 133–160. doi: 10.1007/s10919-019-00293-3
- Kim, J.-M. (2018). Expressive elaboration in personal narratives and fictional narratives in typically developing children. *Commun. Sci. Disord.* 23, 60–73. doi: 10.12963/csd.18473
- Kirby, M. S., Spencer, T. D., and Chen, Y. J. I. (2021). Oral narrative instruction improves kindergarten writing. *Read. Writing. Q.* 37, 574–591. doi: 10.1080/10573569.2021.1879696
- Kucirkova, N., Messer, D., Sheehy, K., and Fernández Panadero, C. (2015). Children's engagement with educational iPad apps: insights from a Spanish classroom. *Comput. Educ.* 71, 175–184. doi: 10.1016/j.compedu.2013.10.003
- Laible, D. (2011). Does it matter if preschool children and mothers discuss positive vs. negative events? Relations with mother-reported attachment, family emotional climate, and socioemotional development. *Soc. Dev.* 20, 394–411. doi: 10.1111/j.1467-9507.2010.00584.x

- Lee, D. H. (2019). Predictive power of conference-related factors on citation rates of conference papers. *Scientometrics* 118, 281–304. doi: 10.1007/s11192-018-2943-z
- Leyva, D., Berrocal, M., and Nolivos, V. (2014). Spanish-speaking parent-child emotional narratives and children's social skills. *J. Cogn. Dev.* 15, 22–42. doi: 10.1080/15248372.2012.725188
- Lim, W. M. (2021). Pro-active peer review for premier journals. *Ind. Mark. Manag.* 95, 65–69. doi: 10.1016/j.indmarman.2021.04.004
- Liu, X., and Zhao, W. (2024). The impact of maternal emotional regulation on children's emotional and health behaviors: the mediating role of family emotional expression and maternal age. *Am. J. Health Behav.* 48, 1232–1244. doi: 10.5993/AJHB.48.5.5
- Lucero, A. (2018). Oral narrative retelling among emergent bilinguals in a dual language immersion program. *Int. J. Biling. Educ. Biling.* 21, 248–264. doi: 10.1080/13670050.2016.1165181
- Madya, R., Malik, M., and Suparno, S. (2019). The impact of socioeconomic status (SES) on early childhood language development. *International Conference on Special and Inclusive Education (ICSIE 2018)* (Atlantis Press). doi: 10.2991/icsie-18.2019.26
- Masters, A. S., Hirsh-Pasek, K., Levine, D., and Golinkoff, R. M. (2022). "Parents matter: the cornerstone for children's cognitive and language development," in *The Cambridge Handbook of Parenting: Interdisciplinary Research and Application*, eds A. S. Morris, and J. Mendez Smith (Cambridge, UK: Cambridge University Press), 95–119. doi: 10.1017/9781108891400.007
- Miles, S., and Chapman, R. S. (2002). Narrative content as described by individuals with down syndrome and typically developing children. *J. Speech Lang. Hear. Res.* 45, 175–189. doi: 10.1044/1092-4388(2002/013)
- Miller-Goldwater, H. E., Williams, B. M., Hanft, M. H., and Bauer, P. J. (2024). Contributions of shared book reading to children's learning of new semantic facts through memory integration. *Early Child. Res. Q.* 68, 99–111. doi: 10.1016/j.ecresq.2024.04.005
- Mills, M. T. (2020). Perceptions of black children's narrative language. *Lang. Speech Hear. Serv. Sch.* 51, 1–15. doi: 10.1044/2020_LSHSS-20-00014
- Minami, M. (2008). Telling good stories in different languages: bilingual children's styles of story construction and their linguistic and educational implications. *Narrat. Inq.* 18, 83–110. doi: 10.1075/ni.18.1.05min
- Moher, D., Liberati, A., Tetzlaff, J., and Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ.* 339:b2535. doi: 10.1136/bmj.b2535
- Narimisei, J., Naeim, M., Imannezhad, S., Samian, P., and Sobhani, M. (2024). Exploring emotional intelligence in artificial intelligence systems: a comprehensive analysis of emotion recognition and response mechanisms. *Ann. Med. Surg.* 86, 4657–4663. doi: 10.1097/MS9.0000000000002315
- Neumann, M. M. (2016). Young children's use of touch screen tablets for writing and reading at home: relationships with emergent literacy. *Comput. Educ.* 97, 61–68. doi: 10.1016/j.compedu.2016.02.013
- Nguyen, T. Q. (2023). *A Dimensional Perspective on the Home Environment and Neurocognitive Systems in Children's Reading Development*. Vanderbilt University.
- Noguerón-Liu, S., and Driscoll, K. (2021). Bilingual families' perspectives on literacy resources and supports at home. *Read. Teach.* 75, 17–25. doi: 10.1002/trtr.2001
- Osborne, A., and Ahinkorah, B. O. (2024). The paternal influence on early childhood development in Africa: implications for child and adolescent mental health. *Child Adolesc. Psychiatry Ment. Health* 18:156. doi: 10.1186/s13034-024-00847-4
- Otero-Mayer, A., González-Benito, A., Gutiérrez-de-Rozas, B., and Expósito-Casas, E. (2025). Family involvement in early childhood education: a systematic review of its measurement. *Early Child. Educ. J.* doi: 10.1007/s10643-025-02024-4
- Pedrotti, B. G., Bandeira, D. R., and Frizzo, G. B. (2024). Context of digital media use in early childhood: factors associated with cognitive development up to 36 months of age. *Infant Behav. Dev.* 76:101963. doi: 10.1016/j.infbeh.2024.101963
- Pietropoli, I., and Gracia, P. (2025). Social inequalities in children's cognitive and socioemotional development: the role of home learning environments and early childhood education. *Res. Soc. Stratif. Mobil.* 97:101034. doi: 10.1016/j.rssm.2025.101034
- Polovina, N. (2005). Attachment theory: old and new approach. *Zbornik Inst. Pedagos. Istraz.* 37, 107–124. doi: 10.2298/ZIP10501107P
- Puglisi, M. L., Hulme, C., Hamilton, L. G., and Snowling, M. J. (2017). The home literacy environment is a correlate, but perhaps not a cause, of variations in children's language and literacy development. *Sci. Stud. Read.* 21, 498–514. doi: 10.1080/10888438.2017.1346660
- Qi, D., and Wu, Y. (2020). Family's social economic status and child educational outcomes in China: the mediating effects of parenting practices and children's learning attitudes. *Child. Youth Serv. Rev.* 118:105387. doi: 10.1016/j.childyouth.2020.105387
- Romeo, R. R., Leonard, J. A., Robinson, S. T., West, M. R., Mackey, A. P., Rowe, M. L., et al. (2018). Beyond the 30-million-word gap: children's conversational exposure is associated with language-related brain function. *Psychol. Sci.* 29, 700–710. doi: 10.1177/0956797617742725
- Romero-González, M., Lavigne-Cerván, R., Gamboa-Ternero, S., Rodríguez-Infante, G., Juárez-Ruiz de Mier, R., and Romero-Pérez, J. F. (2023). Active home literacy environment: parents' and teachers' expectations of its influence on affective relationships at home, reading performance, and reading motivation in children aged 6 to 8 years. *Front. Psychol.* 14:1261662. doi: 10.3389/fpsyg.2023.1261662
- Roostin, E. (2018). Family influence on the development of children. *Primary Edu J. Prim. Educ.* 2:1. doi: 10.22460/pej.v1i1.654
- Rowe, M. L. (2012). A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child Dev.* 83, 1762–1774. doi: 10.1111/j.1467-8624.2012.01805.x
- Rowe, M. L. (2022). *Parent-Child Interactions Support Early Language Development in Diverse Families*. Society for Research in Child Development. Available online at: <https://infantstudies.org/parent-child-interactions-support-early-language-development-in-diverse-families/> (Accessed May 15, 2022).
- Schwab, J. F., and Lew-Williams, C. (2016). Language learning, socioeconomic status, and child-directed speech. *Wiley Interdiscip. Rev. Cogn. Sci.* 7, 264–275. doi: 10.1002/wcs.1393
- Schwarzer, C., Grafe, N., Hiemisch, A., Kiess, W., and Poulain, T. (2022). Associations of media use and early childhood development: Cross-sectional findings from the LIFE Child study. *Pediatr. Res.* 91, 247–253. doi: 10.1038/s41390-021-01433-6
- Scott, A., Gillon, G., McNeill, B., and Kopach, A. (2022). The evolution of an innovative online task to monitor children's oral narrative development. *Front. Psychol.* 13:903124. doi: 10.3389/fpsyg.2022.903124
- Sebastian, S., Marie-Louise Mares, M., Anne Bartsch, B., and Maya, G. (2019). Parents, television, and children's emotional expressions: a crosscultural multilevel model. *J. Cross Cult. Psychol.* 50, 22–46. doi: 10.1177/0022022118806585
- Sénéchal, M., and LeFevre, J.-A. (2002). Parental involvement in the development of children's reading skill: a five-year longitudinal study. *Child Dev.* 73, 445–460. doi: 10.1111/1467-8624.00417
- Sher-Censor, E., and Yates, T. M. (2015). Mothers' expressed emotion and narrative coherence: associations with preschoolers' behavior problems in a multiethnic sample. *J. Child Fam. Stud.* 24, 1392–1405. doi: 10.1007/s10826-014-9946-z
- Silinskas, G., Sénéchal, M., Torppa, M., and Lerkanen, M.-K. (2020). Home literacy activities and children's reading skills, independent reading, and interest in literacy activities from kindergarten to grade 2. *Front. Psychol.* 11:1508. doi: 10.3389/fpsyg.2020.01508
- Sinclair, J., Jang, E. E., and Rudzicz, F. (2021). Using machine learning to predict children's reading comprehension from linguistic features extracted from speech and writing. *J. Educ. Psychol.* 113:1088. doi: 10.1037/edu0000658
- Spencer, T., Moran Wilson, M., Thompson, M., Petersen, D., and Restrepo, M. (2020b). Early efficacy of multitiered dual-language instruction: promoting preschoolers' spanish and english oral language. *AERA Open* 6:233285841989788. doi: 10.1177/2332858419897886
- Spencer, T. D., Kajian, M., Petersen, D. B., and Bilyk, N. (2013). Effects of an individualized narrative intervention on children's storytelling and comprehension skills. *J. Early Interv.* 35, 243–269. doi: 10.1177/1053815114540002
- Spencer, T. D., Petersen, D. B., Slocum, T. A., and Allen, M. M. (2020a). Large-scale replication of a storytelling intervention for preschoolers. *J. Speech Lang. Hear. Res.* 63, 1451–1472.
- Strouse, G. A., and Ganea, P. A. (2017). Parent-toddler behavior and language differ when reading electronic and print picture books. *Front. Psychol.* 8:677. doi: 10.3389/fpsyg.2017.00677
- Sundqvist, A., Majerle, N., Heimann, M., and Koch, F.-S. (2024). Home literacy environment, digital media and vocabulary development in early childhood. *J. Early Child. Lit.* 24, 1–20. doi: 10.1177/1476718X241257337
- Takacs, Z. K., Swart, E. K., and Bus, A. G. (2015). Benefits and pitfalls of multimedia and interactive features in technology-enhanced storybooks: a meta-analysis. *Rev. Educ. Res.* 85, 698–739. doi: 10.3102/0034654314566989
- Tamis-LeMonda, C. S., Bornstein, M. H., and Baumwell, L. (2001). Maternal responsiveness and children's achievement of language milestones. *Child Dev.* 72, 748–767. doi: 10.1111/1467-8624.00313
- Tang, Y., Li, S. M., Ma, L., Zheng, Y. X., and Li, Y. (2023). The effects of mother-father relationships on children's social-emotional competence: The chain mediating model of parental emotional expression and parent-child attachment. *Child. Youth Serv. Rev.* 155:107227. doi: 10.1016/j.childyouth.2023.107227
- Teepe, R. C., Molenaar, I., and Verhoeven, L. (2016). Technology-enhanced storytelling stimulating parent-child interaction and preschool children's vocabulary knowledge. *J. Comput. Assist. Learn.* 33, 123–136. doi: 10.1111/jcal.12169
- Torppa, M., Vasalampi, K., Eklund, K., and Niemi, P. (2022). Long-term effects of the home literacy environment on reading development: familial risk for dyslexia as a moderator. *J. Exp. Child Psychol.* 215:105314. doi: 10.1016/j.jecp.2021.105314
- Tulling, M. A. (2022). *Neural and Developmental Bases of Processing Language Outside The Here-and-Now* [Unpublished doctoral dissertation]. New York University.

- Tuyen, N. T. V., Jeong, S., and Chong, N. Y. (2017). "Learning human behavior for emotional body expression in socially assistive robotics." in *2017 14th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI)* (IEEE) pp. 45–50. doi: 10.1109/URAI.2017.7992882
- van der Wilt, F., Boerma, I., van Oers, B., and van der Veen, C. (2019). The effect of three interactive reading approaches on language ability: an exploratory study in early childhood education. *Eur. Early Child. Educ. Res. J.* 27, 566–580. doi: 10.1080/1350293X.2019.1634242
- van der Wilt, F., van der Veen, C., and Michaels, S. (2022). The relation between the questions teachers ask and children's language competence. *J. Educ. Res.* 115, 64–74. doi: 10.1080/00220671.2022.2029806
- Verhoeven, L., Voeten, M., Setten, E. V., and Segers, E. (2020). Computer-supported early literacy intervention effects in preschool and kindergarten: a meta-analysis. *Educ. Res. Rev.* 30:100325. doi: 10.1016/j.edurev.2020.100325
- Wang, Q., and Leichtman, M. D. (2000). Same beginnings, different stories: a comparison of American and Chinese children's narratives. *Child Dev.* 71, 1329–1346. doi: 10.1111/1467-8624.00231
- Wang, Y. (2023). Influence of early family nurturing environment on children's psychological and emotional social development. *Iran. J. Public Health* 52:2138. doi: 10.18502/ijph.v52i10.13852
- Wen, X., Xiaoqing, C., and Yancai, W. (2024). Association between home literacy environment and eye movements during storybook reading: implications for vocabulary and word reading in Chinese preschoolers. *Acta Psychol.* 250:104518. doi: 10.1016/j.actpsy.2024.104518
- Yanxue, W., and Bhaumik, A. (2024). The impact of parenting styles on child development: a comprehensive review. *Prestieesci Res. Rev.* 1. Available online at: <https://prestieesci.com/journal/index.php/pr/article/view/62> (Accessed December 27, 2025).
- Ye, L., Jiang, J., Liu, Y., Ran, Y., Zhang, S., Yin, Y., et al. (2024). Connection is all you need: a multimodal human-AI Co-creation storytelling system to support children's multi-level narrative skills. *arXiv*. [preprint] arXiv:2405.06495. Available online at: <https://arxiv.org/abs/2405.06495>
- Yoon, J., Kim, N., and Chung, E. (2023). Characteristics of scholarly journals published in non-English-speaking countries: an analysis of library and information science SCOPUS journals. *Learn. Publish.* 36, 14–24. doi: 10.1002/leap.1520
- Younger, P. (2010). Using google scholar to conduct a literature search. *Nurs. Stand.* 24:40–46. doi: 10.7748/ns.24.45.40.s51
- Yu, L., Huang, J., Liu, P. D., Yeung, S. S. S., Lin, D., Cheung, H., et al. (2024). How parenting styles affect the development of language skills and reading comprehension in primary school students. *Br. J. Educ. Psychol.* 94, 1245–1270. doi: 10.1111/bjep.12718
- Zhai, Y., Xie, H., Zhao, H., Wang, W., and Lu, C. (2023). Neural synchrony underlies the positive effect of shared reading on children's language ability. *Cereb. Cortex* 33, 10426–10440. doi: 10.1093/cercor/bhad293
- Zhang, X., Lau, C., and Su, Y. (2023). Home environment and development of English as a second/foreign language for young children in Asian contexts: a systematic review and meta-analysis. *Early Educ. Dev.* 34, 274–305. doi: 10.1080/10409289.2021.1981065
- Zheng, Z., Degotardi, S., and Djonov, E. (2021). Supporting multilingual development in early childhood education: a scoping review. *Int. J. Educ. Res.* 110:101894. doi: 10.1016/j.ijer.2021.101894



OPEN ACCESS

EDITED BY

Jens Bölte,
University of Münster, Germany

REVIEWED BY

Yousri Marzouki,
Qatar University, Qatar
Elisa Gavard,
Aix-Marseille Université, France

*CORRESPONDENCE

Alexander Taikh
✉ alexander.taikh@concordia.ab.ca

RECEIVED 05 August 2025

REVISED 13 February 2026

ACCEPTED 17 February 2026

PUBLISHED 04 March 2026

CITATION

Mangat A, Taikh A, Gagne C and
Spalding TL (2026) The influence of
prime characteristics in semantic
priming.
Front. Psychol. 17:1680365.
doi: 10.3389/fpsyg.2026.1680365

COPYRIGHT

© 2026 Mangat, Taikh, Gagne and
Spalding. This is an open-access article
distributed under the terms of the
[Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/)
(CC BY). The use, distribution or
reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication
in this journal is cited, in accordance
with accepted academic practice. No
use, distribution or reproduction is
permitted which does not comply with
these terms.

The influence of prime characteristics in semantic priming

Ajay Mangat¹, Alexander Taikh^{1*}, Christina Gagne² and
Thomas L Spalding²

¹Department of Psychology, Concordia University of Edmonton, Edmonton, AB, Canada, ²Department of Psychology, University of Alberta Faculty of Arts, Edmonton, AB, Canada

The semantic priming effect refers to the finding that processing a target word is facilitated when it is preceded by a related prime compared to an unrelated prime, and the semantic priming paradigm has been a key way to examine how knowledge is organized and retrieved. Both automatic and strategic processing of the prime can contribute to the semantic priming effect, and a major challenge in developing an integrated account of semantic priming is delineating conditions where the effect is driven by automatic or strategic processing of the prime. We incorporate literature reviews, meta-analyses, and studies with original experiments to examine how numerous factors influence the automatic and strategic processing of the prime, and thus the semantic priming effect. Specifically, we discuss the foundational models of semantic priming, and the automatic and strategic mechanisms by which primes are thought to influence their targets. We then discuss how prime-level lexical and semantic properties, prime-target relational factors, modality and presentation variables, and task-level determinants influence the automatic and strategic processing of the prime. We also point out gaps in the literature and suggest next steps in developing an integrated account of semantic priming.

KEYWORDS

automatic spreading activation, lexical decision task (LDT), semantic categorization task, semantic priming, strategic processing, typing task

Introduction

The semantic priming effect (Meyer and Schvaneveldt, 1971) is the finding that processing a target word is facilitated when it is preceded by a prime that is semantically related to its target (e.g., processing of the target word *dog* is facilitated by the related prime *cat* compared to an unrelated prime *table*). The semantic priming paradigm has been used to examine the organization and retrieval of semantic knowledge, and how contextual semantic information influences the word recognition process (e.g., see Jones and Estes, 2012; Neely, 1991 for reviews). Both automatic and strategic processing of the prime can contribute to processing of the target and the semantic priming effect (Neely et al., 1989; Perea and Rosa, 2002). Automatic processing of the prime occurs quickly, and does not require attention or the awareness of the prime (Posner and Snyder, 1975), while strategic processing of the prime requires its conscious identification, takes more time, and is more likely to be influenced by various experimental factors, such as the linguistic properties of the prime and its relationship to the target, how the prime is presented, and the type of task being used. Importantly, strategic processing of the prime can mask the contribution of automatic prime processing to the semantic priming effect.

Studies that use the semantic priming paradigm to examine the nature of semantic representations have generally used factorial designs and have examined the influence of the type of relationship that exists between primes and targets on the semantic priming effect (e.g., Bueno and Frenck-Mestre, 2008; Lucas, 2000; Perea and Rosa, 2002) and the effect of the experimental task and list context effects on the semantic priming effect (e.g., Becker et al., 1997; Bueno and Frenck-Mestre, 2008; de Groot, 1990; de Wit and Kinoshita, 2014, 2015a, 2015b; Grainger, 1998; Perea and Rosa, 2002; Williams, 1996). While the linguistic properties of primes and targets influence their processing, studies have generally focused on the influence of linguistic variables on the recognition of individual words outside of the priming paradigm (e.g., Pexman et al., 2008; Yap et al., 2012; Yap et al., 2011), or have focused on the linguistic properties of the target word (e.g., Becker, 1979; Cortese et al., 1997; Hutchison et al., 2008). However, less direct emphasis has been placed on the processing of the prime itself, which is critical to understanding the semantic priming effect, and distinguishing between the influences of automatic and strategic processing.

The central aim of this review is to examine how various factors that influence the processing of the prime modulate the balance between automatic and strategic processing. Delineating the conditions under which automatic and strategic processing of the prime drive or contribute to the semantic priming effect is critical to developing an integrated account of semantic priming. We draw upon meta-analyses, literature reviews, and studies reporting original experiments. We first provide an overview of the foundational models of semantic priming, which are distinguished by whether the influence of the prime is automatic or strategic, and by whether the prime facilitates the identification of the target or whether the two are evaluated together after identification of the target. We then focus on factors that influence the processing of the prime: the prime-level lexical and semantic properties; prime-target relational factors; modality and presentation variables; and task-level determinants. We then describe an integrated account of semantic priming (Perea and Rosa, 2002) which posits both automatic and strategic processing of the prime, and discuss the gaps in the current literature and the challenges to developing an integrated account of semantic priming. Our aim is not to provide a fully unified model of the semantic priming effect, but to delineate boundary conditions between when semantic priming is driven by automatic or strategic processing and to highlight the joint influence of numerous factors (prime-level properties, prime-target relational factors, modality and presentation variables, and task-level determinants) on the semantic priming effect whenever possible.

Foundational models of semantic priming

A common conception is that a lexical network (i.e., the mental lexicon) stores words based on their form and sound, whereas, the semantic information associated with words is thought to be stored in a separate network. Indeed, there are numerous models (e.g., Anderson, 1983; Collins and Loftus, 1975; Seidenberg and McClelland, 1989; Plaut, 1995; Plaut and Booth, 2000) that make this distinction when describing how words and their associated meanings are stored and retrieved. For example, according to localist models (e.g., Collins and Loftus, 1975), the concepts

associated with words are stored as units in a semantic network and are connected based on associations that are learned over time (e.g., *cat* and *dog*). Activation of one conceptual node (i.e., *cat*) results in activation of the connected nodes (i.e., *dog*). According to distributed models (e.g., Plaut, 1995), a semantic network is composed of features, and a concept corresponds to a set of related features and other concepts (e.g., the features of fur, a tail, four legs... are associated with both the concept of cat and the concept of dog).

In experiments examining the semantic priming effect, the prime target pairs can be related associatively, where the occurrence of one word is predictive of the occurrence of the other, and/or semantically, where the prime and target overlap in features or have any kind of a meaning relationship (e.g., synonyms, antonyms...). Localist models thus generally explain semantic priming as activation automatically spreading from the node that represents the prime to related nodes and thus facilitating processing of those words (e.g., the node of the target word, *dog*, receives activation from the node of the prime word, *cat*, facilitating the processing of the target word). Distributed models, in contrast, assume that the features activated by the prime will help facilitate the processing of the target to the extent that the prime and target have overlapping features.

While numerous mechanisms have been proposed to explain the semantic priming effect, they generally involve two key distinctions. The first distinction is whether processing the semantic prime influences the lexical activation of the target word or whether the prime and the target are processed together after the lexical selection of the target word. The second distinction is whether the process responsible for the priming effect is automatic or strategic. In this section, we briefly review three types of accounts of semantic priming that highlight these distinctions.

Automatic processing of the prime

Automatic spreading activation, as first explained by Collins and Loftus (1975; see also Posner and Snyder, 1975), proposes that the activation from the lexical representation of the prime spreads to the target's lexical representation. Thus, the prime pre-activates the target's lexical representation and influences the lexical selection of the target. According to localist accounts of semantic memory, such as the one proposed by Collins and Loftus, the facilitation of the target occurs through direct links or connections within semantic memory because the prime and target are associated with each other. Thus, there is no concurrent processing of the prime and target. Spreading activation occurs quickly and automatically (Posner and Snyder, 1975) in that it does not require attention, or even explicit awareness of the prime. We should also note that distributed accounts of semantic memory (Fernandino and Conant, 2024; Fischler, 1977; Plaut, 1995) also potentially explain fast acting and automatic semantic priming effects. For example, de Wit and Kinoshita (2014, 2015a, 2015b) propose that the features of the prime are extracted and facilitate the processing of the target, although this process depends on participants having a task that relies on encoding features (we discuss these studies and the effect of task type, subsequently). The processes of expectancy generation and retrospective post-lexical integration (see Neely, 1991 for a review) which we describe next, can contribute to the semantic priming effect concurrently, and thus obscure the influence of automatic spreading activation.

Expectancy generation

According to expectancy accounts of semantic priming (Becker, 1980; Neely, 1977), participants predict which target words are likely to follow the prime, or the relevant features that define the target words. An expected set of words (or features) should facilitate the recognition of a target related to the prime, but also delay responding to an unrelated target, as participants would first have to reject the words in their expectancy set and then consider the unrelated target word. Expectancy can be a strategic process which requires attentional resources, and the formation of expectancy sets is thought to require enough time between the onset of the prime and the onset of the target (i.e., stimulus onset asynchrony, SOA), and is encouraged by having a greater proportion of primes and targets be related to each other (i.e., relatedness proportion). The SOA needed for expectancy generation, or other strategic processes, to occur is unclear and likely varies between tasks (Hutchison et al., 2001), but has been estimated to be about 300 ms (Neely, 1991).

Post-lexical integration

Retrospective post-lexical and meaning integration accounts posit that participants process the prime and target together. The prime does not influence the lexical selection of the target. Instead, participants are thought to engage in a semantic matching process where they determine whether the prime and target are semantically related following lexical access of the target word but prior to the response of the experimental task (de Groot, 1984; de Wit and Kinoshita, 2014, 2015a, 2015b; Forster, 1981; McKoon and Ratcliff, 1992; Neely et al., 1989). For example, in the compound cue model (McKoon and Ratcliff, 1992) participants form a compound cue made up of the prime and target, and evaluate its familiarity. Familiar cues are ones where prime and target are associatively or semantically related. Similar to expectancy accounts, this process is under strategic control, at least to some degree, and requires processing the meaning of the prime (Neely and Keefe, 1989), which can occur before or after the presentation of the target. In the lexical decision task (LDT), where participants judge whether the target letter string is a word or not, retrospective integration accounts explain the semantic priming effect as a bias. Specifically, participants are biased to make a “word” response when the prime and target are related. The detection of a relationship between the prime and target, such as a familiar prime-target compound cue, would require that the target be a word. Thus, participants are biased to make a “nonword” response when prime and target are unrelated, and are slowed in their response to words following unrelated primes. Increasing the relatedness proportion thus also encourages the use of retrospective integration to guide the LDT decision, because “word” responses are more likely to come after related primes. Thus, retrospective matching can be potentially confounded with expectancy generation.

Importantly, multiple mechanisms can potentially contribute to the semantic priming effect because the processing of the prime can continue when it is no longer being shown and can overlap with the processing of the target. For example, when the duration of the prime and the SOA are sufficient (i.e., over 300 ms), the priming effect can potentially be explained by expectancy generation, given that there is enough time to develop expectancy set, by retrospective matching, since the processing of the prime can continue after the presentation of the target, and potentially, automatic spreading activation since its

influence may persist throughout the trial. Indeed, Perea and Rosa (2002) propose a hybrid model of semantic priming, which includes spreading activation and post-lexical integration accounts (see also Neely et al., 1989). We discuss Perea and Rosa in more detail subsequently, but briefly, the authors propose that automatic spreading activation drives semantic priming when the prime is briefly presented and the onset of the target follows quickly enough so that automatic spreading activation does not fade. Post-lexical integration would then happen after accessing the target word.

As the brief overview of accounts of semantic priming demonstrates, the influence of automatic prime processing on the semantic priming effect is difficult to rule out, but may be obscured by the strategic processing of the prime. Strategic prime processing is more flexible and more likely to be influenced by numerous factors that can be divided into prime-level lexical and semantic properties, prime-target relational factors, modality and presentation variables, and task-level determinants (though these variables can potentially influence the automatic processing of the prime as well). We first focus on prime-level lexical and semantic properties, which influence identifying the prime word and extracting its semantic information that drives the semantic priming effect.

Prime-level lexical and semantic properties

Most studies on word recognition have examined the influence of the linguistic item-level properties of individual words, outside of the semantic priming paradigm. Although these studies do not directly use a semantic priming paradigm, how a word's lexical and semantic information influences its own recognition may provide insight into how that word would be processed as a prime and thus influence the processing of a subsequently presented target. For example, properties that facilitate the recognition of the word and access to its semantic information might make it a more effective semantic prime.

Lexical properties

The influence of lexical factors on word recognition has been well studied (see Yap and Balota, 2015 for a review) and a review of this topic is beyond the scope of the current paper. However, some key lexical factors are frequency (e.g., Whaley, 1978), where words that occur more frequently in the language are recognized more quickly, and length, where shorter words are recognized more quickly (although length and frequency are strongly inversely correlated). Words with a greater orthographic neighborhood size (i.e., the number of words that can be obtained by substituting a single letter of the target word) are responded to more quickly in naming and the lexical decision task, especially for low frequency words (see Andrews, 1997 for a review). These lexical properties influence how quickly the lexical representation of a word is accessed, and thus how quickly the semantic information of a word is accessed. Indeed, Taikh et al. (2015) found that access to meaning in words first requires lexical access, which depends on lexical variables. Taikh et al. examined the influence of lexical and semantic variables for words and their corresponding images on a semantic classification task (living vs. non-living). Lexical variables did not predict response latencies for images, but they did predict response latencies for words.

Less is known about the influence of linguistic properties of primes and targets on the semantic priming effect. Becker (1979) showed that semantic priming effects were larger for lower frequency targets. Hutchison et al. (2008) collected lexical decision and naming responses to 300 semantic prime-target pairs from over 200 participants. This early semantic priming megastudy was extended by the Semantic Priming Project (Hutchison et al., 2001), which we will discuss in more detail subsequently. Using a large set of items and participants allowed Hutchison et al. (2008) to examine the influence of both prime and target item-level variables on the semantic priming effect. In their experimental paradigm, the authors used primes that were semantically related to their targets, semantically unrelated, or were neutral primes (i.e., the prime was the word “BLANK”). The SOA was either short (250 ms) or long (1,250 ms). The responses to targets following neutral primes were used as an indication of how difficult the target was to recognize in isolation. Priming was greater for targets that had longer latencies in the neutral priming condition. However, in the LDT, priming was greater for shorter targets and targets with fewer orthographic neighbors when the SOA was long, which is somewhat contradictory to their finding of greater priming effects for words that were responded to more slowly in the neutral priming condition, and to findings that greater orthographic neighborhoods facilitate the recognition of individual words (Andrews, 1997). In general, however, Hutchison et al. argue that a related prime may have a greater impact on the target when the target is processed more slowly, allowing more time for the related (vs. unrelated) prime to influence the recognition of the target.

In terms of the characteristics of the prime, Hutchison et al. (2008) found that in both the lexical decision task and naming, priming was greater when related primes were short and had fewer orthographic neighbors. These results are stronger at short SOA's, where priming was also greater for high-frequency related primes (we discuss the influence of SOA on the priming effect subsequently). The authors argue that prime words that are recognized and identified more quickly can facilitate the recognition of their targets to a greater extent, especially at short SOA when the influence of strategic processes may be reduced. Interestingly, and in contrast to these findings, Hutchison et al. also found that priming in the LDT, but not in naming, was greater for prime words that were responded to more slowly in the ELP database, and this effect was greater at long SOA. The authors did not offer an explanation for this contradictory finding.

Semantic properties

Similarly, words with more semantic information associated with them are processed faster, especially in tasks that require semantic information about the word (see Balota et al., 1991; Pexman, 2012, 2020 for reviews). The retrieval of meaning is facilitated when the word has a richer semantic representation (e.g., Barca et al., 2016; Pexman et al., 2008; Sidhu et al., 2016; Taikh et al., 2015; Yap et al., 2011). For example, responses to words are facilitated when the word has a greater semantic neighborhood size (i.e., the number of semantically related words or concepts that a word has) (Buchanan et al., 2001), when the word has more semantic features associated with it (McRae et al., 2005), and when it elicits more first associates in a free association task (Nelson et al., 1998).

Studies have also examined semantic priming where the primes are novel words learned by the participants during the experiment (e.g., Bakker et al., 2015; Bakker-Marshall et al., 2018; Balass et al.,

2010; Ball et al., 2024; Borovsky et al., 2012; Perfetti et al., 2005; Tamminen and Gaskell, 2013), with mixed results on whether newly learned words can be effective semantic primes. Novel words, even if adequately learned in the context of the experiment, seem to be processed differently as primes from familiar words. In general, novel words semantically primed their targets when they were able to be processed strategically because of a long enough SOA and a task that encourages strategic processing. We discuss the influence of both of these factors on novel word semantic primes subsequently. In general, however, semantic information associated with novel words is likely not integrated in the semantic network and thus not connected with semantic representations of familiar words (Davis and Gaskell, 2009), meaning that these words, as primes, could not automatically influence their targets. Semantic information associated with novel words should be activated more slowly and they are only able to act as semantic primes if they can be processed strategically.

In summary, the literature suggests that, in general, quicker processing of the prime predicts a greater semantic priming effect, especially at shorter SOAs where automatic processing of the prime is thought to play a greater role. At longer SOAs, the linguistic properties of the prime and strategic processes may override the relatively subtle facilitation provided by the prime being easy to process. Slower processing of the targets likewise predicts a greater priming effect because it allows the related prime to exert its influence. However, the way semantic information from the prime and from the target are processed and influence responding to the target depends on the prime-target relationship. Specifically, information extracted from the prime is most useful if it has a strong associative relationship to the target and can prospectively facilitate its processing. We focus on the influence of prime-target relationships next, which help determine what information from the prime becomes relevant in processing the target and influencing the semantic priming effect.

Prime-target relational factors

As noted previously, the relationship between a prime and target can be associative (Thompson-Schill et al., 1998), where the occurrence of one word is predictive of the occurrence of the other, and/or the relationship can be semantic, where the prime and target overlap in features or have any kind of a meaning relationship (e.g., synonyms, antonyms, and so on). The direction and strength of association can also influence the semantic priming effect. For example, priming in prime-target pairs with a strong forward association (e.g., cigar-smoke) may be driven by a different mechanism than prime-target pairs with a backward association (e.g., herd-cow).

Associative and semantic prime-target relationships

Studies have attempted to determine whether automatic semantic priming is driven by the associative relationship between the prime and target, which is consistent with localist theories of how knowledge is stored, or by featural overlap, which is consistent with distributed theories of how knowledge is stored. One way that studies have attempted to distinguish between these drivers of semantic priming is by examining the effect of the type of relationship between prime and target on the semantic priming effect. Examining the effect of

relationship type on semantic priming provides information on how primes are processed. For example, finding that featural overlap drives semantic priming would suggest that while processing the prime, its features are extracted and used to help process the target.

In a meta-analysis, Lucas (2000) analyzed priming effects from studies that apparently controlled the association strength in their prime-target pairs. Lucas found priming effects when the prime and target were related semantically in the absence of any associative relationship, but no priming based on just association (i.e., when there was no feature overlap). However, the effect was also stronger when the prime and target had both a semantic and an associative relationship. Priming effects driven purely by a semantic relationship were not influenced by SOA and relatedness proportion, suggesting they were driven by an automatic pre-activation mechanism rather than a strategic mechanism. One such effect was between category coordinates (i.e., where the prime and target belong to the same category) that are not associated with one another. Lucas concluded that semantic priming effects, at least those that are automatic, are driven primarily by feature overlap rather than semantic association. Similarly, Perea and Rosa (2002, Experiment 2) found priming in pairs that were semantically related (prime and target were synonyms, antonyms, or category coordinates) but not associated with each other according to free-association norms. We discuss Perea and Rosa in more depth subsequently.

Hutchison (2003) examined the studies used by Lucas (2000) and likewise concluded that priming for semantically related pairs was likely to be automatic, and that the priming effect was larger when prime-target pairs had both a semantic and associative relationship. However, in contrast to Lucas (2000), Hutchison concluded that there was strong evidence for semantic priming effects based on just association between prime and target. For example, Hutchison argued that in studies that Lucas used as evidence of priming between pairs that are semantically but not associatively related (e.g., category coordinates), the prime and target pairs actually were associatively related. Hutchison also argues that Perea and Rosa (2002) defined their prime target pairs as not associated based on the absence of only forward association strength, where the target is likely to be a response when the prime is a cue. The association between primes and targets can also be backward and symmetrical (we cover this subsequently). In general, Hutchison argued that isolating the influences of association and feature overlap is often difficult because words that are associated often also overlap on features, making it difficult to determine whether an observed priming effect is driven by an associative relationship or the feature overlap between primes and targets.

Hutchison (2003) further supports the presence of automatic associative priming (i.e., priming due to spreading activation) with mediated priming studies, which Lucas (2000) did not include in her meta-analysis. In the mediated priming paradigm, a target is related to a prime via a connecting word or concept (e.g., the target *stripes* is indirectly related to the prime *lion* through the mediating concept of *tiger*). Mediated priming has been found in the LDT (e.g., McNamara and Altarriba, 1988; McKoon and Ratcliff, 1992; Livesay and Burgess, 1998) and naming (e.g., Balota and Lorch, 1986; Livesay and Burgess, 1998), although see de Groot (1983) for evidence against mediated priming. Hutchison argues that mediated priming effects are likely automatic because strategic processing of the prime, while useful for direct associates, would not be useful for mediated associates. Specifically, using the retrospective matching strategy on mediated associates should bias participants to make nonword responses

because the compound cue formed by the prime and target are not directly related, similarly to what would happen in a trial when the prime is unrelated to the target. A similar logic would apply to expectancy generation. Hutchison likewise argues that mediated priming is likely not due to feature overlap because the prime and target share little or no overlap in semantic features, although we are unaware of any studies directly examining the influence of feature overlap on mediated priming.

Other researchers argue that mediated priming could be explained by post-lexical matching strategies. For example, McKoon and Ratcliff (1992), argue that mediated priming effects may be due lexical co-occurrence, where the prime and target may not be associatively related, but may nonetheless co-occur in natural language. The familiarity of the prime-target compound cue would thus facilitate the processing of the target. However, Hutchison (2003) points to studies where the degree of association did not influence the mediated priming effect (Balota and Lorch, 1986; McNamara, 1992), and where lexical co-occurrence did not influence the mediated priming effect (Livesay and Burgess, 1998). Hutchison also mentions the findings of Hayes and Bissett (1998), where participants formed associations between novel words (i.e., nonwords). Specifically, participants were trained to associate stimuli from list A to corresponding stimuli from list B, and to corresponding stimuli from list C. Using a double LDT, where two letter strings are simultaneously presented and participants decide whether both are words or nonwords, Hayes and Bissett found that participants responded more quickly to pairs from lists B and C that were associated to each other though list A stimuli (although list B stimuli were never directly associated with list C stimuli). The findings of Hayes and Bissett are consistent with automatic spreading activation facilitating concepts that are linked in the semantic network, even if they do not co-occur.

More recently, Jones (2010) found evidence for a strategic post-lexical integration account of mediated priming. Jones used prime target pairs where there was no associative relationship between the prime and the mediator, and the mediator and the target. Jones found a mediated priming effect in the double and standard LDT tasks, but not the continuous LDT task, which precludes strategic processing. We discuss the continuous LDT in more detail subsequently. However, Jones argues that the search for a relationship between prime and target, as posited by semantic matching models (e.g., Neely et al., 1989), could include searching for a mediator (i.e., searching for *tiger* when shown the prime *lion* and the target *stripes*). Jones concludes that mediated priming is retrospective and strategic, although not in a way that is consistent with the compound-cue model (Ratcliff and McKoon, 1988).

Direction and strength of association

Associative strength can be measured as the likelihood that the target word is produced as a response to the prime cue in a free association task (De Deyne et al., 2019; Nelson et al., 1998), and it can be directional. Specifically, the forward associative strength (FAS) is the proportion of participants who responded with the target when cued with the prime, while the backward associative strength (BAS) is the proportion of participants who produce the prime when cued with the target. According to the model of semantic priming proposed by Neely and Keefe (1989), stronger FAS should result in greater priming at longer SOAs, where participants would have more time to generate a list of expected targets given the prime. Stronger BAS was likewise

predicted to increase the semantic priming effect at longer SOAs because participants are more likely to evaluate the target and the prime together using retrospective matching at longer SOAs. Hutchison et al. (2008) found that at a short SOA, greater FAS predicted a greater semantic priming effect. At long SOA, greater semantic priming was predicted by both stronger FAS and BAS. We will discuss these findings in more detail when considering task type. However, these findings are consistent with the idea of both retrospective matching and expectancy generation being applied at long SOA, but retrospective matching being less likely at a short SOA. One possibility for why a shorter SOA would reduce retrospective matching is that it does not allow enough time to process the meaning of the prime, as argued by Neely and Keefe (1989). However, the meaning of the prime may also be processed while the target is being processed.

The direction of association between prime and target can also influence the nature of the semantic priming effect. For example, Thomas et al. (2012) tested the semantic priming effect with pairs that only had a strong forward association (e.g., cigar—smoke), only a strong backward association (e.g., herd - cow), or were symmetrical (e.g., girl—boy). The targets were visually degraded. A common finding is that the semantic priming effect is greater for visually degraded targets (e.g., Balota et al., 2008; Borowsky and Besner, 1993; Stolz and Neely, 1995) because the longer time needed to process the target allows for more influence from the (related) semantic prime. Thomas et al. found a stronger semantic priming effect for visually degraded targets only when the prime and target pairs were backward associated or symmetrical. The authors argue that retrospective use of the prime is encouraged by the visually degraded target when there is a backward association.

Associative strength can also be measured in terms of the co-occurrence of the prime and target words, which is how likely they are to appear together in text. The latent semantic analysis (LSA) model (Landauer et al., 1998) assumes that semantically similar words co-occur within the same contexts. LSA values have been found to predict semantic priming effects (e.g., Chwilla and Kolk, 2002; Landauer and Dumais, 1997; Lund et al., 1995, 1996). Hutchison et al. (2008), however, found that LSA did not predict semantic priming effects. More recently, Fernandino and Conant (2024) compared how well multiple types of models predicted the automatic semantic priming effect in the LDT. In their experiment, primes were presented for 150 ms and were followed by a 50 ms mask, prior to the presentation of the target, thus precluding expectancy generation, although retrospective matching would still be possible since the primes were visually identified. Fernandino and Conant found that experiential similarity between primes and targets (i.e., sensory-motor and affective features of the words) was the primary driver of the semantic priming effect. Additionally, the authors did not find independent contribution of the co-occurrence measures or of taxonomic information (i.e., semantic similarity between prime and target based on their category information). The authors argue that sensori-motor and affective features of primes and targets are automatically activated during the LDT, which does not explicitly require that kind of information.

In summary, priming that is driven more by automatic processing of the prime seems to be more robust for semantically related pairs, however there is disagreement about whether automatic priming exists for pairs that are purely associatively related. Strong FAS produces robust priming at both long and short SOAs, consistent with a priming mechanism that is automatic and prospective (e.g., spreading activation). Priming driven by a strong BAS seems robust at long (but

not short) SOAs, suggesting that BAS encourages the strategic and retrospective processing of the prime. However, the influence of the prime-target relationship on the priming effect depends on modality and presentation variables. For example, Perea and Rosa (2002) found that the effect of prime-target strength of association depends on SOA. We discuss Perea and Rosa in more detail subsequently.

Modality and presentation variables

We first address the question of whether semantic priming occurs when the prime duration is too brief for it to be identified. A semantic priming effect without prime identification is consistent with the automatic accounts of semantic priming, such as spreading activation from the prime (Marcel, 1983), or accumulation of features from the prime (de Wit and Kinoshita, 2014, 2015a, 2015b), because strategic use of the prime (either via expectancy generation or retrospective matching) would be ruled out. We then examine studies where the prime can be consciously identified, and thus where its strategic processing is possible.

Prime duration

Numerous studies have found semantic priming without prime identification (e.g., Carr et al., 1982; Cheesman and Merikle, 1986; Hirshman and Durante, 1992), while other studies found no semantic priming when the prime was not identified (Dark, 1988; Dark and Benson, 1991; de Wit and Kinoshita, 2015b). Specifically, Hirshman and Durante presented participants with prime - target pairs and asked participants to only identify the prime on some trials, and to only respond to the target in a LDT on others. The authors argue that measuring prime identification when participants are asked to also respond to targets can alter prime identification because of retrospective priming mechanisms (Dark and Benson, 1991), where information from the target helps participants identify a briefly presented prime. Masked primes that were presented for 33 ms were almost never identified, although a semantic priming effect was found. A semantic priming effect was also found when masked primes were presented for 50 ms, and these primes were more likely to be identified (about 50% when the prime was related to the target and about 30% when the prime was unrelated). The authors conclude that semantic priming can occur without identifying the prime (i.e., at 33 ms durations), although identification of primes can occur even when they are presented briefly and masked (i.e., at 50 ms durations). Importantly, even brief prime duration and a short SOA allows for the possibility of retrospective matching, since primes were more likely to be identified in prime-target pairs where a semantic priming effect was found, and when the prime and target were related. The semantic priming effect found by Hirshman and Durante when the prime was not identified is consistent with automatic spreading activation contributing to the semantic priming effect. However, post-lexical mechanisms may contribute to the effect when the prime is identified, even when SOA is short enough to not allow for the generation of expectancy sets.

Studies where primes are visible (where the prime duration is typically 100 ms or more) often manipulate the relatedness proportion, which is the proportion of related prime-target pairs in the experimental stimulus list (Tweedy et al., 1977), often in concert with manipulating the SOA. A semantic priming effect that is driven by

automatic spreading activation, as conceptualized by Collins and Loftus (1975) should not be influenced by relatedness proportion. However, relatedness proportion should influence expectancy generation and retrospective matching. Participants are more likely to generate an expectancy set of the target, and evaluate the target and prime together, when they see the target when the relatedness proportion is high (Becker, 1980; Neely, 1977). Consistent with this notion, numerous studies (Grossi, 2006; Hutchison et al., 2001; Neely et al., 1989; Pecher et al., 2002; Perea and Rosa, 2002) have found that greater relatedness proportion increases the semantic priming effect at longer SOAs (see Hutchison, 2007) for a review. Effects found at short SOAs (under 300 ms) are thus thought to reflect automatic spreading activation (Neely, 1977), overlap of features between prime and target (de Wit and Kinoshita, 2014, 2015a, 2015b), and/or retrospective matching (e.g., Taikh and Lupker, 2020), but not expectancy generation. Of note, priming effects found at longer SOAs can be potentially explained by expectancy generation, retrospective matching, and automatic spreading activation, since it is unclear how long the influence from spreading activation can persist. However, retrospective matching requires enough time to process the meaning of the prime, as argued by Neely and Keefe (1989).

Relatedness proportion

Numerous studies reported no influence of relatedness proportion on the semantic priming effect at short SOA (e.g., Grossi, 2006; Hutchison et al., 2001; Neely et al., 1989; Neely, 1977; Pecher et al., 2002; Perea and Rosa, 2002), which is consistent with the idea that semantic priming effects at short SOAs are due to automatic spreading activation or feature overlap as there is no time to form expectancy sets. Other studies have found that relatedness proportion does influence the semantic priming effect at short SOAs. For example, de Groot (1984) varied relatedness proportion (0.25, 0.50, 0.75, and 1.00) at different SOAs (240, 540, and 1,040 ms) and found an influence of relatedness proportion on the semantic priming effect when the SOA was 240 ms (the effect was bigger in the 0.75 condition than the 0.25 condition). de Wit and Kinoshita (2014, 2015a) also found that relatedness proportion modulated the semantic priming effect at SOAs considered too short for expectancy generation. de Wit and Kinoshita examined the effect of relatedness proportion on semantic priming using the lexical decision and semantic categorisation tasks (we address the influence of task type on the semantic priming effect in the subsequent section). The findings that relatedness proportion influences the semantic priming effect at short SOAs are consistent with the idea that semantic priming, at least in part, is driven by post-lexical meaning and integration accounts. When participants detect a relationship between a prime and target, they are biased to make a “word” response more quickly (Neely et al., 1989).

Stimulus onset asynchrony (SOA)

Studies have also examined the semantic priming effect in associative and semantic prime-target pairs at varying SOAs. For example, Perea and Rosa (2002, Experiments 1 and 2) found that the semantic priming effect in the LDT was driven by association between prime and target at their shortest SOA (66 ms) but not at longer SOAs (83 ms and longer). Specifically, in their Experiment 1, the associative strength of semantically related prime target pairs influenced the magnitude of the priming effect only at the 66 ms SOA. In Experiment 2,

Perea and Rosa used prime target pairs that were semantically related but not associatively related (at least based on association norms). Semantic priming was not found in the 66 ms SOA, but was found in the longer SOAs. The authors point to a hybrid model that combines automatic spreading activation and integration accounts of priming that may better explain such findings. The hybrid model proposes that at short SOAs (~60 ms), priming is brought on by automatic spreading activation, whereas at longer SOAs, integration of features plays a role. Importantly, the presence of priming that is automatic and driven by the associative relationship between prime and target is consistent with the conclusion of Hutchison (2003).

Bueno and Frenck-Mestre (2008, Experiments 5–7) likewise investigated the differences between semantically and associatively related primes in a masked priming paradigm using the LDT at various prime durations and backward masks yielding SOAs of 42 ms, 57 ms, 99 ms, and 256 ms. Their prime-target pairs were either associatively and semantically related, or just semantically related (i.e., feature overlap) but were not forward-associated according to word association norms. The authors found no semantic priming effect for either the associatively or semantically related pairs at the two shortest SOAs (Experiment 5), but did find a priming effect for both types of prime target pairs at the 99 ms SOA (Experiment 6), and at the 256 ms SOA (Experiment 7). Bueno and Frenck-Mestre, like Perea and Rosa (2002, Experiment 2), failed to find a semantic priming effect in the LDT for pairs that were only semantically but not associatively related when the SOA was short (42 ms and 57 ms SOAs for Bueno and Frenck-Mestre, and 66 ms SOA for Perea and Rosa). However, while Perea and Rosa found that the semantic priming effect was driven by prime-target association strength at their 66 ms SOA, Bueno and Frenck-Mestre did not find any priming based on association at their 42 ms and 57 ms SOAs. While Perea and Rosa used Spanish stimuli, Bueno and Frenck-Mestre used French stimuli. Perea and Rosa also did not use backward masks after presenting primes, meaning that in the 66 ms SOA condition, the prime was presented for 66 ms. Bueno and Frenck-Mestre used 14 ms backward masks in their two shortest SOA conditions, meaning that those primes were presented for 28 ms and 43 ms. Nonetheless, Bueno and Frenck-Mestre argue that their findings were consistent with those of Lucas (2000) and that automatic semantic priming does not depend on association strength.

Ball et al. (2024) examined the effect of SOA on semantic priming from primes that were newly learned words. The authors compared semantic priming effects from familiar word primes and novel word primes learned during the course of the study at varying SOAs (450 ms, 1,000 ms, 200 ms). In all cases, primes were presented for 200 ms, and the interstimulus intervals were varied to produce differing SOAs. The authors found that at the 450 ms and 200 ms SOAs, novel words did not prime their targets while familiar words did. At the 1,000 ms SOA, novel words did prime their targets, but familiar words did not. Ball et al. argue that because the semantic representations of novel words are not yet integrated into the semantic network (Davis and Gaskell, 2009), they engage in semantic processing through controlled and strategic semantic processes, which require greater SOAs (see also Bakker et al., 2015; Balass et al., 2010; Perfetti et al., 2005). Familiar words, in contrast, could semantically prime their targets automatically (i.e., at shorter SOAs).

Ball et al. (2024) link the strategic processing of primes (including retrospective integration) to longer SOAs. Specifically, Ball et al. argue that at a 1,000 ms SOA, the lack of semantic priming from familiar word primes was because of the low prime-target BAS. In contrast, the

authors say that at the 1,000 ms SOA, retrospective integration may have contributed to the priming from the novel words as the BAS ratings between these primes and their targets were greater. These findings are potentially consistent with those of [Hutchison et al. \(2008\)](#), where greater BAS predicted a stronger semantic priming effect at a long SOA but not a short one. However, these findings contradict [de Wit and Kinoshita \(2015a, 2015b\)](#), where semantic priming was found in the LDT at a short enough SOA to preclude expectancy generation (we discuss de Wit and Kinoshita in more detail subsequently). De Wit and Kinoshita argue these effects are driven by retrospective matching because they were modulated by relatedness proportion and the priming effect was greater for slower trials (see also [Taikh and Lupker, 2020](#)). Based on [Neely's \(1991\)](#) three process model of semantic priming, retrospective integration should be possible at shorter SOAs because the processing of the prime could continue during the processing of the target. Retrospective integration of novel word primes and their targets would thus allow priming at shorter SOAs, suggesting that the effects found by Ball et al. were driven by expectancy generation.

In summary, the literature suggests temporal boundary conditions for the mechanism driving semantic priming. The contribution of strategic prime processing to the semantic priming effect can only be completely ruled out when the prime is presented for less than 50 ms and thus cannot be identified. However, ruling out the influence of automatic processing of the prime is difficult because it is unclear how long it persists. When the prime can be identified, retrospective processing of the prime can also contribute to the semantic priming effect, and the use of this strategy is encouraged by a greater relatedness proportion, and stronger BAS. An SOA of approximately 300 ms is thought to be required for expectancy generation and is encouraged by a strong FAS. Variables that encourage faster processing of the prime and slower processing of the target (e.g., length and frequency) predict greater semantic priming effects.

The experimental task type influences the effects of modality and presentation variables. For example, tasks that require producing the targets (e.g., naming or typing the target words) seem to minimize the use of retrospective matching even if the SOA and BAS are sufficient to encourage retrospective matching. We address the effect of task-level determinants next.

Task-level determinants

The semantic priming effect has been studied using numerous tasks ([Balota et al., 2008](#); [de Wit and Kinoshita, 2014](#); [de Wit and Kinoshita, 2015a, 2015b](#)), including semantic decision, naming, and relatedness judgment, although the lexical decision task (LDT) is the most common. In the LDT, participants decide whether a target letter string is a real word or a nonword. Most often, participants respond only to the target, not the prime. However, in the single-presentation LDT (e.g., [Davies, 1998](#); [Jones, 2010](#); [McNamara and Altarriba, 1988](#)), participants respond to both the prime and the subsequently presented target. Responses to the LDT are thought to be driven mainly by lexical level activity (e.g., [Pexman et al., 2002](#)), although post lexical strategies such as semantic matching may influence responses as well (e.g., [Taikh and Lupker, 2020](#)). However, several studies have shown limitations in using a lexical decision task. Some studies have shown the LDT to be a less reliable indicator of semantic processing than a

semantic categorization task ([Becker et al., 1997](#); [Bueno and Frenck-Mestre, 2008](#); [de Groot, 1990](#); [Grainger, 1998](#); [Williams, 1996](#)).

Single presentation-LDT

In the single-presentation LDT, participants respond to both the prime and the target sequentially. The single-presentation LDT is argued to be driven solely by automatic processes, thus excluding the influence of expectancy generation or retrospective matching because participants are thought to be less likely to notice relationships between the subsequently presented words. [Perea and Rosa \(2002\)](#), Experiments 3 and 4) tested the semantic priming effect with prime - target items that were associatively and semantically related, or just semantically related. When the interval between a response and the subsequently presented stimulus was short (200 ms in Experiment 3), Perea and Rosa found a significant semantic priming effect for prime - target pairs that were associatively and semantically related, and prime - target pairs that were just semantically related. When the response stimulus interval was long (1750 ms in Experiment 4), there was only a priming effect for the associatively related pairs. Perea and Rosa argued that priming driven by a semantic (but not an associative) relationship may persist for a short time (i.e., over 200 ms but under 1750 ms). In summary, it seems that associative information may drive priming automatically, via automatic spreading activation, or strategically, via expectancy generation and/or retrospective matching.

Semantic categorization task

In semantic categorization tasks, participants make a judgement about the target word that relies on its meaning, for example indicating whether the target is an animal or an object. In contrast to the LDT, semantic categorization tasks require participants to attend to some semantic dimension of the targets. Semantic priming with this task has been found to be greater at long SOAs, where more time is allowed to process the relevant semantic information of the prime (e.g., [Becker et al., 1997](#)). As with the LDT, some researchers (e.g., [Jared and Seidenberg, 1991](#)) argue that semantic categorization may be contaminated by other processes, such as expectancy generation ([Hutchison, 2003](#); [Lucas, 2000](#)). However, results from [Bueno and Frenck-Mestre \(2008\)](#) show that semantic categorization tasks are not necessarily strategy-dependent and may reflect automatic spreading activation. As noted, Bueno and Frenck-Mestre examine the semantic priming effect in semantically and associatively related pairs with various SOAs, in both the LDT (Experiments 5–7) and semantic categorization task (Experiments 1–4). The authors found a semantic priming effect in pairs that were semantically but not associatively related in the semantic categorization task at the 42 ms and 57 ms SOAs (Experiment 1), which contrasts with their Experiment 5, where no semantic priming effect was found for semantically but not associatively related pairs at these same SOAs when using the LDT. Using prime-target pairs that were associatively related but weakly semantically related, the authors found no priming at SOAs of 42 ms and 57 ms (Experiment 2), but did find priming at an SOA of 99 ms (Experiment 3) and an SOA of 256 ms (Experiment 4).

Considering their LDT and semantic categorization experiments, [Bueno and Frenck-Mestre \(2008\)](#) concluded that semantic priming increases when the task requires processing of semantic information. Specifically, priming effects occurred at shorter SOAs in the semantic categorization task than the LDT. Additionally, just having an

associative relationship was not enough to produce priming at short SOAs, suggesting that associative information could drive priming via expectancy generation and/or retrospective matching. In contrast, a semantic relationship, which includes feature overlap, could be used quickly and automatically to drive semantic priming in the semantic categorization task. Thus, Bueno and Frenck-Mestre argue that the lexical decision task is not as sensitive to early semantic processing. Similarly, [McRae and Boisvert \(1998\)](#) argue that strategic processes in semantic categorization tasks can be reduced if researchers use broad categories.

In a series of experiments, [de Wit and Kinoshita \(2014, 2015a, 2015b\)](#) examined semantic priming effects in both the lexical decision task and the semantic categorization task. The authors likewise conclude that the mechanisms driving semantic priming are different across tasks. In the lexical decision task, semantic priming effects are due to retrospective semantic matching mechanisms, even when the SOA is too short to allow generation of expectancy sets. In semantic categorization tasks, semantic priming effects are due to automatic facilitation of the target by the prime, although differently from the way posited by spreading activation accounts. In a semantic categorization task (where participants make a judgement about the target word that relies on its meaning), the semantic prime facilitates its target via evidence accumulation and source confusion. Specifically, if the judgement is whether the target is an animal or not, participants accumulate evidence in the form of diagnostic features (e.g., “has four legs”) to make their decision (see also [Norris, 2006](#)). De Wit and Kinoshita propose that evidence accumulated from the prime gets combined with the evidence accumulated from the target (i.e., source confusion) when the two are presented in close temporal proximity. Thus, the strength of the featural relationship (i.e., how much of an overlap of features there is between the prime and target), not necessarily of the associative relationship, influences how strongly the prime can facilitate the target in a semantic categorization task. Retrospective matching should also be discouraged in the semantic categorization task ([de Wit and Kinoshita, 2014](#)). When the SOA between prime and target is short (eliminating the possibility of expectancy set formation) the evidence accumulated from the prime gets confused for evidence accumulated from the target, resulting in the related prime facilitating the target. Although, similar to the argument made by [Jones \(2010\)](#), retrospective matching could consist of attending to a semantic dimension of the target and prime, for example, deciding whether the prime and target belong to the “animal” category. Again, this would be retrospective matching, but not in a way that is consistent with the compound-cue model ([McKoon and Ratcliff, 1992](#)).

de Wit and Kinoshita examined the semantic priming effect in the semantic categorization task ([de Wit and Kinoshita, 2014](#)) and the lexical decision task ([de Wit and Kinoshita, 2015a](#)) as a function of relatedness proportion, with a 240 ms SOA (the prime was presented for 200 ms and there was a 40 ms interstimulus interval). An SOA of 240 ms should preclude the generation of expectancy sets, allowing for the prime to facilitate the target via automatic activation and/or a retrospective strategy. In both tasks, the authors found a greater semantic priming effect with a high relatedness proportion, ruling out an automatic spreading activation as posited by [Collins and Loftus \(1975\)](#) as the only mechanism driving semantic priming. De Wit and Kinoshita then examined the latency distributions of responses. Specifically, a semantic priming effect that is greater on slower trials would be consistent with a retrospective strategy being applied after lexical access to the target, since the prime would have more time to influence the response during slower trials. Of

note, this is similar to the logic used by [Thomas et al. \(2012\)](#), who found a greater semantic priming effect for visually degraded targets (which took longer to process, thus allowing more time for the prime to influence the target), but only the prime and target pair had a backward association. de Wit and Kinoshita found that the semantic priming effect was indeed greater for slower trials in high relatedness proportion condition in the lexical decision task, but was similar across fast and slow trials for the semantic categorization task, suggesting that retrospective matching operated during the LDT but not semantic categorization.

Similarly, [de Wit and Kinoshita \(2015b\)](#) examined the semantic priming effect in the lexical decision and semantic categorization tasks when the primes were able to be consciously identified by participants (the prime was presented for 200 ms followed by a 40 ms blank) or not (prime presented for 50 ms and was immediately followed by the target). In both tasks, a semantic priming effect was found when the primes were visible. Making the prime unavailable for conscious identification eliminated the semantic priming effect in the lexical decision task but not in the semantic categorization task. Additionally, in the lexical decision task, the semantic priming effect was greater for trials with slower responses, while in the semantic categorization task, the size of the priming effect was similar between slower and faster trials. The authors argue their results are consistent with semantic priming being driven by a retrospective matching process in the lexical decision task and by pre-activation in the semantic categorization task. Specifically, eliminating the visibility of the prime (and thus eliminating the use of retrospective matching) eliminated the semantic priming effect in the lexical decision task but not the semantic categorization task. Although we should note that a 50 ms prime presentation may not entirely preclude prime identification ([Hirshman and Durante, 1992](#)).

The collective findings of [de Wit and Kinoshita \(2014, 2015a, 2015b\)](#) are not consistent with the previously discussed findings of [Fernandino and Conant \(2024\)](#), who found that the overlap in sensory and affective features was the primary driver of the automatic semantic priming effect in the LDT (the authors used a 200 ms SOA). It should be noted that while Fernandino and Conant only used the LDT, de Wit and Kinoshita directly compared the LDT and semantic categorization tasks.

The findings of [Taikh and Lupker \(2020\)](#) are also consistent with the semantic priming effect in the lexical decision task being driven by a retrospective process. Taikh and Lupker examined the concurrent effects of semantic primes (at long and short SOAs) and masked orthographic primes on target responses. Specifically, a semantic prime (*mutton*) and then an orthographic prime (*lkmb*) preceded a target (*LAMB*). While the mechanism(s) driving semantic priming are debated, the effect of the masked orthographic prime (which is usually not consciously identified) is thought to be automatic and pre-lexical. The effects of the semantic and orthographic primes did not interact, indicating that the two types of primes influenced separate processes according to [Sternberg's \(1969\)](#) additive factors logic. Additionally, consistent with [De Wit and Kinoshita \(2014, 2015b\)](#), the semantic priming effect increased during slower trials. In contrast, the orthographic priming effect remained consistent between slow and fast trials.

Naming

In naming tasks, participants read the target word aloud. Compared to the LDT, naming tasks may provide a purer measure of

the influence of automatic spreading activation on semantic priming (Balota and Lorch, 1986; Seidenberg et al., 1984), as retrospective matching is less likely to be used for the demands of the naming task. Consistent with this, Hutchison et al. (2008) found that BAS was a stronger predictor of semantic priming in the lexical decision task than in naming at long SOA (see also Kahan et al., 1999). Overall, though, Hutchison et al. found reliable semantic priming effects in both the LDT and naming, although the effect was smaller in naming (Lupker, 1984; Neely, 1991).

Hutchison et al. (2008) found that FAS predicted the semantic priming effect in both the naming task and the LDT, while the BAS predicted the effect in only the LDT. The authors argued the retrospective matching mechanism contributes to semantic priming in the LDT, but not in naming. Heyman et al. (2016) also examined whether the mechanisms driving semantic priming in the LDT and naming are the same using a large semantic priming dataset. Heyman et al. used data from the Semantic Priming Project (Hutchison et al., 2013), a dataset of naming and LDT responses from 768 participants for 1,661 targets following related and unrelated semantic primes at both short (200 ms) and long (1,200 ms) SOAs. The authors reasoned that if the priming effects of the two tasks are driven by the same mechanisms, then item-level priming effects should be correlated. When controlling for prime and target properties (i.e., length, frequency, orthographic neighbourhood size, baseline response latencies in the English Lexicon Project), Heyman et al. found that item-level priming effects were correlated between the two tasks. However, correlations were stronger at the short SOA than the long SOA. When analyzing only items with a strong FAS and no BAS, these findings remained the same, suggesting that in both tasks, primes pre-activate their targets quickly and automatically via automatic spreading activation (Collins and Loftus, 1975) or the encoding of features (de Wit and Kinoshita, 2014, 2015a, 2015b).

Typing

More recently, the typing task has been used to examine the processing and production of language (see Gagné et al., 2023a; Yamaguchi, 2019 for reviews). Typing a word requires multiple processes, including identifying it, accessing its semantic information, planning the sequence of keystrokes, and executing that sequence. Unlike the LDT and semantic categorization tasks, the typing production of a word unfolds over multiple keystrokes. Keystroke latencies can be recorded with precision, thus allowing researchers to examine how linguistic information influences the processes involved in typing. Specifically, the latency of the first keystroke is thought to reflect accessing the word, planning its motor output, and storing the sequence of letters in the graphemic output buffer. The latencies of the non-initial letters are thought to reflect the execution of the motor plan.

Modular theories of typing posit that typing processes are sequential, and are informationally isolated from one another. For example, the Context Retrieval and Updating Model (Logan, 2018; Logan and Crump, 2009, 2011) posits a word level, where a lexical item is selected, and a letter level, where the item is decomposed into letters and these letters are typed (see also Yamaguchi et al., 2017). Importantly, according to modular theories, linguistic information is not present during, and does not influence the execution of the motor plan. Information from semantic primes should thus influence the initial keystroke, but not the non-initial ones. Interactive theories (e.g.,

Gagné and Spalding, 2016) posit that typing processes are cascaded, and thus allow for linguistic information to also influence the execution of the motor plan (i.e., the initial and non-initial keystrokes).

Few studies have examined the influence of semantic priming on typing, and it is not clear whether semantic primes influence the execution of the motor plan in addition to recognizing the word and planning the keystrokes. For example, Chen et al. (2021) analyzed word typing times from TypeRacer, an online game where participants type short passages. The authors found that words that were more predictable based on their context sentence, were typed more quickly. However, Chen et al. did not differentiate between the latencies of the initial and non-initial keystrokes, and thus their findings do not indicate whether contextual predictability influenced keystroke execution. In the TypeRacer game, and other naturalistic copy-typing tasks where the entire passage is available, it is difficult to control the delay between a target word and its preceding contextual information, so the influences of strategic processing of semantic information are not clear.

Scaltritti et al., 2017 examined the influence of semantic primes (duration and SOA were 100 ms) on the typing of auditorily presented targets. Scaltritti et al. argue that semantic priming influences only the word recognition processes (i.e., only the first stage of the Context Retrieval and Updating model) and wanted to test whether information from semantic primes persisted during the execution of the motor plan during the typing of the target words. The authors found that related primes facilitated both the initial keystrokes and the interkey intervals between the first and second keystrokes of auditorily presented targets, but not the average latencies of the non-initial keystrokes. Scaltritti et al. argue that their findings are potentially consistent with the cascaded typing processes as posited by interactive theories of typing, although the effect on the interkey intervals between the first and second keystrokes was small relative to the effect on the initial keystrokes. The short SOA used by Scaltritti et al. precluded the use of expectancy generation, leaving spreading activation and retrospective lexical matching to potentially influence the execution of the motor plan.

Mangat and Taikh (2024) examined the influence of having more time to process semantic primes on typing targets, using a short SOA (275 ms) and a long SOA (600 ms). In both cases, the prime was presented for 200 ms and the duration of the interstimulus interval mask varied (i.e., 75 ms or 400 ms), and targets were visually presented using the type-to-copy paradigm, where the target remains on the screen as it is being typed. Related primes facilitated the initiation of the target words, but there was no effect on any non-initial key latencies, which is consistent with modular theories of typing. Interestingly, both the initial and non-initial target keystrokes were typed faster at the long SOA. Mangat and Taikh suggest that the prime (whether related or unrelated) may have interfered with the production of the target, where the prime was automatically encoded and its letters stored in the graphemic output buffer interfering with planning and executing the keystrokes of the target. A longer interstimulus interval may have allowed participants to clear the graphemic output buffer prior to seeing the target word. Mangat and Taikh also noted that using a type-to-copy task may reduce reliance on linguistic information, including semantic information from the prime because participants may copy target letters.

Current studies thus do not provide strong support for semantic primes facilitating the execution of the motor plan while typing targets. However, findings from studies examining the typing of compound words in a semantic priming paradigm (e.g., Gagné and

Spalding, 2014, 2016; Libben et al., 2021) suggest that semantic primes influence how the semantic information of the constituent parts of the compound is integrated during its typing. Numerous studies (e.g., Gagné et al., 2023b; Gagné and Spalding, 2014, 2016; Gow et al., 2024; Sahel et al., 2008; Will et al., 2006) have found a delay in typing between the words embedded in compounds and pseudo-compounds (e.g., a pause between the final *h* of *high* and the *l* of *light* when typing *highlight*). This boundary effect suggests that embedded words, rather than the entire word, are planning units of production, and that morphological information is active during the execution of the motor plan. Additionally, Taikh et al. found that, in typing compound words the linguistic properties of the second constituent influenced the typing of the first constituent. When the second constituent was more semantically transparent (i.e., the constituent maintained its stand-alone word meaning to a greater extent, such as *berry* in *blueberry*), the first constituent was typed more quickly. The findings of Taikh et al. suggest that keystroke execution is influenced by semantic information, where the meanings of the individual constituents are accessed and integrated during typing.

Several studies (Gagné and Spalding, 2014, 2016; Libben et al., 2021) have examined the typing of compound word targets in the semantic priming paradigm. These studies have confirmed the boundary effect, and the influence of the semantic information of the constituents on the production of the compound targets. In general, semantically opaque constituents (e.g., *wash* in *hogwash*) were typed more slowly because their representations might be suppressed in favor of the meaning of the whole compound. Importantly, the effect of the semantic prime was influenced by the semantic transparency of the compound's constituents. For example, Gagné and Spalding (2016) found that facilitating access to the meaning of the first constituent (e.g., the related prime *chicken* facilitating *egg* in the compound *egg-beater*), slowed down typing of the second constituent. These studies suggest that the semantic information from the prime influences accessing and integrating the semantic information of the compound's constituent morphemes during the typing of the compound.

In summary, task-level determinants can either encourage or preclude strategic processing. For example, the single presentation-LDT is thought to preclude strategic processing, while the typical LDT likely reflects the influence of both automatic and strategic processing (although researchers disagree on what drives semantic priming in the LDT). In general, binary response tasks such as the typical LDT and the semantic categorization likely encourage retrospective matching processes, such that stronger BAS predicts a greater semantic priming effect. Production tasks, such as naming and typing, may rely less on retrospective matching and thus may be better suited for examining the influences of prospective priming mechanisms (i.e., automatic spreading activation and expectancy generation).

Toward an integrated account of semantic priming

Numerous integrated accounts (e.g., Neely et al., 1989; Perea and Rosa, 2002) allow for semantic priming to be driven by both automatic and strategic processes. For example, Perea and Rosa propose an integrated account of semantic priming, where the effect is driven by automatic spreading activation at short SOAs (60 ms or shorter), and by post-lexical integration at longer SOAs. To continue to develop an

integrated understanding of semantic priming a key step is to separate the contributions of automatic and strategic processing of the prime, and similarly, the influences of multiple concurrent mechanisms that can drive the semantic priming effect. Our review of the literature on the semantic priming effect shows that numerous factors (prime-level lexical and semantic properties, prime-target relational factors, modality and presentation variables, and task-level determinants) jointly influence the processing of the prime and the semantic priming effect, and the effects of these factors are often conflated. The influence of these factors is generally greater when strategic processing is possible. In this section, we integrate findings about the influence of these factors on both the automatic and strategic processing of the prime, and examine the key challenges in developing an integrated understanding of semantic priming and suggest future directions.

Considering automatic processing of the prime

Automatic processing of the prime occurs quickly and does not require attention or explicit awareness of the prime (Posner and Snyder, 1975), and has been conceptualized as automatic spreading activation of the prime pre-activating the lexical representation of the target (Collins and Loftus, 1975) or as the extraction of semantic features from the prime that facilitate the processing of the target (e.g., de Wit and Kinoshita, 2014, 2015a, 2015b). Studies that find the semantic priming effect when the prime is presented too briefly to be identified (typically less than 50 ms) are consistent with automatic prime processing (e.g., de Wit and Kinoshita, 2015b; Hirshman and Durante, 1992). Specifically, not identifying the primes does not allow for expectancy generation or post-lexical integration. However, ruling out the influence of post-lexical integration when the prime is visible is difficult. As Hirshman and Durante noted, even at 50 ms durations, primes can be identified and processed concurrently with the target, and they found evidence of post-lexical integration at those durations.

It is also not clear how long the influence of automatic spreading activation persists, although studies have attempted to estimate its duration. For example, Perea and Rosa found a semantic priming effect for semantically related pairs in the single presentation LDT at a short SOA (200 ms), but not a longer SOA (1750 ms). The authors suggested that automatic spreading activation, which they argued is what drives priming for semantically related pairs, persists for less than 1750 ms. However, the authors also note that the absence of a priming effect for these pairs at a long SOA could have been due to expectancy generation of items that are associatively (but not purely semantically) related to the prime. Items generated based on expectancy would thus interfere with the semantically (but not associatively) related pairs. Thus, it is difficult to rule out the influence of automatic spreading activation at an SOA of 1750 ms for purely semantically related pairs, and it is still unclear how long automatic spreading activation persists in prime-target pairs with a different relationship or in a different task.

Studies disagree on how the type of prime-target relationship influences automatic processing of the prime. Lucas (2000) concluded that there was no semantic priming effect driven only by automatic prime processing when the primes and targets were not semantically related but only associatively related (see Perea and Rosa, 2002 for a similar conclusion based on an experiment that also manipulated SOA). In contrast to Lucas, Hutchison (2003) argued that there was evidence for automatic priming based on association alone, and

further argued that mediated priming is an example of this. Similarly, it is unclear how task demands influence automatic prime processing. [de Wit and Kinoshita \(2015b\)](#) found that at SOAs of 50 ms, when the prime supposedly could not be identified, there was a semantic priming effect in the semantic categorization task but not the LDT, arguing that feature extraction (i.e., distributed but not localist accounts of semantic memory) explain automatic priming effects. However, [Hirshman and Durante \(1992\)](#) found a semantic priming effect in the LDT using prime durations of 33 ms. Similarly, [Bueno and Frenck-Mestre \(2008\)](#) found a semantic priming effect for associatively related pairs in the LDT at SOAs of 42 ms, when primes were unlikely to be identified.

An important direction for future studies attempting to develop an integrated account of semantic priming is thus to delineate when automatic processing of the prime drives or contributes to the semantic priming effect. The difficulty ruling out retrospective matching, and to some extent, expectancy generation, and the fact that these two strategies can obscure the influence of automatic processing of the prime make this challenging. However, studies using the single presentation-LDT could examine the influence of numerous factors (e.g., semantic vs. automatic prime-target relationship, and prime-level lexical and semantic properties). Similarly, studies extending [De Wit and Kinoshita \(2015b\)](#) could examine the effect of primes that cannot be consciously identified (i.e., prime duration being 33 ms instead of 50 ms where primes can still be identified, see [Hirshman and Durante, 1992](#)) in numerous tasks. Additionally, [Hutchison \(2003\)](#) argues that mediated priming precludes the use of expectancy generation and retrospective matching, although other researchers (e.g., [Jones, 2010](#); [McKoon and Ratcliff, 1992](#)) argue that mediated priming can be explained by strategic processing of the prime. Examining mediated priming using a single presentation-LDT, or in a standard LDT with short durations, would clarify the mechanism(s) driving mediated priming. Such studies would provide boundary conditions under which automatic processing of the prime influences the semantic priming effect.

Considering strategic processing of the prime

With greater prime duration and SOA, strategic prime processing plays a greater role in driving the semantic priming effect. Studies examining the semantic priming effect suggest that numerous variables allow and encourage the use of expectancy generation and post-lexical integration strategies. Generating expectancy sets is possible if the prime is consciously identified, and SOA is long enough (estimated to be 300 ms or longer, see [Neely, 1977](#)). The use of expectancy generation should be more likely when the stimulus list has a high relatedness proportion, and when the prime-target relationship is consistent and obvious (e.g., prime and target are members of the same category). Retrospective processing of the prime should also be possible whenever the prime can be visually identified. In theory, retrospective processing does not require a minimum SOA because the prime is being processed concurrently with the target. However, retrospective processing of the prime makes a greater contribution to the semantic priming effect when processing of the target words is slowed down, either due to visual degradation of the target words (e.g., [Balota et al., 2008](#); [Borowsky and Besner, 1993](#); [Stolz and Neely, 1995](#); [Thomas et al., 2012](#)), or when the prime is a higher frequency word ([Hutchison](#)

[et al., 2008](#)). In general, [Hutchison et al.](#) argue that strategic processing of the prime has a greater impact on the target when the prime is processed more quickly, and when the target is processed more slowly (allowing the more time for the related prime to influence the processing of the target).

The demands of the experimental tasks also influence how the prime is processed and thus the contributions of expectancy generation and retrospective matching to the semantic priming effect. Naming or typing targets requires multiple sequential responses (i.e., articulating a sequence of sounds or executing a sequence of keystrokes), while semantic categorization and the LDT require a single response for the target. [Balota and Lorch \(1986](#); see also [Seidenberg et al., 1984](#)) argue that the demands of the naming task are much less likely to require retrospective matching (although expectancy generation is still possible). The findings of [Hutchison et al. \(2008\)](#), that BAS was a weaker predictor of priming in naming than the lexical decision task at a long SOA, are consistent with this notion. Using data from the Semantic Priming Project ([Hutchison et al., 2013](#)), [Heyman et al. \(2016\)](#); found that naming and LDT item-level priming effects were correlated, but this correlation was stronger at the short SOA. Together, the findings of [Hutchison et al. \(2008\)](#) and [Heyman et al. \(2015\)](#) suggest that naming reduces or eliminates retrospective processing of the prime. In the naming task, the semantic priming effect may be driven by automatic processing of the prime (i.e., automatic spreading activation or the extraction of features), or by expectancy generation, but the prime does not seem to be processed while the target is being produced. Studies examining the semantic priming effect in typing (e.g., [Mangat and Taikh, 2024](#); [Scaltritti et al., 2017](#)) likewise suggest that the influence of the prime on the production of the target may be prospective, influencing the identification of the target and the planning of its production (i.e., the latency of the first keystroke), but not the execution of the motor plan (i.e., the non-initial keystrokes). However, the finding by [Mangat and Taikh](#) that initial and non-initial target keystrokes were typed more quickly when the SOA was long (following related and unrelated primes) suggests that in the typing task, the prime may be processed while the target is being typed. Specifically, the letters of the prime, related or unrelated, may have been encoded and stored in preparation for their output (although the prime was not typed in this task). Thus, processing the prime as a stimulus that may need to be typed may have interfered with the concurrent output of the target. We note that there have been relatively few studies examining the semantic priming paradigm in typing.

Developing an integrated account of semantic priming requires delineating when strategic processes, such as expectancy generation and retrospective matching, drive the semantic priming effect. One challenge to this is accounting for or eliminating the influence of automatic processing of the prime. Boundary conditions under which automatic processing contributes to semantic priming, as discussed above, would be informative. Similarly, it is difficult to rule out the influence of retrospective matching strategies to isolate the influence of expectancy generation. Using prime-target pairs that have a strong backward association but no forward association ([Thomas et al., 2012](#)) may allow for examining the influence of expectancy generation in various tasks. In addition to numerous contradicting results and arguments in the literature that were noted in the present review (e.g., [Lucas, 2000](#) vs. [Hutchison, 2003](#)), another challenge to developing an integrated account of semantic priming is that many factors that influence semantic priming are studied in isolation (e.g., [Hutchison et al.,](#)

2013). One solution to this problem is to use large datasets of semantic priming, which we discuss next. First, though, we note that some factors cannot be studied concurrently and their interactions cannot be observed. For example, it would not be possible to control the SOA between prime and target, or the time participants spend processing the prime, in studies where participants respond to both the prime and the target sequentially (e.g., single presentation-LDT).

Using large datasets of semantic priming

The majority of studies examining the semantic priming effect utilize factorial designs, which makes examining the factors driving the effect more challenging (Hutchison et al., 2013). Specifically, in such studies, the researchers select stimuli based on a variable of interest (e.g., associative vs. categorial prime-target relations) and attempt to equate the stimuli on other variables that influence word recognition and semantic priming. Because many variables can potentially influence word recognition and the semantic priming effect, stimuli in semantic priming studies cannot be effectively equated on all the other relevant variables. For example, as Hutchison (2003) notes, the influence of prime-target co-occurrence is difficult to separate from the influence of semantic relatedness because prime-target pairs that are semantically related often occur in the same linguistic context. Hutchison et al. also note that factorial designs that use unusual items (e.g., very high vs. very low word frequency) may lead participants to change the way they process primes and targets, reducing the generalizability of these studies. Hutchison et al. (see also Buchanan et al., 2026) argue that large datasets of semantic priming data, such as their Semantic Priming Project, can overcome some of these limitations. The Semantic Priming Project includes naming and LDT responses to 1,661 target words from 768 participants. The target words follow both related and unrelated primes, allowing Hutchison et al. to estimate item-level semantic priming effects when averaged over participants. Hutchison et al. note that datasets that include many items would allow item characteristics to vary continuously.

We note, however, that in a dataset with many items and thus variation on many variables, the effect of a specific variable of interest (e.g., type of prime-target relationship) may be obscured by the other variables. Sibley et al. (2009) attempted to replicate effects of numerous linguistic variables on naming latencies that are reliably obtained in factorial experiments using large datasets. For example, the robust finding that words with irregular (vs. regular) pronunciations are named more slowly, but only when they are low frequency (e.g., Seidenberg, 1985; Taraban and McClelland, 1987), was generally not replicated using datasets from megastudies such as the English Lexicon Project (Balota et al., 2008). Sibley et al. also found only moderate item-level correlations in naming latencies between megastudy datasets. The authors argued that large databases may not be sensitive enough to detect subtle, but theoretically significant effects that can be detected using factorial studies. Factorial studies are used to deliberately create conditions where the variable of interest can be examined. Using large datasets is thus not always a suitable replacement for factorial experiments.

Another challenge of examining semantic priming effects is that they are generally not reliable. In addition to examining the reliability of item-level priming effects across tasks (LDT vs. naming) in the Semantic Priming Project dataset (Hutchison et al., 2013), Heyman et al. (2016) examined the reliability of the semantic priming effects

within each task and SOA grouping (i.e., short and long SOA groups in the naming and LDT tasks). The authors split the participant sample of each of the four groupings of data in half and computed correlations of the item-level priming effects. The correlation coefficients were similar for the two tasks and were greater when the SOA was short (spearman-brown estimates were about 0.3) than when SOA was long (spearman-brown estimates were about 0.2). Heyman et al. (2016) note that estimates indicate that semantic priming effects have low reliability, and are similar to reliability estimates computed in other studies (e.g., Heyman et al., 2015). Similarly, Heyman et al. (2018) examined the reliability of item-level semantic priming effects across two different times. In their study, participants completed two sessions that were 4 weeks apart, where the stimuli used in each session were identical. There were 134 critical targets, where the prime-target relationships were diverse (e.g., synonyms, antonyms, subordinates...). The SOA was 200 ms (the prime was shown for 150 ms followed by a 50 ms ISI prior to the target), which was meant to discourage strategic processing of the primes. Heyman et al. (2018) found low item-level test-retest reliability (less than 0.30), which is consistent with the findings of Heyman et al. (2015) and other studies examining the reliability of the semantic priming effect (e.g., Heyman et al., 2016).

Heyman et al. (2016, 2018) note that increasing sample size would increase the reliability estimates of item-level semantic priming effects. Individual participants may vary on how strongly two concepts (i.e., the prime and its target) are related, and on strategies they may use when the prime duration and SOA allow for strategic processing. While individual differences in semantic priming effects are beyond the scope of the present paper (see Yap et al., 2017), averaging over participants to compute item-level effects introduces variability, and increasing the number of participants would reduce this variability. For example, Heyman et al. (2018) computed the split-half reliability of semantic priming effects from randomly selected participants from data from Tan and Yap (2016), who tested the masked priming effect and used a sample of 240 participants. Hutchison et al. (2013) obtained a split-half reliability of 0.70 from the complete Tan and Yap dataset, but only an estimate of 0.28 from a subset of the data that included 40 participants.

Summary and concluding remarks

The semantic priming paradigm has been a key way in which researchers have tried to understand the representation and retrieval of knowledge. Given that the semantic priming effect in any particular situation can result from automatic and/or strategic processes, our aim is to, whenever possible, delineate boundary conditions between when semantic priming is driven by automatic vs. strategic processing of the prime. Consistent semantic priming effects are most likely when strategic influences are minimized. Most notably, greater semantic priming is predicted by primes that are processed more quickly (i.e., short and frequent words with few orthographic neighbors) when the SOA is short and automatic processes are more likely to drive the semantic priming effect. However, it is still not clear whether automatic processing drives the semantic priming effect only when the prime and target are semantically related (e.g., Lucas, 2000; Perea and Rosa, 2002) or if it is also possible when the prime and target are associatively related

(Hutchison, 2003). It is also unclear whether automatic processing can drive semantic priming in the lexical decision task (e.g., Bueno and Frenck-Mestre, 2008; Hirshman and Durante, 1992), or not (e.g., de Wit and Kinoshita, 2015b).

In contrast, the influence of strategic processing is more likely to result in inconsistencies of the semantic priming effect because strategic processing is more likely to be influenced by the factors reviewed here. Examining how numerous factors jointly influence the semantic priming effect, especially when it is driven by strategic processing, is important for delineating boundary conditions under which automatic and strategic processing drive the semantic priming effect. For example, novel word primes are only effective in priming their targets when the SOA is long enough (i.e., 1,000 ms) to process them strategically (Ball et al., 2024), as they are not yet integrated into the semantic network. Similarly, Hutchison et al. (2008) found that at long SOA (1,250 ms), when strategic processing was more likely, priming was predicted by both FAS and BAS. However, at short SOA (250 ms), when strategic processing was less likely, Hutchison et al. found that priming was predicted only by FAS. Similarly, while relatedness proportion was found to influence semantic priming at shorter SOAs (e.g., de Groot, 1984; de Wit and Kinoshita, 2014; de Wit and Kinoshita, 2015a), its influence was stronger at longer SOAs, and numerous studies did not find the influence of relatedness proportion at short SOAs (e.g., Perea and Rosa, 2002). Collectively, such findings help to narrow down the boundary conditions under which

semantic priming is driven by automatic or strategic processing of the prime.

However, as Hutchison et al. (2013) points out, many factors that influence semantic priming are studied in isolation. Continuing to examine the joint influence of numerous factors would contribute to delineating conditions where automatic and strategic prime processing contributes to the semantic priming effect. We note some outstanding questions about the automatic and strategic drivers of the semantic priming effect as a function of time available to process the prime in Table 1. For example, the field would benefit from studies that examine the joint influence of a prime's semantic richness and relatedness proportion. Few studies have examined semantic priming in the typing task. As in the naming task, typing a target unfolds sequentially, rather than ending after a single keystroke as in the LDT, and this likely influences the strategic processing of the prime and how long this information persists. Again, we note that some factors, and their interactions, cannot be studied concurrently. For example, in a paradigm where the prime and the target are sequentially presented and the participant responds to both (such as the single presentation-LDT), it is impossible to control the SOA between prime and target and thus to examine the influence of SOA in such a task. While using large datasets of semantic priming may facilitate examining the joint effects of multiple variables on semantic priming (Hutchison et al., 2013), they may not be sensitive enough to detect subtle effects that are

TABLE 1 Outstanding questions about automatic and strategic influences on the semantic priming effect at different prime processing times.

Time to process prime	Driver(s) of semantic priming	Outstanding questions
Prime not visible (prime duration under 50 ms)	Automatic processing of prime	Do priming effects driven purely by automatic processing of the prime require a semantic prime - target relationship? Is automatic semantic priming driven by automatic spreading activation or feature overlap? What tasks allow automatic priming effects? This is especially pertinent to the standard LDT.
Prime is visible and SOA is short (under about 300 ms)	Both automatic and strategic processing of prime (although automatic processing is likely to be the dominant contributor)	Do retrospective integration mechanisms drive semantic priming in tasks that require producing the targets (e.g., typing or naming)? Do retrospective integration mechanisms drive semantic priming in mediated priming paradigms? How much time is required to process the meaning of the prime, thus allowing for prospective mechanisms to drive semantic priming? Are the contributions of retrospective matching and automatic prime processing to the semantic priming effect additive? Do lexical and semantic properties of the prime that facilitate its processing (e.g., greater frequency) facilitate both automatic and strategic drivers of semantic priming?
Prime is visible and SOA is medium or long (over 300 ms)	automatic and strategic processing of prime, although strategic processing becomes more dominant with longer SOA	Is expectancy generation possible in mediated priming paradigms? How long does the influence of automatic prime processing persist? What factors influence this period of time when automatic processing of the prime influences processing of the target? Are the contributions of expectancy generation, retrospective matching, and automatic prime processing to the semantic priming effect additive? What is the minimum SOA needed for novel word primes to produce a semantic priming effect?

still theoretically important (Sibley et al., 2009). Specifically, factorial experiments are deliberately created to examine these variables by using specific types of stimuli, prime-target relationships, presentation variables, and task types. Thus factorial experiments and large datasets are complementary in continuing to examine the joint influences of multiple variables on the semantic priming effect.

In conclusion, both automatic and strategic processing of the prime can concurrently contribute to the semantic priming effect, where strategic processing can obscure the influence of automatic processing. Strategic processing of the prime is more likely than automatic processing to be influenced by a wide array of factors, such as prime-level lexical and semantic properties, prime-target relational factors, modality and presentation variables, and task-level determinants. To develop an integrated account of semantic priming, it is critical to delineate conditions when automatic or strategic processing is contributing to the semantic priming effect. Our review aimed to delineate these boundary conditions through highlighting the joint influence of various factors on the semantic priming effect when possible. We also note contradictory findings in the literature, and the limits of integrating multiple task types, modalities, and temporal parameters, and suggest future directions.

Author contributions

AM: Investigation, Project administration, Writing – original draft, Writing – review & editing. AT: Investigation, Project administration, Supervision, Writing – original draft, Writing – review & editing. CG: Writing – review & editing. TS: Writing – review & editing.

References

- Anderson, J. R. (1983). A spreading activation theory of memory. *J. Verbal Learn. Verbal Behav.* 22, 261–295. doi: 10.1016/S0022-5371(83)90201-3
- Andrews, S. (1997). The effect of orthographic similarity on lexical retrieval: resolving neighborhood conflicts. *Psychon. Bull. Rev.* 4, 439–461. doi: 10.3758/BF03214334
- Bakker, I., Takashima, A., van Hell, J. G., Janzen, G., and McQueen, J. M. (2015). Tracking lexical consolidation with ERPs: lexical and semantic-priming effects on N400 and LPC responses to newly-learned words. *Neuropsychologia* 79, 33–41. doi: 10.1016/j.neuropsychologia.2015.10.020
- Bakker-Marshall, I., Takashima, A., Schoffelen, J. M., Van Hell, J. G., Janzen, G., and McQueen, J. M. (2018). Theta-band oscillations in the middle temporal gyrus reflect novel word consolidation. *J. Cogn. Neurosci.* 30, 621–633. doi: 10.1162/jocn_a_01240
- Balass, M., Nelson, J. R., and Perfetti, C. A. (2010). Word learning: an ERP investigation of word experience effects on recognition and word processing. *Contemp. Educ. Psychol.* 35, 126–140. doi: 10.1016/j.cedpsych.2010.04.001
- Ball, L. V., Brusini, P., and Bannard, C. (2024). Revisiting novel word semantic priming: the role of strategic priming mechanisms. *Q. J. Exp. Psychol.* 78, 2284–2298. doi: 10.1177/17470218241306747
- Balota, D., Ferraro, F., and Connor, L. (1991). “On the early influence of meaning in word recognition: a review of the literature” in *The psychology of word meanings*. ed. P. J. Schwanenflugel (Hillsdale, NJ: Erlbaum), 187–222.
- Balota, D. A., and Lorch, R. F. (1986). Depth of automatic spreading activation: mediated priming effects in pronunciation but not in lexical decision. *J. Exp. Psychol. Learn. Mem. Cogn.* 12, 336–345. doi: 10.1037/0278-7393.12.3.336
- Balota, D. A., Yap, M. J., Cortese, M. J., and Watson, J. M. (2008). Beyond mean response latency: response time distributional analyses of semantic priming. *J. Mem. Lang.* 59, 495–523. doi: 10.1016/j.jml.2007.10.004
- Barca, L., Benedetti, F., and Pezzulo, G. (2016). The effects of phonological similarity on the semantic categorisation of pictorial and lexical stimuli: evidence from continuous behavioural measures. *J. Cogn. Psychol.* 28, 159–170. doi: 10.1080/20445911.2015.1101117
- Becker, C. A. (1979). Semantic context and word frequency effects in visual word recognition. *J. Exp. Psychol. Hum. Percept. Perform.* 5, 252–259. doi: 10.1037/0096-1523.5.2.252
- Becker, C. A. (1980). Semantic context effects in visual word recognition: an analysis of semantic strategies. *Mem. Cogn.* 8, 493–512. doi: 10.3758/BF03213769
- Becker, S., Moscovitch, M., Behrmann, M., and Joordens, S. (1997). Long-term semantic priming: a computational account and empirical evidence. *J. Exp. Psychol. Learn. Mem. Cogn.* 23, 1059. doi: 10.1037/0278-7393.23.5.1059
- Borovsky, A., Elmann, J. L., and Kutas, M. (2012). Once is enough: N400 indexes semantic integration of novel word meanings from a single exposure in context. *Lang. Learn. Dev.* 8, 278–302. doi: 10.1080/15475441.2011.614893
- Borowsky, R., and Besner, D. (1993). Visual word recognition: a multistage activation model. *J. Exp. Psychol. Learn. Mem. Cogn.* 19, 813. doi: 10.1037//0278-7393.19.4.813
- Buchanan, E. M., Cuccolo, K., Heyman, T., van Berkel, N., Coles, N. A., Iyer, A., et al. (2026). Measuring the semantic priming effect across many languages. *Nat. Hum. Behav.* 10, 182–201. doi: 10.1038/s41562-025-02254-x
- Buchanan, L., Westbury, C., and Burgess, C. (2001). Characterizing semantic space: neighborhood effects in word recognition. *Psychon. Bull. Rev.* 8, 531–544. doi: 10.3758/BF03196189
- Bueno, S., and Frenck-Mestre, C. (2008). The activation of semantic memory: effects of prime exposure, prime-target relationship, and task demands. *Mem. Cogn.* 36, 882–898. doi: 10.3758/MC.36.4.882
- Carr, T. H., McCauley, C., Sperber, R. D., and Parmelee, C. M. (1982). Words, pictures, and priming: on semantic activation, conscious identification, and the automaticity of information processing. *J. Exp. Psychol. Hum. Percept. Perform.* 8, 757–777. doi: 10.1037/0096-1523.8.6.757

Funding

The author(s) declared that financial support was not received for this work and/or its publication.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that Generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Cheesman, J., and Merikle, P. M. (1986). Distinguishing conscious from unconscious perceptual processes. *Canad. J. Psychol.* 40, 343–367. doi: 10.1037/h0080103
- Chen, R., Levy, R., and Eisape, T. (2021). On factors influencing typing time: insights from a viral online typing game. *Proceedings of the annual meeting of the cognitive science society*.
- Chwilla, D. J., and Kolk, H. H. (2002). Three-step priming in lexical decision. *Mem. Cogn.* 30, 217–225. doi: 10.3758/BF03195282
- Collins, A. M., and Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychol. Rev.* 82, 407–428. doi: 10.1037/0033-295X.82.6.407
- Cortese, M. J., Simpson, G. B., and Woolsey, S. (1997). Effects of association and image-ability on phonological mapping. *Psychon. Bull. Rev.* 4, 226–231. doi: 10.3758/BF03209397
- Dark, V. J. (1988). Semantic priming, prime reportability, and retroactive priming are interdependent. *Mem. Cogn.* 16, 299–308. doi: 10.3758/BF03197040
- Dark, V. J., and Benson, K. (1991). Semantic priming and identification of near threshold primes in a lexical decision task. *Q. J. Exp. Psychol. Sec. A* 43, 53–78. doi: 10.1080/14640749108400999
- Davies, P. D. M. (1998). Automatic semantic priming: the contribution of lexical-and semantic-level processes. *Eur. J. Cogn. Psychol.* 10, 389–412. doi: 10.1080/713752286
- Davis, M. H., and Gaskell, M. G. (2009). A complementary systems account of word learning: neural and behavioural evidence. *Philos. Trans. R. Soc. B Biol. Sci.* 364, 3773–3800. doi: 10.1098/rstb.2009.0111
- De Deyne, S., Navarro, D. J., Perfors, A., Brysbaert, M., and Storms, G. (2019). The “small world of words” English word association norms for over 12,000 cue words. *Behav. Res. Methods* 51, 987–1006. doi: 10.3758/s13428-018-1115-7
- de Groot, A. M. (1983). The range of automatic spreading activation in word priming. *J. Verbal Learn. Verbal Behav.* 22, 417–436. doi: 10.1016/S0022-5371(83)90273-6
- de Groot, A. M. B. (1984). Primed lexical decision: combined effects of the proportion of related prime-target pairs and the stimulus-onset asynchrony of prime and target. *Q. J. Exp. Psychol. Sec. A* 36, 253–280. doi: 10.1080/14640748408402158
- de Groot, A. M. B. (1990). “The locus of the associative priming effect in the mental lexicon” in *Comprehension processes in reading*, eds. D. A. Balota and G. B. Flores d’Arcais (Hillsdale, NJ: Erlbaum), 101–123.
- de Wit, B., and Kinoshita, S. (2014). Relatedness proportion effects in semantic categorization: reconsidering the automatic spreading activation process. *J. Exp. Psychol. Learn. Mem. Cogn.* 40, 1733–1744. doi: 10.1037/xlm0000004
- de Wit, B., and Kinoshita, S. (2015a). The masked semantic priming effect is task dependent: reconsidering the automatic spreading activation process. *J. Exp. Psychol. Learn. Mem. Cogn.* 41, 1062–1075. doi: 10.1037/xlm0000074
- de Wit, B., and Kinoshita, S. (2015b). An RT distribution analysis of relatedness proportion effects in lexical decision and semantic categorization reveals different mechanisms. *Mem. Cogn.* 43, 99–110. doi: 10.3758/s13421-014-0446-6
- Fernandino, L., and Conant, L. L. (2024). The primacy of experience in language processing: semantic priming is driven primarily by experiential similarity. *Neuropsychologia* 201:108939. doi: 10.1016/j.neuropsychologia.2024.108939
- Fischler, I. (1977). Semantic facilitation without association in a lexical decision task. *Mem. Cogn.* 5, 335–339. doi: 10.3758/BF03197580
- Forster, K. I. (1981). Priming and the effects of sentence and lexical contexts on naming time: evidence for autonomous lexical processing. *Q. J. Exp. Psychol.* 33, 465–495. doi: 10.1080/14640748108400804
- Gagné, C. L., and Spalding, T. L. (2014). Typing time as an index of morphological and semantic effects during English compound processing. *Lingue Linguaggio* 13, 241–262. doi: 10.1418/78409
- Gagné, C. L., and Spalding, T. L. (2016). Effects of morphology and semantic transparency on typing latencies in English compound and pseudocompound words. *J. Exp. Psychol. Learn. Mem. Cogn.* 42, 1489–1495. doi: 10.1037/xlm0000258
- Gagné, C. L., Spalding, T. L., and Taikh, A. (2023a). Carport and carpet: effects of compound and pseudocompound word structures on typing. *J. Exp. Psychol. Learn. Mem. Cogn.* 49:2003. doi: 10.1037/xlm0001299
- Gagné, C., Spalding, T. L., and Taikh, A. (2023b). “Impact of morphology on written word production: an overview of empirical evidence and theoretical implications” in *Linguistic morphology in the mind and brain*, ed. Crepaldi, E. Routledge, 96–107.
- Gow, K., Taikh, A., and Islam, M. (2024). The influence of uppercase letter location on typing multiword passphrases. *Ment. Lexicon* 19, 161–169. doi: 10.1075/ml.24014.gow
- Grainger, J. (1998). Masked priming by translation equivalents in proficient bilinguals. *Lang. Cogn. Process.* 13, 601–623. doi: 10.1080/016909698386393
- Grossi, G. (2006). Relatedness proportion effects on masked associative priming: an ERP study. *Psychophysiology* 43, 21–30. doi: 10.1111/j.1469-8986.2006.00383.x
- Hayes, S. C., and Bissett, R. T. (1998). Derived stimulus relations produce mediated and episodic priming. *Psychol. Rec.* 48, 617–630. doi: 10.1007/BF03395293
- Heyman, T., Bruninx, A., Hutchison, K. A., and Storms, G. (2018). The (un) reliability of item-level semantic priming effects. *Behav. Res. Methods* 50, 2173–2183. doi: 10.3758/s13428-018-1040-9
- Heyman, T., De Deyne, S., Hutchison, K. A., and Storms, G. (2015). Using the speeded word fragment completion task to examine semantic priming. *Behav. Res. Methods* 47, 580–606. doi: 10.3758/s13423-015-0932-2
- Hirshman, E., and Durante, R. (1992). Prime identification and semantic priming. *J. Exp. Psychol. Learn. Mem. Cogn.* 18:255. doi: 10.1037/0278-7393.18.2.255
- Heyman, K. A., and Storms, G. (2016). Uncovering underlying processes of semantic priming by correlating item-level effects. *Psychonomic Bulletin and Review*, 23:540–547. doi: 10.3758/s13423-015-0932-2
- Hutchison, K. A. (2003). Is semantic priming due to association strength or feature overlap? A microanalytic review. *Psychon. Bull. Rev.* 10, 785–813. doi: 10.3758/BF03196544
- Hutchison, K. A. (2007). Attentional control and the relatedness proportion effect in semantic priming. *J. Exp. Psychol. Learn. Mem. Cogn.* 33, 645–662. doi: 10.1037/0278-7393.33.4.645
- Hutchison, K. A., Balota, D. A., Cortese, M. J., and Watson, J. M. (2008). Predicting semantic priming at the item level. *Q. J. Exp. Psychol.* 61, 1036–1066. doi: 10.1080/17470210701438111
- Hutchison, K. A., Balota, D. A., Neely, J. H., Cortese, M. J., Cohen-Shikora, E. R., Tse, C. S., et al. (2013). The semantic priming project. *Behav. Res. Methods* 45, 1099–1114. doi: 10.3758/s13428-012-0304-z
- Hutchison, K. A., Neely, J. H., and Johnson, J. D. (2001). With great expectations, can two “wrongs” prime a “right”? *J. Exp. Psychol. Learn. Mem. Cogn.* 27, 1451–1463. doi: 10.1037/0278-7393.27.6.1451
- Jared, D., and Seidenberg, M. S. (1991). Does word identification proceed from spelling to sound to meaning? *J. Exp. Psychol. Gen.* 120:358. doi: 10.1037/0096-3445.120.4.358
- Jones, L. L. (2010). Pure mediated priming: a retrospective semantic matching model. *J. Exp. Psychol. Learn. Mem. Cogn.* 36, 135–146. doi: 10.1037/a0017517
- Jones, L. L., and Estes, Z. (2012). “Lexical priming: associative, semantic, and thematic influences on word recognition” in *Visual word recognition, Vol. 2: Meaning and context, individuals and development*, ed. J. S. Adelman (London, UK: Taylor and Francis), 44–72.
- Kahan, T. A., Neely, J. H., and Forsythe, W. J. (1999). Dissociated backward priming effects in lexical decision and pronunciation tasks. *Psychon. Bull. Rev.* 6, 105–110. doi: 10.3758/BF03210816
- Landauer, T. K., and Dumais, S. T. (1997). A solution to Plato’s problem: the latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychol. Rev.* 104:211. doi: 10.1037/0033-295X.104.2.211
- Landauer, T. K., Foltz, P. W., and Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Process.* 25, 259–284. doi: 10.1080/01638539809545028
- Libben, G., Gallant, J., and Dressler, W. U. (2021). Textual effects in compound processing: a window on words in the world. *Front. Commun.* 6:646454. doi: 10.3389/fcomm.2021.646454
- Livesay, K., and Burgess, C. (1998). Mediated priming in high-dimensional semantic space: no effect of direct semantic relationships or co-occurrence. *Brain Cogn.* 37, 102–105.
- Logan, G. D. (2018). Automatic control: how experts act without thinking. *Psychol. Rev.* 125, 453–485. doi: 10.1037/rev0000100
- Logan, G. D., and Crump, M. J. (2009). The left hand doesn’t know what the right hand is doing: the disruptive effects of attention to the hands in skilled typewriting. *Psychol. Sci.* 20, 1296–1300. doi: 10.1111/j.1467-9280.2009.02442.x
- Logan, G. D., and Crump, M. J. (2011). “Hierarchical control of cognitive processes: the case for skilled typewriting” in *Psychology of learning and motivation*, Academic Press, 54, 1–27.
- Lucas, M. (2000). Semantic priming without association: a meta-analytic review. *Psychon. Bull. Rev.* 7, 618–630. doi: 10.3758/BF03212999
- Lund, K., Burgess, C., and Atchley, R. A. (1995). Semantic and associative priming in high-dimensional semantic space. *Proceedings of the annual meeting of the cognitive science society*
- Lund, K., Burgess, C., and Audet, C. (1996). Dissociating semantic and associative word relationships using high-dimensional semantic space. *Proceedings of the annual meeting of the cognitive science society*
- Lupker, S. J. (1984). Semantic priming without association: a second look. *J. Verbal Learn. Verbal Behav.* 23, 709–733. doi: 10.1016/S0022-5371(84)90434-1
- Mangat, A., and Taikh, A. (2024). The influence of semantic primes on the typing of word targets. *Ment. Lexicon* 19, 126–134. doi: 10.1075/ml.24015.man
- Marcel, A. J. (1983). Conscious and unconscious perception: experiments on visual masking and word recognition. *Cogn. Psychol.* 15, 197–237. doi: 10.1016/0010-0285(83)90009-9
- McKoon, G., and Ratcliff, R. (1992). Spreading activation versus compound cue accounts of priming: mediated priming revisited. *J. Exp. Psychol. Learn. Mem. Cogn.* 18:1155. doi: 10.1037//0278-7393.18.6.1155
- McNamara, T. P. (1992). Theories of priming: I. Associative distance and lag. *J. Exp. Psychol. Learn. Mem. Cogn.* 18:1173. doi: 10.1037/0278-7393.18.6.1173
- McNamara, T. P., and Altarriba, J. (1988). Depth of spreading activation revisited: semantic mediated priming occurs in lexical decisions. *J. Mem. Lang.* 27, 545–559. doi: 10.1016/0749-596X(88)90025-3

- McRae, K., and Boisvert, S. (1998). Automatic semantic similarity priming. *J. Exp. Psychol. Learn. Mem. Cogn.* 24:558. doi: 10.1037/0278-7393.24.3.558
- McRae, K., Cree, G. S., Seidenberg, M. S., and McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behav. Res. Methods* 37, 547–559. doi: 10.3758/BF03192726
- Meyer, D. E., and Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: evidence of a dependence between retrieval operations. *J. Exp. Psychol.* 90, 227–234. doi: 10.1037/h0031564
- Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: roles of inhibitionless spreading activation and limited-capacity attention. *J. Exp. Psychol. Gen.* 106, 226–254. doi: 10.1037/0096-3445.106.3.226
- Neely, J. H. (1991). “Semantic priming effects in visual word recognition: a selective review of current findings and theories” in *Basic processes in reading: Visual word recognition*. eds. D. Besner and G. W. Humphreys (Hillsdale, NJ: Erlbaum), 264–336.
- Neely, J. H., and Keefe, D. E. (1989). “Semantic context effects on visual word processing: a hybrid prospective/retrospective processing theory” in *The psychology of learning and motivation: Advances in research and theory*. ed. G. H. Bower, vol. 24 (New York, NY: Academic Press), 207–284.
- Neely, J. H., Keefe, D. E., and Ross, K. L. (1989). Semantic priming in the lexical decision task: roles of prospective prime-generated expectancies and retrospective semantic matching. *J. Exp. Psychol. Learn. Mem. Cogn.* 15, 1003–1019. doi: 10.1037/0278-7393.15.6.1003
- Nelson, D. L., McKinney, V. M., Gee, N. R., and Janczura, G. A. (1998). Interpreting the influence of implicitly activated memories on recall and recognition. *Psychol. Rev.* 105, 299–324. doi: 10.1037/0033-295X.105.2.299
- Norris, D. (2006). The Bayesian reader: explaining word recognition as an optimal Bayesian decision process. *Psychol. Rev.* 113:327. doi: 10.1037/0033-295X.113.2.327
- Pecher, D., Zeelenberg, R., and Raaijmakers, J. G. W. (2002). Associative priming in a masked perceptual identification task: evidence for automatic processes. *Q. J. Exp. Psychol. A Hum. Exp. Psychol.* 55A, 1157–1173. doi: 10.1080/02724980244000143
- Perea, M., and Rosa, E. (2002). The effects of associative and semantic priming in the lexical decision task. *Psychol. Res.* 66, 180–194. doi: 10.1007/s00426-002-0086-5
- Perfetti, C. A., Landi, N., and Oakhill, J. (2005). “The Acquisition of Reading Comprehension Skill” in *The science of reading: a handbook*. eds. M. J. Snowling and C. Hulme (Oxford: Blackwell Publishing), 227–247.
- Pexman, P. M. (2012). “Meaning-based influences on visual word recognition” in *Visual word recognition volume 2*, ed. Adelman, J. (Psychology Press), 24–43.
- Pexman, P. M. (2020). How does meaning come to mind? Four broad principles of semantic processing. *Canad. J. Exp. Psychol.* 74, 275–283. doi: 10.1037/cep0000235
- Pexman, P. M., Hargreaves, I. S., Siakaluk, P. D., Bodner, G. E., and Pope, J. (2008). There are many ways to be rich: effects of three measures of semantic richness on visual word recognition. *Psychon. Bull. Rev.* 15, 161–167. doi: 10.3758/PBR.15.1.161
- Pexman, P. M., Lupker, S. J., and Hino, Y. (2002). The impact of feedback semantics in visual word recognition: number-of-features effects in lexical decision and naming tasks. *Psychon. Bull. Rev.* 9, 542–549. doi: 10.3758/BF03196311
- Plaut, D. C. (1995) Semantic and associative priming in a distributed attractor network. In J. D. Moore and J. F. Lehman (Eds.), *Proceedings of the 17th annual conference of the cognitive science society* (pp. 37–42)
- Plaut, D. C., and Booth, J. R. (2000). Individual and developmental differences in semantic priming: empirical and computational support for a single-mechanism account of lexical processing. *Psychol. Rev.* 107, 786–823. doi: 10.1037/0033-295X.107.4.786
- Posner, M. I., and Snyder, C. R. (1975). “Attention and cognitive control” in *Information processing and cognition*. ed. R. L. Solso (Hillsdale, NJ: Erlbaum), 55–85.
- Ratcliff, R., and McKoon, G. (1988). A retrieval theory of priming in memory. *Psychological review*, 95:385.
- Sahel, S., Nottbusch, G., Grimm, A., and Weingarten, R. (2008). Written production of German compounds: effects of lexical frequency and semantic transparency. *Written Lang. Lit.* 11, 211–227. doi: 10.1075/wll.11.2.06sah
- Scaltritti, M., Pinet, S., Longcamp, M., and Alario, F. X. (2017). On the functional relationship between language and motor processing in typewriting: an EEG study. *Lang. Cogn. Neurosci.* 32, 1086–1101. doi: 10.1080/23273798.2017.1283427
- Seidenberg, M. S. (1985). The time course of phonological code activation in two writing systems. *Cognition* 19, 1–30. doi: 10.1016/0010-0277(85)90029-0
- Seidenberg, M. S., and McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychol. Rev.* 96:523. doi: 10.1037/0033-295X.96.4.523
- Seidenberg, M. S., Waters, G. S., Sanders, M., and Langer, P. (1984). Pre- and post-lexical loci of contextual effects on word recognition. *Mem. Cogn.* 12, 315–328. doi: 10.3758/BF03198291
- Sibley, D. E., Kello, C. T., and Seidenberg, M. S. (2009). Error, error everywhere: A look at megastudies of word reading. *Proceedings of the Annual Meeting of the Cognitive Science Society*. Amsterdam, The Netherlands.
- Sidhu, D. M., Heard, A., and Pexman, P. M. (2016). Is more always better for verbs? Semantic richness effects and verb meaning. *Front. Psychol.* 7:798. doi: 10.3389/fpsyg.2016.00798
- Sternberg, S. (1969). The discovery of processing stages: extensions of Donders' method. *Acta Psychol.* 30, 276–315. doi: 10.1016/0001-6918(69)90055-9
- Stolz, J. A., Besner, D., and Carr, T. H. (1995). Implications of measures of reliability for theories of priming: activity in semantic memory is inherently noisy and uncoordinated. *Vis. Cogn.* 12, 284–336. doi: 10.1037/0278-7393.21.3.596
- Taikh, A., Hargreaves, I. S., Yap, M. J., and Pexman, P. M. (2015). Semantic classification of pictures and words. *Q. J. Exp. Psychol.* 68, 1502–1518. doi: 10.1080/17470218.2014.975728
- Taikh, A., and Lupker, S. J. (2020). Do visible semantic primes preactivate lexical representations? *J. Exp. Psychol. Learn. Mem. Cogn.* 46:1533. doi: 10.1037/xlm0000825
- Tamminen, J., and Gaskell, M. G. (2013). Novel word integration in the mental lexicon: evidence from unmasked and masked semantic priming. *Q. J. Exp. Psychol.* 66, 1001–1025. doi: 10.1080/17470218.2012.724694
- Tan, L. C., and Yap, M. J. (2016). Are individual differences in masked repetition and semantic priming reliable? *Vis. Cogn.* 24, 182–200. doi: 10.1080/13506285.2016.1214201
- Taraban, R., and McClelland, J. L. (1987). Conspiracy effects in word pronunciation. *J. Mem. Lang.* 26, 608–631. doi: 10.1016/0749-596x(87)90105-7
- Thomas, M. A., Neely, J. H., and O'Connor, P. (2012). When word identification gets tough, retrospective semantic processing comes to the rescue. *J. Mem. Lang.* 66, 623–643. doi: 10.1016/j.jml.2012.02.002
- Thompson-Schill, S. L., Kurtz, K. J., and Gabrieli, J. D. (1998). Effects of semantic and associative relatedness on automatic priming. *J. Mem. Lang.* 38, 440–458. doi: 10.1006/jmla.1997.2559
- Tweedy, J. R., Lapinski, R. H., and Schvaneveldt, R. W. (1977). Semantic-context effects on word recognition: influence of varying the proportion of items presented in an appropriate context. *Mem. Cogn.* 5, 84–89. doi: 10.3758/BF03209197
- Whaley, C. P. (1978). Word–nonword classification time. *J. Verbal Learn. Verbal Behav.* 17, 143–154. doi: 10.1016/S0022-5371(78)90110-X
- Will, U., Nottbusch, G., and Weingarten, R. (2006). Linguistic units in word typing: effects of word presentation modes and typing delay. *Written Lang. Liter.* 9, 153–176. doi: 10.1075/wll.9.1.10wil
- Williams, J. N. (1996). Is automatic priming semantic? *Eur. J. Cogn. Psychol.* 8, 113–162. doi: 10.1080/095414496383121
- Yamaguchi, M. (2019). What 100 years of typing research can tell us. *Cogn. Psychol. Bull.* 1, 25–33. doi: 10.53841/bpsocg.2019.1.4.25
- Yamaguchi, M., Randle, J. M., Wilson, T. L., and Logan, G. D. (2017). Pushing typists back on the learning curve: memory chunking improves retrieval of prior typing episodes. *J. Exp. Psychol. Learn. Mem. Cogn.* 43, 1432–1447. doi: 10.1037/xlm0000385
- Yap, M. J., and Balota, D. A. (2015). Visual word recognition. In A. Pollatsek & R. Treiman (Eds.), *Oxford library of psychology. The Oxford handbook of reading*. Oxford University Press. 26–43.
- Yap, M. J., Hutchison, K. A., and Tan, L. C. (2017). Individual differences in semantic performance: Insights from the semantic priming project. in *Big data in cognitive science*, Routledge/Taylor & Francis Group, 203–226.
- Yap, M. J., Pexman, P. M., Wellsby, M., Hargreaves, I. S., and Huff, M. J. (2012). An abundance of riches: cross-task comparisons of semantic richness effects in visual word recognition. *Front. Hum. Neurosci.* 6:72. doi: 10.3389/fnhum.2012.00072
- Yap, M. J., Tan, S. E., Pexman, P. M., and Hargreaves, I. S. (2011). Is more always better? Effects of semantic richness on lexical decision, speeded pronunciation, and semantic classification. *Psychon. Bull. Rev.* 18, 742–750. doi: 10.3758/s13423-011-0092-y



OPEN ACCESS

EDITED BY

Thomas L. Spalding,
University of Alberta, Canada

REVIEWED BY

Meral Ceylan,
Anadolu University, Türkiye

*CORRESPONDENCE

Bo Yu
✉ yubo307@126.com

RECEIVED 02 February 2026

REVISED 21 April 2026

ACCEPTED 22 April 2026

PUBLISHED 11 May 2026

CITATION

Yu B (2026) Chinese as a second language anxiety and intervention in a technology-assisted environment: a mini-review. *Front. Psychol.* 17:1801765. doi: 10.3389/fpsyg.2026.1801765

COPYRIGHT

© 2026 Yu. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Chinese as a second language anxiety and intervention in a technology-assisted environment: a mini-review

Bo Yu*

International College of Chinese Language and Culture, Chongqing Normal University, Chongqing, China

This mini-review examines research on anxiety among learners of Chinese as a second language (CSL) in technology-assisted environments. It synthesizes findings from empirical studies published between 2015 and 2025, focusing on the unique manifestations of anxiety in digital settings, key influencing factors (e.g., technological features, individual differences, instructional design), and evidence-based intervention strategies. Results indicate that technology can both alleviate and exacerbate anxiety: tools such as VR and gamification show promise in reducing oral and communicative anxiety, whereas technical challenges and emotional isolation may heighten distress. Notably, the same platform (e.g., video-conferencing) can produce opposite effects depending on task design and learner proficiency. The review also highlights methodological limitations, including small sample sizes and short-term designs, and proposes a Technology-Affect Interaction Framework to guide future research. The review recommends longitudinal, mixed-methods approaches and learner-centered technological ecosystems that systematically support emotional as well as cognitive dimensions of CSL learning.

KEYWORDS

Chinese as a second language (CSL), intervention strategies, language learning anxiety, online learning, technology-assisted language learning

1 Introduction

1.1 Background: the rise of technology in CSL and affective considerations

In the 21st century, where globalization and digitalization are deeply integrated, technology-assisted language learning (Computer-Assisted Language Learning, CALL), especially mobile-assisted language learning (Mobile-Assisted Language Learning, MALL), has become an irreversible trend in the field of language education (Kukulska-Hulme and Viberg, 2018). The widespread use of mobile devices such as smartphones and tablets, as well as the empowerment of technologies like artificial intelligence and big data, have greatly expanded the temporal and spatial boundaries of language learning, providing learners with an unprecedented personalized and immersive learning experience. At the same time, the vigorous development of International Chinese Language Education has led to an increasingly large group of learners of Chinese as a second language (CSL), and the emotional factors in their learning process, especially language learning anxiety, have also received increasing attention from researchers.

Language learning anxiety refers to a distinctively strong anxiety emotion that learners experience due to the uniqueness of the language learning process. It is a key variable that affects the effectiveness of second language acquisition. For Chinese language learners, the unique writing system of Chinese characters, the complex tone and the profound cultural connotations may cause them to face greater cognitive load and psychological pressure compared to learners of other languages, leading to language anxiety and impacting their performance (Qian, 1999).

1.2 Conceptual clarification: defining “anxiety” in this review

In the language learning literature, “anxiety” is a multifaceted construct. For the purpose of this review, the framework established by Horwitz et al. is adopted and adapted, which defines foreign language classroom anxiety (FLCA) as a distinct complex of self-perceptions, beliefs, feelings, and behaviors related to classroom language learning arising from the uniqueness of the language learning process (Horwitz et al., 1986). The focus of this review is specifically on situated, state-level anxiety—the subjective, transient apprehension experienced by learners in technology-assisted CSL learning environments. While acknowledging the influence of more stable trait anxiety (a general personality predisposition to anxiety), the analysis centers on how technological and pedagogical factors evoke or modulate situation-specific anxiety states.

Anxiety is further differentiated based on the language skill domain. The manifestations are examined distinctly in relation to listening, speaking, reading, and writing in digital contexts. A primary aim of this synthesis is to uncover whether and how technology impacts anxiety differently across these skill modalities.

Attention is also given to the theoretical distinction between debilitating anxiety (which impairs performance) and facilitating anxiety (which may motivate heightened effort). The literature reviewed herein primarily reports on anxiety as a negative, performance-interfering emotion. Consequently, this review predominantly addresses anxiety in its debilitating form, while noting that certain technological designs might theoretically transform anxiety into a more facilitative experience—a potential avenue for future inquiry. By adopting these conceptual clarifications, consistency and precision are ensured in interpreting and synthesizing the findings across the diverse studies included in this mini-review.

1.3 Aims and scope of the present mini-review

This study aims to conduct a mini-review of the research on Chinese second language anxiety in a technology-assisted environment. The reason for adopting this mini-review format is that although there have been many empirical explorations in this field, the research conclusions are often scattered or even contradictory, and there is a lack of systematic integration and critical analysis. A comprehensive systematic review is undoubtedly

ideal, but given that the current research volume is still in the accumulation stage and the quality varies, a mini-review that focuses on core issues, aims to outline the key context and highlight contradictions, can provide timely and concentrated insights for clarifying the current research status, identifying methodological flaws, and guiding future directions.

This mini-review will be conducted around the following three core questions:

1. In a technology-assisted environment, what are the unique manifestations of Chinese second language learners' anxiety? What are the main influencing factors (such as technological characteristics, teaching design, individual differences)? The technology environment not only may change the form of traditional classroom anxiety, but its interaction with the learning difficulties of Chinese itself is more likely to generate unique anxiety experiences, which urgently need to be deeply analyzed.
2. What technology intervention strategies have been used in existing research to alleviate Chinese second language anxiety? What are the empirical effects of these strategies, and what is their mechanism? The effectiveness evidence of current intervention measures (such as using VR/AR to create a safe context, based on AI for immediate feedback, gamification incentives, etc.) is uneven, and their targeted effects on different types of anxiety (such as oral anxiety, writing anxiety) are not clear, and require systematic evaluation.
3. What are the main methodological limitations and theoretical gaps in this field of research? What paths should future research prioritize? A large number of studies have problems such as small sample size, short cycles, and unrigorous experimental design (Burston, 2015), and most have not deeply revealed the mediating mechanism of technology on anxiety. Identifying these limitations and gaps is crucial for enhancing the scientific nature and practical guidance value of future research.

It is worth noting that the interaction between the technological environment and second language anxiety in Chinese presents a highly complex and contradictory nature, which constitutes the core contradiction that this review needs to address. On the one hand, technology is regarded as a potential remedy for anxiety. Technologies such as virtual reality (VR) and augmented reality (AR) can create safe simulated communication scenarios; adaptive learning systems can provide customized content, reducing feelings of frustration. However, on the other hand, technology itself may also become a new source of anxiety. Network connections, software operations, the sense of coldness in human-computer interaction, and the social gaze pressure from online public expression can all trigger or exacerbate anxiety (Stockwell, 2013). This dual-edged nature of technology has been highlighted in recent empirical studies. Research indicates that, compared to traditional face-to-face teaching, the use of CALL and MALL in instruction can lead to a significant increase in learners' anxiety levels (Dong et al., 2022). This finding contrasts with many studies reporting positive effects, profoundly revealing the complexity of technology's involvement in the emotional domain. It warns us to abandon the simplistic assumption that technology will inevitably reduce anxiety and instead calls for more careful and detailed

examinations. Moreover, how individual differences of learners (such as cultural background, motivation type, and self-efficacy) interact with the technological environment and jointly influence the anxiety experience is also an indispensable dimension for understanding this phenomenon.

This mini-review will sort out the existing evidence, integrate and analyze the manifestations, causes, interventions, and research quality of Chinese second language anxiety in a technology-assisted environment, aiming to clarify the current contradictions discovered and provide a clear road map for future construction of more rigorous and effective emotion-enhanced technology-based Chinese language learning environments.

2 Methodology

In order to ensure that this review can fully reflect the research progress of Chinese second language anxiety and intervention in the technology assisted environment, this review follows clear literature selection criteria.

2.1 Literature search strategy and selection criteria

To ensure a transparent and reproducible review process, a systematic literature search was conducted. The search was performed across major academic databases. Chinese-language publications were primarily retrieved from the China National Knowledge Infrastructure (CNKI). English-language publications were sourced from databases including Web of Science Core Collection, ERIC, and PsycINFO.

The search strategy employed a combination of keywords related to three core concepts: (1) the target language (e.g., “Chinese as a second language,” “CSL,” “汉语二语”), (2) the psychological construct (e.g., “anxiety,” “焦虑”), and (3) the learning context (e.g., “technology-assisted,” “online,” “CALL,” “MALL,” “计算机辅助,” “线上”). Boolean operators (AND, OR) were used to refine the search. An illustrative search string for English databases was: (“Chinese as a second language” OR “CSL”) AND (anxiety) AND (“online” OR “technology-assisted” OR “CALL” OR “MALL”). The publication timeframe was set from 2015 to 2025 to capture the evolution of research concurrent with the rapid adoption of educational technology. All the retrieved records were imported into the literature management software, and the duplicate documents were removed.

2.2 Study screening, selection, and final inclusion

The screening and selection process followed a structured multi-stage protocol, detailed in the PRISMA flow diagram (see Figure 1).

Initial Screening: All retrieved records were compiled, and duplicates were removed. The titles and abstracts of the remaining records were screened against broad relevance criteria.

Records were excluded if they were clearly unrelated (e.g., not focusing on CSL learners, or not involving a technology-assisted learning environment).

Full-Text Screening and Eligibility Assessment: The full texts of the remaining records were obtained and rigorously assessed against the following pre-defined eligibility criteria:

Study Type: Empirical research (quantitative, qualitative, or mixed-methods). Theoretical or review articles were excluded but saved for background reading.

Participants: Learners of Chinese as a second or foreign language.

Context: The learning environment must be technology-assisted (e.g., online platforms, mobile applications, virtual reality).

Focus: The study’s core variables must include CSL learning anxiety and/or an intervention aimed at it.

Language and Date: Publications in Chinese or English, published between 2015 and 2025.

This process was carried out independently by two researchers, and the screening results were cross checked to ensure the reliability of the decision.

Final Inclusion and Data Extraction: Studies meeting all eligibility criteria were included. For each included study, key data (e.g., authors, technology environment, main findings related to anxiety) were extracted and synthesized. This process resulted in the final corpus of 25 publications for this mini-review.

A Note on Source Types and Rationale: The final corpus comprises 20 empirical studies and 5 review articles. Notably, among the 20 empirical studies, 10 are peer-reviewed journal articles and 10 are Master’s theses. This composition aligns with the purpose of a mini-review, which is to provide a timely, focused synthesis and mapping of a developing field rather than an exhaustive systematic review. Including Master’s theses allows for a more comprehensive capture of early exploratory work and practitioner-led inquiries. In the synthesis and discussion, we maintain a critical perspective: primary conclusions are grounded in findings from peer-reviewed journal articles, while insights from theses are treated as valuable preliminary evidence and illustrative cases. This treatment also enables us to consider the potential quality differences caused by differences in research methods (such as peer-reviewed research and Dissertations) in the comprehensive analysis.

2.3 Overview of review literature

Research on Chinese second language learning anxiety in the past decade has shown the characteristics of multi-dimensional perspective and technology integration. The study found that motivation is negatively correlated with anxiety, while learning strategies positively predict anxiety, and technology tools had complex regulatory effects on anxiety, motivation, strategies and learning outcomes (Lan, 2025). Compared with Chinese and foreign studies, scholars in mainland China mostly use quantitative methods, while international scholars prefer

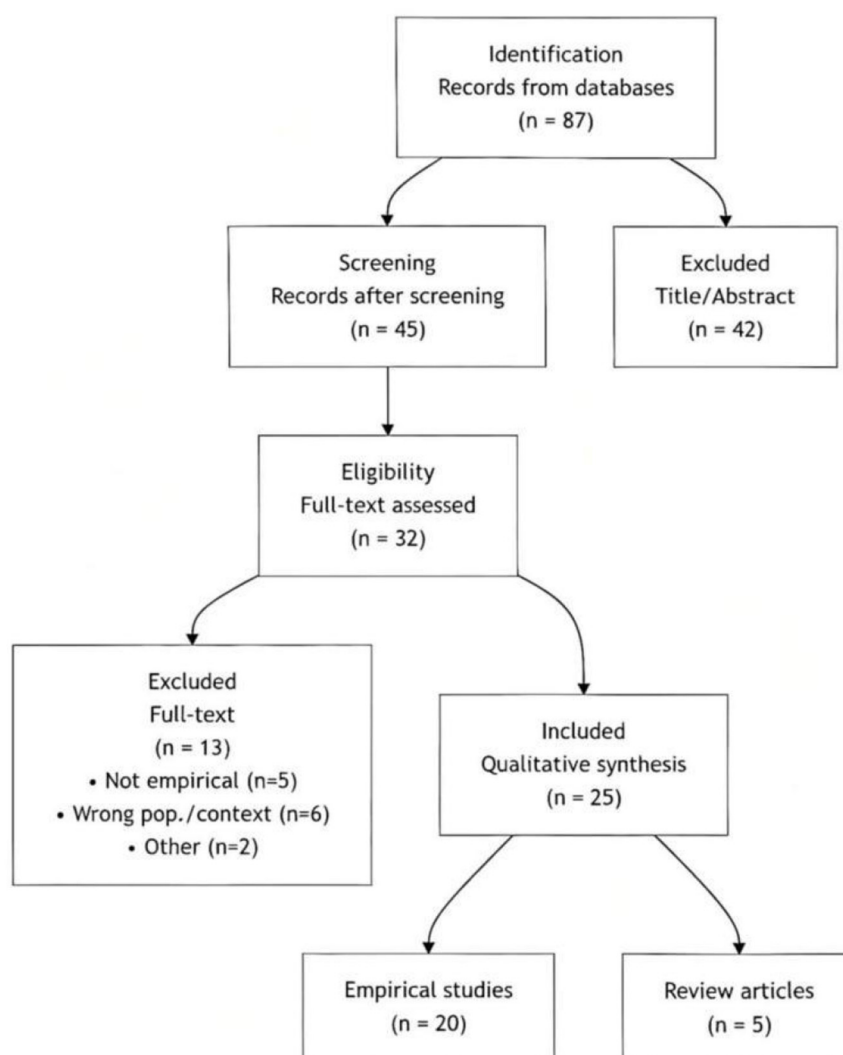


FIGURE 1
PRISMA flow diagram of the literature search and selection process.

mixed research methods; Gender, language level and cultural background were the significant influencing factors of anxiety (Yao et al., 2022). In addition, the meta-analysis results show that anxiety is negatively correlated with Chinese learning achievement in a small to medium degree, among which beginners are most affected by anxiety, and the proportion of women significantly regulates the relationship between the two (Qiao et al., 2025).

In terms of research on technology assisted Chinese teaching, the research found that mobile applications, Web 2.0 technology, network conference technology and so on are widely used to support Chinese teaching, and mobile applications have become the mainstream research tools in the past decade (Lyu and Qi, 2020). The research also points out that technology application has advantages in improving learning motivation and providing real context, but it also faces challenges such as teachers' and students' lack of technical ability, students' low attendance and retention

rate, and difficulties in connecting learning inside and outside the classroom.

To sum up, the existing research has constructed the correlation framework of anxiety, motivation, strategy and technology, but it is necessary to strengthen cross-cultural comparison, long-term tracking and the application of mixed methods to deepen the empirical research of technology intervention anxiety.

2.4 Overview of empirical research literature

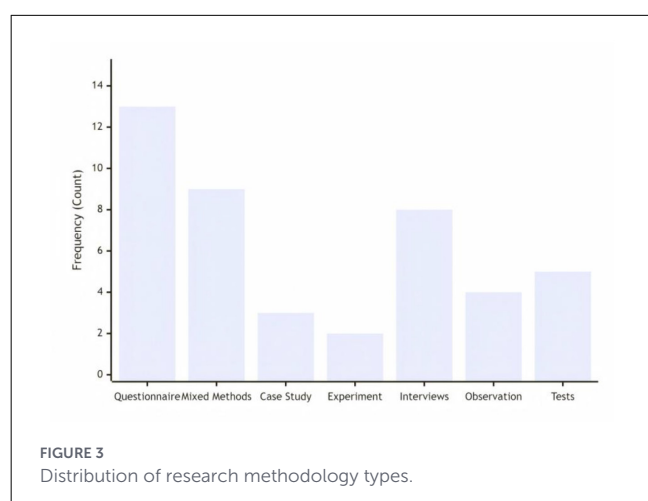
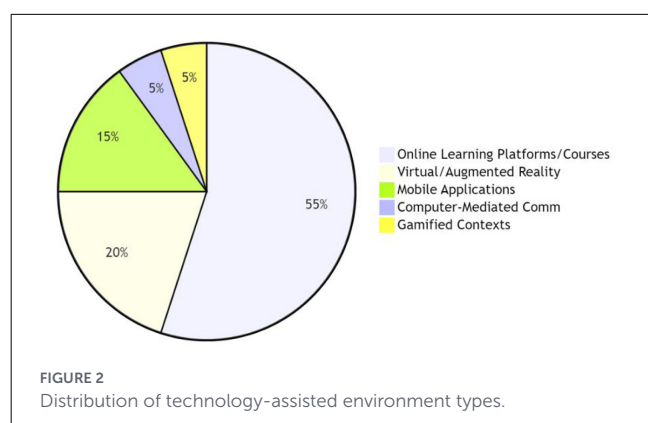
Twenty empirical studies were identified, including 10 Chinese publications and 10 English publications. These studies mainly explore the influence of variables such as anxiety, motivation, learning strategies and Chinese learning outcomes (such as language skills, academic adaptation, etc.), as well as the role of

TABLE 1 Included empirical studies: Chinese-language publications.

Author	Title	Technology-assisted environment	Main findings (related to anxiety)
Watanabe (2020)	A study on Chinese learning anxiety of Japanese beginner-level learners in online classes	Online Chinese class	Anxiety levels were generally moderately high. The dimension of “Anxiety in Teacher-Student Classroom Communication” scored highest. Learning background (e.g., ethnicity, proficiency) was a primary influencing factor.
Guo (2021)	A Study on the anxiety in online spoken Chinese classes of international students at the primary stage	Online spoken Chinese class	Anxiety was prevalent among beginner learners. Individual factors (e.g., gender, proficiency) influenced anxiety levels. Students often employed positive coping strategies.
Chen C. (2021)	A study on the anxiety of junior and intermediate international students in Chinese online classes based on meta-cognitive strategies	Online Chinese class	Anxiety levels were moderate. Usage of meta-cognitive strategies showed a negative correlation with anxiety levels. Student individual factors and teacher instructional methods jointly influenced anxiety.
Xie and Luo (2021)	An analysis of the causes and adjustment strategies of anxiety in online Chinese language learning among international students	Online Chinese learning	Anxiety levels were higher compared to previous studies. Anxiety showed a negative correlation with HSK level and learning duration. High anxiety primarily stemmed from the course and exams themselves, rather than communication apprehension.
Sun W. R. (2022)	A study on the anxiety of online Chinese listening and speaking classes for high school students in Thailand	Online Chinese listening and speaking class	Moderate-to-high anxiety was prevalent. Anxiety was highest during the information processing stage. Listening and speaking materials were the most anxiety-inducing external factor.
Sun J. (2022)	A survey of Chinese learners’ learning anxiety under network teaching environment	Network-based teaching environment	Anxiety levels were moderate. Affective deficiency anxiety (loneliness, lack of interaction) was most prominent. Age, HSK level, feelings of loneliness, and network factors significantly influenced anxiety.
Hong (2022)	A study on online writing anxiety and self-adjustment strategies of intermediate Chinese L2 learners	Online writing	Writing anxiety was at a moderate level. Anxiety was positively correlated with the frequency of self-regulated strategy use (learners with higher anxiety used strategies more frequently). Age and gender influenced specific anxiety dimensions.
Liu (2023)	A study on online Chinese language learning anxiety and coping ways of Pakistani students	Online Chinese learning	Anxiety levels were moderately high. Anxiety showed a negative correlation with the tendency to adopt positive coping strategies. Gender and self-rated language proficiency significantly affected anxiety.
Chen (2023)	A survey on online Chinese learning anxiety of primary level learners	Online Chinese learning	Anxiety levels were moderately high. Anxiety mainly stemmed from the external environment (e.g., limited teacher-student interaction, untimely feedback, network issues). Gender had an effect, but learning duration showed no significant impact.
Zhang et al. (2024)	The application of gamified virtual contexts in international Chinese language education—a utility perspective of discourse cognition and emotional motivation	Gamified virtual context	Gamified virtual contexts can significantly reduce language use anxiety, enhance willingness to communicate, and improve cognitive comprehension abilities.

TABLE 2 Included empirical studies: English-language publications.

Author	Title	Technology-assisted environment	Main findings (related to anxiety)
Wang et al. (2017)	Chinese language learners' anxiety toward chat partners in computer-mediated communication	Text-based Computer-Mediated Communication (CMC)	Chinese learners experienced significantly higher levels of Foreign Language Anxiety (FLA) when chatting with Native Speakers (NSs) compared to Non-Native Speakers (NNSs). The main mediating factors were language confidence and partner familiarity.
Xie et al. (2019b)	Using interactive virtual reality tools in an advanced Chinese language class: a case study	Virtual reality tools (Google cardboard/expeditions)	Using VR tools for contextualized presentations helped ease nervousness and anxiety during oral presentations.
Xie et al. (2019a)	Effects of using mobile-based virtual reality on Chinese L2 students' oral proficiency	Mobile-based Virtual Reality (VR: Google cardboard and expeditions)	Learners using VR tools for role-play tasks scored significantly higher on the content and vocabulary of their oral presentations than those using PowerPoint. VR reduced presentation anxiety by providing an immersive preparation environment and diverting audience attention.
Chen C. (2021)	The effects of online learning on alleviating students' Chinese language learning anxiety: a study in a Chinese University	Online learning platform	Online learning significantly reduced anxiety in listening, speaking, and writing, but increased anxiety in reading. Reading anxiety became the strongest predictor of general FLA in the online environment. Multiple regression indicated that reading anxiety explained more variance in general FLA for the online group than speaking anxiety did for the traditional group.
Xu et al. (2022)	The roles of motivation, anxiety and learning strategies in online Chinese learning among Thai learners of Chinese as a foreign language	Online learning	Anxiety emerged as the most stable negative predictor of self-rated CFL achievement in online learning.
Sandyaningrum and Kusumaningtyas, 2023	The effects of learning Mandarin online with a native lecturer to University students' anxiety	Online learning	Online learning with native lecturers caused student anxiety, manifesting as reduced learning conditions, self-efficacy, and learning effectiveness.
Srisinthon (2024)	Anxiety and engagement in the online classroom: a case study of Thai learners of Chinese as a foreign language	Online learning	Thai CFL learners' online learning anxiety was at a moderate level, primarily caused by fear of communication. Chinese proficiency affected anxiety, with low-proficiency learners having higher anxiety. Online learning anxiety was significantly correlated with learner engagement in the dimension of social presence.
Zhao and Wang (2024)	Domain-specific L2 grit, anxiety, boredom, and enjoyment in online Chinese learning	Online learning	In online Chinese learning, the Perseverance of Effort (POE) facet of L2 Grit significantly positively predicted learning enjoyment and negatively predicted anxiety. Both anxiety and enjoyment significantly predicted boredom.
Huang et al. (2024)	"How Anxious I am": the effect of different online modalities on Chinese language beginners' classroom anxiety	Video conferencing (zoom) and virtual world (second life)	The effect of different online modalities on anxiety was complex, influenced by factors like gender, task-related language skills, and learners' immediate experience (e.g., engagement, task awareness). Reading anxiety decreased significantly in Zoom, while anxiety in speaking/listening decreased significantly in second life.
Puri et al. (2025)	The impact of mobile-based language learning on speaking and learning anxiety, engagement and achievement in Chinese language learning: the mediating role of cognitive load	Mobile-Based Language Learning (MBLL)	MBLL was significantly associated with lower speaking and learning anxiety. Cognitive load mediated this relationship, meaning MBLL reduces anxiety by lessening cognitive load.



technological aids in it. [Tables 1, 2](#) summarize the main findings of relevant empirical studies.

First of all, from the perspective of technology assisted environment (see [Figure 2](#)), online learning platforms (such as Zoom and various online course platforms) are mainstream research scenarios, which reflects the current trend of normalization of online education. At the same time, virtual reality/augmented reality (VR/AR) and mobile apps, as the intervention means of emerging technologies, have also attracted more attention, reflecting the emerging dynamics of technology development. However, the research on computer mediated communication (CMC) and gamified context is relatively few, and there is still room for further exploration.

Quantitative research using questionnaires is dominant, as anxiety studies typically require large-scale data support. Mixed research methods (combined with questionnaires, interviews, observations, etc.) also account for a considerable proportion, which shows researchers' emphasis on in-depth exploration. Case studies and experimental studies provide strong evidence for in-depth understanding of individual differences and testing the effect of technical intervention (see [Figure 3](#)).

Existing empirical studies have shown that there are complex and diverse relationships between anxiety and Chinese second language learning. At the same time, as an important means of teaching intervention, technical aids have a significant impact on

anxiety and related variables in many ways, and its effect mainly depends on the type of tools, the way of use and the learning situation. These findings provide an important empirical basis for further understanding the internal mechanism of Chinese second language acquisition and the design of effective teaching strategies.

3 Performance and influencing factors of Chinese second language anxiety in technology assisted environment

3.1 Multidimensional performance and skill differences of anxiety

The technology-assisted environment has reshaped the anxiety situation of Chinese second language learners. Research shows that online learners generally have moderately high levels of anxiety, and their manifestations exhibit significant multi-dimensional characteristics. On one hand, traditional classroom anxieties, such as communication anxiety and exam anxiety, still exist in the online environment and may even be amplified; on the other hand, the technological environment has given rise to new dimensions of anxiety, among which “emotional deprivation anxiety” becomes prominent due to physical isolation and insufficient social presence.

The technical operation anxiety in online Chinese language learning is increasingly prominent, involving multiple challenges such as network stability, platform operation, and adaptation to digital texts ([Xie and Luo, 2021](#)). The anxiety responses to language skills in the technical environment show significant differences. Listening comprehension often becomes an important source of anxiety for students due to the instantaneous and non-repetitive nature of information input in the online environment. Studies have shown that beginner learners frequently report that listening materials are too fast and content is difficult to understand, both of which directly contribute to listening anxiety. This issue is particularly pronounced in tasks with higher listening demands ([Guo, 2021](#)). In terms of oral expression, the “being watched” sensation during synchronous video interaction significantly increases students' psychological pressure. When learners express themselves in real time in front of the camera, they are prone to a strong sense of being evaluated, which further aggravates oral anxiety ([Sun W. R., 2022](#)). Writing learning also faces challenges in technical adaptation, such as difficulty in inputting Chinese characters and concerns about file saving, all of which may trigger writing anxiety ([Hong, 2022](#)). For intermediate-level learners, they not only need to adapt to the online writing environment but also overcome technical obstacles such as input method usage and format adjustment, which all increase the possibility of anxiety. Interestingly, reading anxiety may take on different forms in the online environment. Some research indicates that although online learning has a certain alleviating effect on anxiety in listening, speaking, and writing, differences in reading habits of digital texts, visual adaptation to screen reading, and other issues may also trigger or exacerbate reading anxiety ([Chen C., 2021](#)).

The specific manifestations of technical operation anxiety in various language skills have become an important psychological factor affecting online Chinese language learning. Future

research can further focus on the construction of corresponding intervention strategies to enhance learners' learning experience and effectiveness in the technology-integrated environment.

The literature reveals a significant tension regarding the manifestation of communication anxiety, a core dimension of FLCAS, in online environments. On one hand, studies (e.g., [Watanabe, 2020](#); [Srisinthon, 2024](#)) indicate that "Anxiety in Teacher-Student Classroom Communication" or fear of communication remains highly prevalent online, often cited as a primary or significant source of distress. This suggests a direct transference of traditional classroom anxieties to the digital space. On the other hand, other research ([Chen D., 2021](#); [Wang et al., 2017](#)) presents a more nuanced picture, where the nature of the online partner (native vs. non-native speaker) or the specific task design (synchronous performance vs. asynchronous practice) critically mediates this anxiety. This contradiction points to a critical, underexamined variable: the specific affordances and constraints of the technological platform itself may reconfigure, rather than merely replicate, social evaluation dynamics. For instance, the "being watched" sensation on a synchronous video platform like Zoom may amplify the performative aspect of communication anxiety, while text-based CMC with delayed response could potentially mitigate the immediacy of social threat. Future syntheses must move beyond simply confirming the presence of communication anxiety online to dissecting how different technological mediation architectures reshape its intensity and qualitative experience.

The above discussion reveals the complexity of the effect of technology on anxiety. In order to understand these contradictory results more systematically, we need to examine several key regulatory variables. One is the technical mode. Synchronous video conferencing may aggravate performance anxiety, while asynchronous text communication or virtual environment may reduce social threats. Second is task design. Impromptu oral tasks are more likely to induce anxiety than prepared speeches, while game tasks can relieve pressure by reducing the cost of trial and error. Third is learners' level. Beginners' dual strangeness to technical operation and language itself may magnify anxiety, while middle and advanced learners are more affected by fear of evaluation. Finally, research design should be considered. Cross-sectional surveys can only capture static anxiety levels, while longitudinal or experimental designs can reveal the dynamic effects of technical intervention. Future research should clarify these regulatory factors in the design and explore the interaction between them, so as to more accurately explain the contradictory effect of technology.

3.2 Differentiated impact of individual factors

There is a deep intertwined relationship between learners' individual characteristics and anxiety experience. As a basic influencing factor, language proficiency has a significant effect on learners' anxiety. Primary learners often feel high anxiety because they are unable to understand and express effectively due to their immature language ability ([Srisinthon, 2024](#)). Intermediate

learners, such as those who have reached HSK level 3–4, may encounter learning bottlenecks and face specific psychological difficulties, which will also cause anxiety ([Sun J., 2022](#)).

Self-cognition plays a key role in anxiety experience. Learners' underestimation of their language ability and uncertainty about the effectiveness of online learning will significantly aggravate their anxiety. This negative evaluation of self-ability often leads to learners' lack of self-confidence in the process of learning, and then increases anxiety. The influence of gender differences on anxiety experience in specific situations cannot be ignored. Female learners may be under greater pressure in terms of evaluation sensitivity and technology adaptability, which may result from excessive attention to others' evaluation and inadaptability to new technologies. For example, Pakistani female Chinese learners tend to show a higher level of anxiety in the face of technical operation and evaluation ([Liu, 2023](#)).

In addition, individual factors such as personality traits, motivation type, cultural background, and prior experience in a target-language country jointly shape differentiated anxiety responses. Introverted learners may be more likely to fall into an anxiety state because they are less likely to actively seek help and express themselves ([Watanabe, 2020](#)). However, learners with instrumental motivation, that is, those who learn Chinese mainly for practical purposes, may have higher anxiety levels due to excessive attention to learning outcomes ([Xu et al., 2022](#)).

In the future, when designing and implementing online Chinese teaching, we should fully consider these individual differences and adopt targeted teaching strategies to effectively alleviate learners' anxiety and improve their learning effect.

While the reviewed studies consistently identify individual factors (e.g., proficiency, gender, and motivation) as significant correlates of anxiety, the prevailing analytical mode tends to treat these as static, pre-existing learner attributes that predispose individuals to higher or lower anxiety in the technological environment. This perspective, while valuable, risks technological determinism by overlooking how the technological environment actively co-constructs these "individual" experiences. For example, the finding that female learners or beginners report higher anxiety should not be taken as a simple reflection of inherent traits. Instead, a critical analysis must ask: Does the design of the technology or the pedagogical approach enacted through it inadvertently create or exacerbate challenges that disproportionately affect these groups? A beginner's anxiety might be less about low proficiency itself more about the platform's lack of scaffolding or adjustable difficulty levels. Similarly, gender differences in anxiety could be linked to how online public performance or technical troubleshooting is socially framed within the learner's culture or the class dynamic. Therefore, future research should shift from merely cataloging which individual factors matter to investigating the interaction between learner characteristics and specific technological-pedagogical designs. This requires a design that can capture dynamic interactions, such as qualitative interviews, diary research and other methods, to deeply understand how different learners perceive and respond to specific technical characteristics, and how this perception changes over time. This would reveal whether technology is merely a neutral backdrop against which fixed differences play out, or an active

agent that differentially modulates the anxiety landscape for diverse learners.

3.3 Interaction between teaching and technology environment

The interaction between teaching environment and technology design has a profound impact on the generation and evolution of learners' anxiety. The quality of interaction mode is a key variable affecting the level of anxiety. Although online communication with native speakers offers valuable language practice opportunities, it often leads to higher anxiety due to language proficiency gaps and social unfamiliarity (Wang et al., 2017). On the contrary, the lack or inefficiency of interaction between teachers and students and students directly leads to the lack of emotional support and the increase of learning frustration (Chen, 2023). This lack of emotional support further aggravates learners' anxiety and affects their learning enthusiasm and effect. Task design also plays an important role in the interaction between teaching and technology environment. Inappropriate task difficulty, insufficient preparation time and forced immediate performance may trigger learners' processing and output anxiety.

For example, in the online Chinese listening and speaking class, if the task design is too complex or the time arrangement is too tight, students often feel anxious because they cannot fully prepare (Sun W. R., 2022). Therefore, designing task difficulty reasonably, providing sufficient preparation time and avoiding forced immediate performance are effective ways to alleviate learners' anxiety.

The technological tool itself has duality, and whether its design is reasonable or not directly affects the anxiety level of learners. Improper technology design will become a source of anxiety, increase the cognitive burden and operation difficulty of learners, and cause anxiety. However, making good use of technical tools can become a pressure reducer. By reducing cognitive load, mobile learning provides learners with a more flexible and convenient way of learning, which indirectly alleviates the anxiety of learners, thus indirectly alleviating learning anxiety. In addition, virtual reality technology disperses learners' excessive attention to self-expression and reduces their anxiety through situational immersion (Xie et al., 2019a).

In order to effectively alleviate the anxiety of learners and improve their learning effect, teachers should pay attention to optimizing the interaction mode, reasonably designing the task difficulty, and making full use of the advantages of technical tools to create a positive and supportive learning environment for learners.

3.4 Complex interaction between emotion and cognition

Anxiety is not an isolated emotional phenomenon, but closely related to the cognitive process, which jointly affects learners' online learning experience. From the perspective of positive psychology, learners' psychological resilience, especially their efforts to adhere to this quality, can effectively buffer the negative effects of anxiety and promote the generation of positive emotional experience

(Zhao and Wang, 2024). Specifically, when facing the challenges in online learning, learners with high psychological resilience are more able to maintain a positive attitude and turn anxiety into a driving force for progress, so as to improve the learning effect. Cognitive load theory further reveals the influence mechanism of technological environment on learners' anxiety level. In the technological environment, information overload is a common problem, which directly causes the anxiety reaction of learners. When the amount of information learners need to process exceeds the carrying capacity of their cognitive resources, cognitive overload will occur, which will lead to the rise of anxiety. Therefore, a good teaching design should focus on optimizing cognitive resources, help learners effectively manage cognitive load and reduce anxiety level by reasonably designing learning tasks and providing clear learning guidance. The development of meta-cognitive ability is particularly critical in anxiety management. Meta-cognition refers to learners' ability to recognize and regulate their cognitive process, including the use of strategies such as planning, monitoring and regulation. Research shows that learners who are good at using planning, monitoring, and adjustment strategies can show stronger adaptability and significantly lower anxiety level when facing the uncertainty of online learning (Chen D., 2021). These learners can make reasonable learning plans according to their own learning objectives and actual situation; In the process of learning, they can monitor their learning status in time, find problems, and adjust them in time; When facing difficulties and challenges, be able to actively seek solutions and maintain a positive learning attitude.

These findings all point to an important understanding: anxiety management in the technological environment needs to go beyond mere emotional comfort, but should go deep into the cognitive level and build a supportive learning ecology. This means that educators should not only pay attention to learners' emotional needs and provide emotional support; It is more necessary to help learners effectively manage cognitive load and enhance psychological resilience by optimizing teaching design and improving learners' meta-cognitive ability, so as to achieve a more efficient and pleasant learning experience in the technical environment. These cognitive and affective mechanisms not only explain the nature of anxiety in digital environments but also point toward potential leverage points for technological intervention, which we turn to next.

4 Strategies and empirical effects of technology intervention on Chinese second language anxiety

4.1 The interventional role of technology-enhanced environments in reducing anxiety

Immersive virtual environment provides an innovative path to alleviate anxiety through psychological transfer and situational embedding. The application of virtual reality technology shows that when learners act as virtual guides in VR scenes, their oral expression anxiety is significantly reduced. Its core mechanism is

that the 360 degree immersion environment diverts the audience's attention from the speaker, providing a valuable psychological buffer (Xie et al., 2019b). This defocusing effect makes learners pay more attention to the task itself rather than self-expression.

Gamification design further expands the dimension of technical intervention. The study found that the anxiety level of primary Chinese learners who learned in the game virtual situation was significantly reduced after the intervention, and their willingness to communicate was improved (Zhang et al., 2024). Its utility mechanism is multiple. Game elements (such as progress visualization and achievement system) transform learning into a more attractive and secure exploration process by providing immediate positive feedback and controllable challenges; Virtual situation creates a low-risk language practice field, allowing learners to make bold attempts in an environment with low cost of making mistakes.

The design of interactive mode needs to be based on the refined consideration of technical characteristics. Comparative studies have found that different online environments have different effects on anxiety: learners who complete communication tasks in the three-dimensional virtual environment show lower overall anxiety than learners in the traditional video conference environment, especially in speaking and listening (Huang et al., 2024). This is because the avatars and situations provided by the virtual environment reduce the pressure to evaluate the real face and allow more flexible collaboration.

Even in the conventional platform, promoting structured social interaction through technical means can effectively alleviate the anxiety of emotional loss. These practices show that technology is not only a channel for information transmission, but also a catalyst for high-quality social interaction.

4.2 Technology-supported pathways for cognitive and meta-cognitive strategies

The intervention effect of mobile learning technology has been strongly supported by cognitive science theory. Well-designed mobile language learning applications can significantly reduce learners' oral anxiety and general learning anxiety, with cognitive load playing a key mediating role (Puri et al., 2025). This discovery has important teaching implications. The effectiveness of technological intervention depends not only on the richness of functions, but also on whether it conforms to the laws of human cognitive processing. When mobile applications can optimize the information processing process through simple navigation, timely support, personalized content push, etc., they can effectively reduce the external cognitive load, thus reducing the frustration and anxiety caused by it.

The combination of technical tools and strategy training provides an endogenous solution for anxiety management. Chen D. (2021) finds that in online Chinese learning, the frequency of meta-cognitive strategy use is significantly negatively correlated with the level of anxiety. Therefore, teachers can systematically integrate strategy training into the technical environment: guide students to use digital calendar and task management tools for learning planning (planning strategies); The learning process

monitoring and self-assessment (monitoring strategy) are carried out by using the recording, broadcasting and playback function; Real time problem solving and path adjustment (adjustment strategy) using online resource library and help system. This training mode of technology empowerment strategy can enhance learners' sense of control over the online environment and fundamentally improve their adaptability.

4.3 Building a systematic technology-supported ecosystem

Effective technical intervention needs multi-level systematic support. For specific skill anxiety, targeted technical solutions should be adopted: to alleviate listening anxiety, multimodal audio and video resources with adjustable speech speed and optional subtitles can be provided; In order to reduce writing anxiety, online model article library, collaborative editing platform and intelligent feedback system can be built (Hong, 2022). In view of the general sense of technical unease, it is necessary to establish a perfect technical support system, including clear operation guidelines, timely technical assistance channels, and the recording and broadcasting backup of key teaching contents (Sandyaningrum and Kusumaningtyas, 2023). Furthermore, an ideal technology ecology should have integration and adaptability. It needs to organically combine immersive environment, mobile applications, interactive platform, and strategy training tools to form complementary technical solutions. At the same time, the ecosystem should be able to provide differentiated support paths according to learners' anxiety types, language levels, etc., such as automatically recommending learning resources or strategy guidance that match their current state according to learners' behavior data on the platform.

It is worth noting that the term "technology" in this review covers many levels, from hardware devices (such as VR head mounted display) to software platforms (such as zoom) and then to specific functions (such as asynchronous discussion and intelligent feedback). The effect is not single, but closely related to the specific implementation of these levels. The empirical studies reviewed, while demonstrating the potential of various technological interventions, collectively point to a crucial gap between technical feasibility and pedagogical sustainability in anxiety alleviation. A critical synthesis suggests that the effectiveness of any tool—be it VR for situational exposure, mobile apps for spaced repetition, or AI for personalized feedback—is not an inherent property of the technology itself. Instead, it is contingent upon its embeddedness within a coherent pedagogical design that addresses the core anxiety sources identified in Section 3. For instance, a gamified VR environment might successfully lower the cognitive load and immediate social threat for speaking practice (addressing performance anxiety), yet fail to mitigate a learner's underlying "fear of communication" if the activity is not followed by supportive, formative feedback that builds self-efficacy. Similarly, asynchronous tools that reduce evaluation apprehension must be carefully balanced with opportunities for developing communicative competence to prevent the reinforcement of avoidance strategies. Therefore, future intervention research should

move beyond proving that a technology “can reduce anxiety” in a controlled setting. The priority should shift to investigating how these tools can be integrated into longitudinal, pedagogically-grounded learning sequences that not only provide temporary relief but also foster the psychological and strategic resources (e.g., enhanced self-regulation, positive attribution) needed for durable anxiety management. This necessitates study designs that evaluate not just pre-post anxiety scores, but also the mediating role of factors like growing familiarity with the tech environment, improved meta-cognitive awareness, and the quality of teacher/peer scaffolding within the digital space.

5 Conclusion

This mini review reveals the key findings in this field by combing the research on Chinese second language anxiety and its intervention in the technology assisted environment. The technological environment does not simply exacerbate or alleviate anxiety, but has a complex interaction with learners’ individual characteristics, instructional design and language skills.

The study found that anxiety in the technology assisted environment showed obvious duality. On the one hand, the forms of communication fear and test anxiety in the traditional classroom are amplified in the online environment; On the other hand, the technological environment has given birth to new anxiety dimensions, such as emotional loss anxiety caused by physical isolation and technological operation anxiety caused by digital adaptation. The response of different language skills to the technical environment is significantly different. The transience of listening comprehension, the feeling of being watched during oral expression, the technical adaptability of writing and the habit of reading digital text together shape the differentiated anxiety experience. This complexity shows that the impact of technology on anxiety cannot be generalized, but needs to be analyzed in combination with specific skill areas.

In terms of intervention effect, research has confirmed the potential of technical means in alleviating anxiety, but its effectiveness highly depends on the specific implementation methods. Virtual reality and game design provide an effective solution to oral anxiety by creating a low-risk practice environment and a distraction mechanism; Mobile learning technology indirectly reduces learning pressure by optimizing cognitive load; The refined design of online interaction can alleviate the sense of social isolation. It is worth noting that the technological empowerment of meta-cognitive strategies shows its unique value. It enhances the learners’ ability to cope with anxiety from the internal mechanism by enhancing the learners’ sense of control over the learning process. These findings support the view that the fit with cognitive and emotional processes is key to the success of technical intervention (Golonka et al., 2014).

However, the current research has exhibits notable methodological limitations. A large number of studies are limited to the experimental design of small samples and short periods, relying too much on self-reported questionnaires, and lacking in-depth capture of the dynamic process of anxiety. At the theoretical level, the mediating mechanism of technology

influencing anxiety has not been fully revealed, especially the path of how technology characteristics affect anxiety experience through variables such as cognitive load and self-efficacy needs to be further clarified. In addition, the existing studies focus on the negative effects of anxiety, and pay less attention to the complex relationship between anxiety and motivation. Some types of anxiety may promote learning, and the performance of this dimension in the technical environment is worth further discussion.

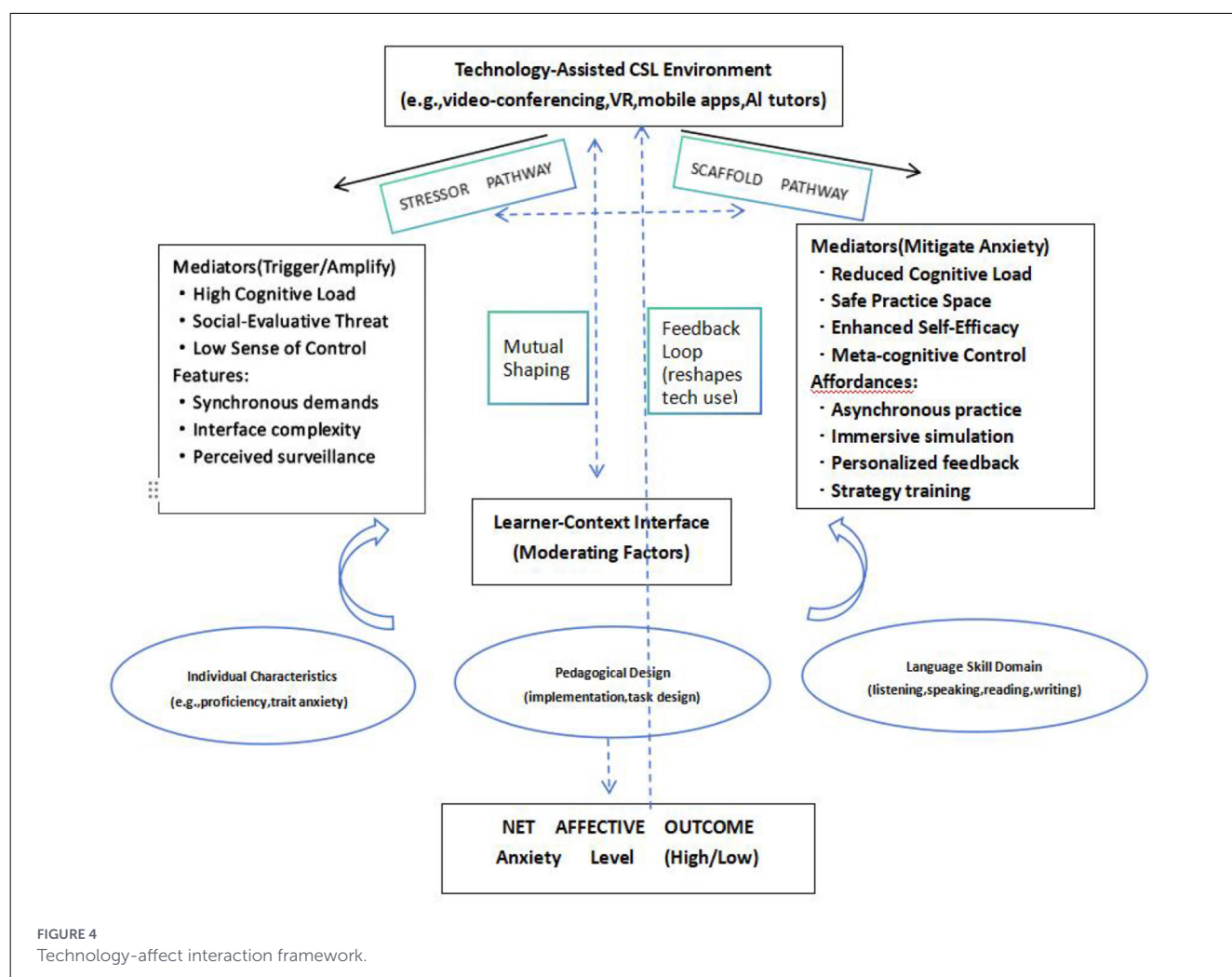
In general, the research on Chinese second language anxiety in the technology assisted environment is at a critical stage from phenomenon description to mechanism exploration. Future research needs to achieve breakthroughs in the preciseness of methods, theoretical depth and practical pertinence, so as to provide solid support for the construction of an efficient and humanized technology enhanced Chinese learning environment.

To integrate these findings and guide future research beyond descriptive phenomenology, we propose a Technology-Affect Interaction Framework for conceptualizing anxiety in technology-assisted CSL learning (See Figure 4). This framework posits that learner anxiety in digital environments is not a direct result of technology *per se*, but the net outcome of a dynamic interaction between two core processes:

1. The Technology-as-Stressor Pathway: Certain technological features (e.g., synchronous performance demands, interface complexity, and perceived surveillance) can trigger or amplify anxiety by increasing cognitive load, heightening social-evaluative threat, and undermining the learner’s sense of control.
2. The Technology-as-Scaffold Pathway: Conversely, other technological affordances (e.g., asynchronous practice, immersive simulation, personalized feedback, and strategy training tools) can mitigate anxiety by reducing extraneous cognitive load, offering a safe space for practice, and fostering self-efficacy and meta-cognitive control. The predominance of either pathway—and thus the ultimate affective outcome—is critically moderated by the learner-context interface: the learner’s individual characteristics (e.g., proficiency, trait anxiety) interacting with the specific pedagogical design and the language skill domain in focus. This framework explains the documented duality of effects: the same technological environment (e.g., a video-conferencing platform) can activate different pathways depending on how its features are pedagogically implemented and filtered through individual and situational factors. Moving forward, this model directs research toward disambiguating these pathways, testing specific mediational chains (e.g., from technology feature X, through mediator Y, and to anxiety outcome Z), and designing interventions that systematically enhance the scaffolding pathway while mitigating the stressor pathway. Such a shift is essential for building the efficient and humanized technology-enhanced learning environments envisioned above.

6 Implications

Based on the analysis of existing literature, this study puts forward the following enlightenment from three aspects: theoretical construction, method innovation and practical application.



Theoretical construction needs to break through the traditional perspective that regards anxiety as a purely negative factor and establish a more dialectical theoretical framework. Future research can draw lessons from the research on the distinction between facilitating anxiety and inhibiting anxiety and their correlation with motivation (Luo et al., 2020), so as to further explore the dynamic transformation mechanism between them in the technological environment. At the same time, it is necessary to integrate cognitive load theory, social presence theory and self-regulated learning theory to build a comprehensive model that can explain how technical characteristics affect anxiety experience through multiple mediation paths. Special attention should be paid to the interaction between the uniqueness of Chinese language (such as Chinese character writing system, tone system) and the technical environment, so as to form an anxiety theory with the characteristics of Chinese second language learning.

Research methods should be diversified and deepened. First, expand the sample size and coverage of the study, and carry out long-term follow-up studies to avoid the result bias caused by novel effects. Secondly, using mixed research methods, such as the illustrative sequential design (Putri and Degeng, 2024), quantitative data are combined with qualitative data such as in-depth interviews and classroom observations to comprehensively capture the

complex performance of anxiety. In terms of measurement tools, it is necessary to develop a new anxiety scale suitable for the technology assisted environment, which covers the dimensions of technology operation anxiety and digital social anxiety, and is supplemented by objective data such as physiological indicators, so as to improve the research validity.

Finally, the key to teaching practice is to build a learner centered technology enhanced learning ecology. Teachers and instructional designers should pay attention to the adjustment of intervention setting and duration, reasonably balance the learning arrangement between formal classroom and informal environment, and ensure the continuity and stability of technical intervention. According to the characteristics of different language skills, differentiated support strategies are designed, such as using VR technology to alleviate oral anxiety, providing adjustable listening materials to reduce listening pressure, and establishing online writing communities to enhance writing confidence. Meta-cognitive strategy training is integrated into the process of using technical tools to cultivate learners' ability to plan, monitor and adjust learning, and fundamentally improve their adaptability and anti-anxiety ability in the technical environment. Correctly guide learners to understand the pressure in the technology learning environment, and transform moderate anxiety into learning motivation through goal setting and positive

feedback, so as to avoid the negative effects of excessive anxiety. For technology developers, they should pay attention to user experience design, reduce the threshold of technology operation, and provide personalized learning support and emotional feedback. Technology design should serve the goal of teaching method rather than the opposite, and avoid technology itself becoming a new source of anxiety.

Future research can be carried out from the following aspects: first, to explore the mediating mechanism of technology affecting anxiety and clarify its internal path; The second is to carry out cross-cultural and cross-age comparative research to reveal the differences and similarities between different groups; The third is to explore the application potential of artificial intelligence technology in personalized anxiety intervention and promote practical innovation; Fourth, on this basis, it is committed to building a comprehensive Chinese learning support system integrating technical support, teaching design, and emotional care.

Author contributions

BY: Writing – original draft.

Funding

The author(s) declared that financial support was received for this work and/or its publication. This research was supported by Chongqing Normal University [grant number 23XWB013].

References

- Burston, J. (2015). Twenty years of MALL project implementation: a meta-analysis of learning outcomes. *ReCALL* 27, 4–20. doi: 10.1017/S0958344014000159
- Chen, C. (2021). The effects of online learning on alleviating students' Chinese language learning anxiety: a study in a Chinese university. *Front. Educ. Technol.* 4, 85–107. doi: 10.22158/fet.v4n2p85
- Chen, D. (2021). *A study on the anxiety of junior and intermediate international students in Chinese online classes based on meta-cognitive strategies* (Master's thesis). Shanghai Normal University, Shanghai, China.
- Chen, S. (2023). *A survey on online Chinese learning anxiety of primary level learners* (Master's thesis). Yunnan Normal University, Kunming, China.
- Dong, L., Jamal Mohammed, S., Ahmed Abdel-Al Ibrahim, K., and Rezai, A. (2022). Fostering EFL learners' motivation, anxiety, and self-efficacy through computer-assisted language learning- and mobile-assisted language learning-based instructions. *Front. Psychol.* 13:899557. doi: 10.3389/fpsyg.2022.899557
- Golonka, E. M., Bowles, A. R., Frank, V. M., Richardson, D. L., and Freynik, S. (2014). Technologies for foreign language learning: a review of technology types and their effectiveness. *Comput. Assist. Lang. Learn.* 27, 70–105. doi: 10.1080/09588221.2012.700315
- Guo, T. F. (2021). *A study on the anxiety in online spoken Chinese class of international students at the primary stage* (Master's thesis). Zhejiang University of Science and Technology, Hangzhou, China.
- Hong, Y. (2022). *A study on online writing anxiety and self adjustment strategies of intermediate Chinese learners as second language* (Master's thesis). Shanghai International Studies University, Shanghai, China.
- Horwitz, E. K., Horwitz, M. B., and Cope, J. (1986). Foreign language classroom anxiety. *Mod. Lang. J.* 70, 125–132. doi: 10.1111/j.1540-4781.1986.tb05256.x
- Huang, H., Grant, S., and Yan, J. (2024). 'How anxious I am': the effect of different online modalities on Chinese language beginners' classroom anxiety. *Lang. Learn. J.* 52, 539–555. doi: 10.1080/09571736.2024.2365375
- Kukulski-Hulme, A., and Viberg, O. (2018). Mobile collaborative language learning: state of the art. *Br. J. Educ. Technol.* 49, 207–218. doi: 10.1111/bjet.12580
- Lan, X. (2025). A systematic review of anxiety, motivation, and strategy in learning Chinese as a second language from 2008 to 2022. *Chin. J. Appl. Linguist.* 48, 325–351. doi: 10.1515/CJAL-2025-0209
- Liu, J. J. (2023). *A study on online Chinese language learning anxiety and coping ways of Pakistani students: taking Hebei University as an example* (Master's thesis). Hebei University, Baoding, China.
- Luo, Z., Subramaniam, G., and O'Steen, B. (2020). Will anxiety boost motivation? The relationship between anxiety and motivation in foreign language learning. *Malays. J. ELT Res.* 17, 53–71.
- Lyu, B., and Qi, X. (2020). A review of research on technology-assisted teaching and learning of Chinese as a second or foreign language from 2008 to 2018. *Front. Educ. China* 15, 142–163. doi: 10.1007/s11516-020-0006-8
- Puri, B., Mushtaque, I., Fang, S., Chenhe, G., and Younas, A. (2025). The impact of mobile-based language learning on speaking and learning anxiety, engagement and achievement in Chinese language learning: the mediating role of cognitive load. *Acta Psychol.* 259:105400. doi: 10.1016/j.actpsy.2025.105400
- Putri, N. A., and Degeng, P. D. D. (2024). Utilizing Mobile-Assisted Language Learning (MALL) to alleviate speaking anxiety among EFL students. *Engl. Rev. J. Engl. Educ.* 12, 125–136. doi: 10.25134/erjee.v12i1.9352
- Qian, X. (1999). Foreign students' anxiety in learning Chinese. *Lang. Teach. Linguist. Stud.*, 144–154.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Qiao, P., Yuan, Z., and Li, S. (2025). The impact of anxiety on Chinese as a second language achievement: a meta-analytic perspective. *Front. Psychol.* 16:1620275. doi: 10.3389/fpsyg.2025.1620275
- Sandyaningrum, J. N., and Kusumaningtyas, D. P. S. (2023). The effects of learning Mandarin online with a native lecturer to university students' anxiety. *Mandarinable J. Chin. Stud.* 1, 121–129. doi: 10.20961/mandarinable.v2i2.947
- Srisinthon, P. (2024). Anxiety and engagement in the online classroom: a case study of Thai learners of Chinese as a foreign language. *Knowl. Manag. Elearn.* 16, 716–735. doi: 10.34105/j.kmel.2024.16.033
- Stockwell, G. (2013). "Technology and motivation in English-language teaching and learning," in *International Perspectives on Motivation: Language Learning and Professional Challenges*, ed. E. Ushioda (London: Palgrave Macmillan UK), 156–175. doi: 10.1057/9781137000873_9
- Sun, J. (2022). *A survey of Chinese learners' learning anxiety under network teaching environment* (Master's thesis). Dalian University of Foreign Languages, Dalian, China.
- Sun, W. R. (2022). *A study on the anxiety of online Chinese listening and speaking classes for high school students in Thailand* (Master's thesis). Lanzhou Jiaotong University, Lanzhou, China.
- Wang, Y., Crooks, S. M., and Borst, S. (2017). Chinese language learners' anxiety toward chat partners in computer-mediated communication. *Chin. Second Lang.* 52, 127–147. doi: 10.1075/csl.52.2.02wan
- Watanabe, A. (2020). *Beginner level Japanese students online language learning anxiety* (Master's thesis). Beijing Foreign Studies University, Beijing, China.
- Xie, Y., Chen, Y., and Ryder, L. H. (2019a). Effects of using mobile-based virtual reality on Chinese L2 students' oral proficiency. *Comput. Assist. Lang. Learn.* 34, 1–21. doi: 10.1080/09588221.2019.1604551
- Xie, Y., Ryder, L., and Chen, Y. (2019b). Using interactive virtual reality tools in an advanced Chinese language class: a case study. *TechTrends* 63, 251–259. doi: 10.1007/s11528-019-00389-z
- Xie, Y. J., and Luo, T. (2021). An analysis of the causes and adjustment strategies of anxiety in online Chinese language learning among international students. *J. Hunan Mass Media Vocat. Tech. Coll.* 21, 77–81. doi: 10.16261/j.cnki.cn43-1370/z.2021.02.019
- Xu, W., Zhang, H., Sukjairungwattana, P., and Wang, T. (2022). The roles of motivation, anxiety and learning strategies in online Chinese learning among Thai learners of Chinese as a foreign language. *Front. Psychol.* 13:962492. doi: 10.3389/fpsyg.2022.962492
- Yao, S., Zhang, D., and Shen, Q. (2022). Research on anxiety of learning Chinese as a second or foreign language in and outside Mainland China: a systematic review of the literature 1999–2020. *Front. Psychol.* 13:843858. doi: 10.3389/fpsyg.2022.843858
- Zhang, L., Fang, L., and Shang, J. J. (2024). The application of gamified virtual contexts in international Chinese language education: a utility perspective of discourse cognition and emotional motivation. *Eeduc. Res.* 45, 121–128.
- Zhao, X., and Wang, D. (2024). Domain-specific L2 grit, anxiety, boredom, and enjoyment in online Chinese learning. *Asia-Pac. Educ. Res.* 33, 783–794. doi: 10.1007/s40299-023-00777-3



OPEN ACCESS

EDITED BY

Antonio Bova,
Catholic University of the Sacred Heart,
Italy

REVIEWED BY

Hamza Alshenqeeti,
Taibah University, Saudi Arabia
Gerakini Douka,
Aristotle University of Thessaloniki,
Greece

*CORRESPONDENCE

Claudia Marzi
✉ claudia.marzi@cnr.it

RECEIVED 20 January 2026

REVISED 28 May 2026

ACCEPTED 05 June 2026

PUBLISHED 01 July 2026

CITATION

Polo E and Marzi C (2026)
Finger-tracking captures distinct
cognitive signatures in L1 vs. L2 reading.
Front. Psychol. 17:1792437.
doi: 10.3389/fpsyg.2026.1792437

COPYRIGHT

© 2026 Polo and Marzi. This is an
open-access article distributed under the
terms of the [Creative Commons
Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other
forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution
or reproduction is permitted which does
not comply with these terms.

Finger-tracking captures distinct cognitive signatures in L1 vs. L2 reading

Eva Polo¹ and Claudia Marzi^{2*}

¹Department of Philology, Literature and Linguistics, University of Pisa, Pisa, Italy, ²Institute for Computational Linguistics, National Research Council, Pisa, Italy

Introduction: Predictive and integrative mechanisms play a central role in real-time language comprehension. Understanding how native (L1) and non-native (L2) readers process linguistic information in real time is central to psycholinguistic research. While eye-tracking and event-related potentials (ERPs) have long provided insights into predictive and integrative mechanisms, behavioural methods with fine-grained temporal resolution remain comparatively limited. This study introduces a *finger-tracking* paradigm as a novel, fine-grained method for capturing real-time reading behaviour as participants trace sentences on a touchscreen.

Methods: Two groups of young adult participants – native English speakers (L1) and upper-intermediate English L2 learners – read sentences containing either semantic or morphosyntactic violations, or well-formed control sentences. Tracking speed were analysed both at the token (whole word) and symbol (within-word position) levels.

Results: L1 readers were overall faster than L2 readers and exhibited stronger sensitivity to linguistic anomalies, with modulation of tracking speed during violation processing. In contrast, L2 readers showed a reduced and more gradual sensitivity to violations, consistent with differences in the temporal dynamics of processing across the two groups.

Discussion: In line with neurocognitive and eye-tracking evidence, these findings suggest that L1 reading is supported by rapidly deployed predictive mechanisms that are promptly disrupted by violations, whereas L2 reading is characterised by more incremental integration and reduced anticipatory processing. Overall, finger-tracking emerges as a sensitive, fine-grained behavioural method for studying real-time language processing, revealing distinct cognitive signatures in L1 and L2 reading.

KEYWORDS

behavioral proxy of neural activity, finger-tracking, first and second language processing, linguistic violations, predictive processing, reading dynamics

1 Introduction

Reading is one of the most complex cognitive skills humans perform, involving the continuous integration of perceptual, lexical, syntactic, and semantic information over time. Far from being a passive decoding process, reading comprehension goes beyond simple word recognition, requiring readers to coordinate information across word, sentence, and text levels to build coherent mental representations. In doing so, multiple linguistic and cognitive resources must be dynamically engaged, including memory, attention, and predictive mechanisms that allow readers to anticipate upcoming linguistic input and adjust their expectations in real time. Recent theoretical frameworks

have conceptualized language reading and comprehension as a probabilistic and predictive process, in which incoming words are continuously evaluated against internal models of linguistic structure (Kuperberg and Jaeger, 2016; Pickering and Garrod, 2013). These predictive operations make processing and comprehension efficient, but also render the process sensitive to violations of expected patterns—whether semantic or morphosyntactic. When the linguistic input deviates from prediction, the reader must rapidly revise her/his expectations, a process that has been widely documented in electrophysiological research. Event-related potentials (ERPs), which measure brain activity time-locked to specific linguistic events, have been instrumental in revealing the temporal dynamics of language comprehension. Distinct ERP components are associated with different stages of processing: for example, unexpected or incongruent words typically elicit an enhanced N400 component, reflecting increased semantic integration costs (Li et al., 2024; Kuperberg and Jaeger, 2016; Kutas and Federmeier, 2011), whereas morphosyntactic anomalies often evoke a P600 response, linked to structural reanalysis or repair processes (Aurnhammer et al., 2021; Hagoort and Brown, 2000; Osterhout, 1997; Hagoort et al., 1993). Together, this body of evidence suggests that language reading and comprehension involves continuous cycles of prediction, verification, and updating.

Importantly, the timing and amplitude of the neural responses reveal not only whether a violation is detected, but also how the reader processes linguistic information—anticipatorily or reactively. Readers who rely strongly on predictive processing tend to show early and robust ERP responses to violations, whereas those adopting a more bottom-up, integrative strategy exhibit delayed or attenuated effects (Pickering and Gambi, 2018; Van Petten and Luka, 2012). Understanding this balance between predictive and reactive processing therefore offers a powerful window into the dynamics of language processing and comprehension.

Traditional experimental paradigms such as ERPs and eye-tracking have provided compelling evidence for the temporal and hierarchical organization of language processing. Beside the N400 and P600 components as well-established neural markers of lexical-semantic and syntactic prediction errors (Seyednozadi et al., 2021; Kutas and Federmeier, 2011), eye-tracking studies have, in turn, shown that fixation durations increase when readers encounter words that violate syntactic or semantic expectations, reflecting the cognitive cost of integration and reanalysis (Hubers et al., 2020; Staub, 2015).

Importantly, predictive mechanisms in reading are not uniform across readers but vary as a function of linguistic experience, proficiency, and processing efficiency. Differences in how readers generate, maintain, and update predictions offer a crucial testing ground for theories of real-time language comprehension. In particular, populations that differ in the strength and automaticity of their linguistic representations provide a unique opportunity to examine how predictive and reactive mechanisms are weighted during reading. Second language (L2) readers constitute a paradigmatic case in this respect. Compared to native (L1) readers, they typically operate under increased cognitive demands and reduced processing efficiency, which may affect both the timing and the robustness of prediction-based mechanisms.

1.1 Predictive and reactive mechanisms in L1 and L2 reading

A growing body of research suggests that predictive processing mechanisms differ markedly between native (L1) and non-native (L2) speakers. In L1 comprehension, predictions about upcoming words are generated rapidly and automatically, supported by well-entrenched linguistic representations and probabilistic expectations derived from extensive language exposure. In contrast, L2 readers—particularly late or less proficient learners—tend to rely more heavily on local cues and bottom-up information (Kaan, 2014; Hahne, 2001). These differences reflect not only linguistic knowledge but also differences in processing efficiency and resource allocation (Schlenter, 2023; Kuperberg and Jaeger, 2016).

Neurocognitive evidence has consistently shown that L2 readers exhibit attenuated and delayed neural markers of prediction. Event-related potential (ERP) studies report reduced and temporally shifted N400 and P600 components in L2 comprehension, suggesting that prediction and integration occur later or less efficiently (see, for reviews, Kuperberg and Jaeger, 2016; Van Hell and Tokowicz, 2010). It has been shown that L2 learners tend to be less sensitive than L1 speakers to the conflict between syntactic and semantic processing of a same stimulus (Zheng and Lemhöfer, 2019). Similarly, Hopp (2015a) and Kaan (2014) argue that reduced automaticity in lexical and syntactic access constrains L2 readers' ability to pre-activate upcoming information. Consequently, L2 processing often appears more reactive—that is, based on integration after the input is encountered—whereas L1 processing is predominantly predictive, anticipating incoming words before they appear.

Behaviorally, this distinction is reflected in longer and more variable reading times, greater sensitivity to lexical and morphosyntactic complexity, and reduced contextual facilitation effects in L2 compared to L1 readers (Cop et al., 2015; Kuiken and Vedder, 2012; Kotz, 2009). Eye-tracking studies show that L2 readers often integrate information later, with delayed regressions and weaker parafoveal preview effects (Tiffin-Richards, 2024; Nahatame, 2023).

However, while electrophysiological and eye-movement studies have provided compelling evidence for differences between L1 and L2 readers in the timing and strength of predictive mechanisms, relatively few behavioral paradigms can capture these dynamics online during continuous, ecologically valid reading. This limitation is particularly relevant when the goal is to characterize how prediction and integration unfold moment-by-moment under different processing constraints.

1.2 Finger-tracking as a behavioral window into reading dynamics

Finger-tracking offers a simple yet powerful way to capture the continuous temporal unfolding of reading behavior. By requiring readers to trace the text as they read, this paradigm produces a fine-grained, time-resolved behavioral signal that closely follows the progression of linguistic input across words and

within words. Unlike eye movements, which primarily index shifts of overt visual attention and segment reading into discrete fixation events, finger movements provide a continuous motor trace that unfolds alongside ongoing linguistic processing. At the same time, finger-tracking does not impose the physical constraints associated with electrophysiological recordings and can be implemented in relatively naturalistic reading conditions using the touchscreen of an ordinary tablet. The resulting continuous movement trajectory provides a high-resolution record of reading time. Finger-tracking does not require specialized hardware or calibration, making it suitable for classroom or field studies, while still offering precise temporal and spatial information. As such, it provides a low-cost and ecologically valid behavioral measure that complements traditional eye-tracking approaches, and offers new insights into how visual and motor processes are coordinated during text reading.

In detail, participants continuously trace linguistic stimuli presented on a touchscreen, producing a rich stream of kinematic data that reflects the unfolding of lexical access. Previous research on continuous movement measures has shown that movement trajectories can provide a sensitive behavioral index of ongoing cognitive processes, reflecting the temporal dynamics of information processing as cognition unfolds over time (Freeman et al., 2011; Song and Nakayama, 2009). Accordingly, subtle variations in tracking speed and timing can reveal the cost of prediction, integration, and error detection as the linguistic input unfolds. In this sense, finger-tracking captures not only *when* readers slow down, but also *where* within the word or sentence processing difficulty emerges, offering a spatially and temporally continuous view of online processing.

Finger-tracking has recently been applied across a variety of reading populations, providing converging evidence for its sensitivity to the temporal dynamics of reading. Studies in both adult and developing readers have shown that finger-tracking measures strongly correlate with established eye-tracking indices, thereby supporting its use as a behavioral proxy for the temporal dynamics of oculomotor control and lexical processing (Marzi et al., 2025; Nadalini et al., 2023; Crepaldi et al., 2022). Similarly, in clinical populations, including individuals with mild cognitive impairment and early dementia, finger-tracking has proven effective in detecting subtle alterations in online comprehension and reading strategies (Marzi et al., 2026). Together, these evidence support the ecological validity and generalizability of the finger-tracking paradigm as a fine-grained method for investigating reading dynamics.

Here, we apply finger-tracking to investigate how native and non-native readers differ in their real-time processing of linguistic violations. While native readers are typically thought to rely on highly predictive, top-down mechanisms (Pickering and Clark, 2014), L2 learners often show more reactive and locally driven processing patterns (Kaan, 2014). By examining the temporal signatures of finger movement at both the word (token) and within-word (symbol) levels, we aim to capture how predictive and integrative processes interact across different grain sizes of linguistic input during reading. We hypothesize that L1 readers will show earlier and more pronounced sensitivity to linguistic violations, reflecting disruption of anticipatory processing, whereas

L2 readers may exhibit reduced or more gradual effects, consistent with differences in the temporal dynamics of predictive processing.

1.3 The present study

Building on these premises, the present study investigates whether finger-tracking can capture distinct real-time reading dynamics in native (L1) and non-native (L2) English speakers. Participants read English stimuli containing either semantic or morphosyntactic violations, along with matched correct control items. This experimental design enables a direct comparison of reading behaviors at two complementary levels of analysis: the token level (whole words) and the symbol level (within-word positions).

At the whole-word level, we predicted longer tracking times for violation stimuli, reflecting increased processing demands and disruption on predictive flow. At the within-word level, we expected qualitatively distinct temporal profiles for L1 and L2 readers. For L1 speakers, sensitivity to violations should emerge early and consistently across word positions, consistent with predictive processing and rapid anomaly detection. For L2 speakers, by contrast, differences between correct and violation sentences were expected to emerge later and less distinctly within words, reflecting delayed or more locally driven integration processes.

This approach tests whether finger-tracking can serve as a behavioral proxy for the temporal dynamics of predictive vs. reactive processing, complementing neural evidence from ERP research. More broadly, it explores whether fine-grained temporal signatures of reading, captured through an ecological and accessible motor measure, can reveal cognitive differences in processing and comprehension between L1 and L2 readers.

If successful, this method could offer a bridge between controlled laboratory paradigms and naturalistic reading environments. It would also contribute to ongoing discussions on how prediction, automaticity, and proficiency shape the temporal architecture of comprehension in first and second language processing.

2 Materials and methods

2.1 Participants

Thirty-two young adults participated in the study, divided into two groups of sixteen each. The L1 group consisted of 16 native speakers of English, all residing and tested in Oxford (UK), while the L2 group included 16 English learners, whose first language was Italian, tested in Pisa (Italy). L2 proficiency was self-reported and corresponded on average to CEFR upper-intermediate (B2) level, based on at least 5 years of formal English instruction at secondary-school level, as well as ongoing use of English as a language of university study or in course materials. All participants reported normal or corrected-to-normal vision and no history of reading or language disorders.

Participants were enrolled at different educational levels, ranging from the final year of high school to postgraduate studies (12 – 16 years of formal education, mean = 14.02, standard deviation = 1.61). The two groups were comparable in terms of age (in the range 18–35, mean age 25.7 for L1 and 25.9 for L2) and educational background (mean years of formal schooling 13.9 for L1 and 14.2 for L2). Specifically, the sample included 12 high school students (7 L1, 5 L2), 15 bachelor students (6 L1, 9 L2), and 5 master's students (3 L1, 2 L2).

All participants took part voluntarily and provided written informed consent prior to participation. The study adhered to the ethical standards of the previously approved protocol on developmental reading (Ethical approval 0037523/2021, Italian National Research Council—Committee for Research Ethics) and complied with the principles of the Declaration of Helsinki.

2.2 Materials

The experimental materials consisted of 16 English sentences, each constructed in two versions: a correct version and a violated version. Violations were of two types—semantic and morphosyntactic—yielding a total of 32 stimuli. Semantic violations involved contextually anomalous words (e.g., *The volcano was meticulously *eaten by a team of experienced scientists using advanced thermal imaging equipment*, correct version: *observed*), while morphosyntactic violations introduced ungrammatical agreement or tense errors (e.g., *The old house *has broken into while the owners were on vacation in the mountains*, correct version: *was*). These types of manipulations are commonly used in psycholinguistic research to elicit sensitivity to semantic implausibility and morphosyntactic violations, typically producing robust processing costs in both behavioral and neurocognitive measures (e.g., Osterhout 2015; Hagoort 2003; Kutas and Hillyard 1983).

All sentences in their correct versions were designed to be comprehensible for upper-intermediate L2 readers, and lexical selection was controlled to avoid low-frequency or highly specialized lexical items. Many stimuli consisted of two or more clauses, typically separated by a comma or a full stop. Overall, stimuli were relatively long, with an average of 25.56 tokens per sentence (sd = 5.0, range = 17–35) and a mean token length of 4.50 characters (sd = 0.76, range = 3.6–6.1), ensuring that the reading task required sustained processing across multiple words. Each participant was presented with 16 stimuli: eight correct and eight containing a violation. Stimulus assignment was counterbalanced across participants, such that each stimulus appeared equally often in its correct and violated form across the two groups. This ensured that no participant saw both versions of the same item.

Since the experimental design aimed to approximate relatively naturalistic reading conditions while maintaining control over the occurrence of semantic and morphosyntactic violations, no filler items were included, as the study prioritized continuity of online reading dynamics rather than to function as a classical acceptability-judgement paradigm. Whereas filler items are commonly used in highly controlled psycholinguistic paradigms to reduce strategic responding (see, for a recent review,

Covey and Gabriele, 2023), the inclusion of well-formed control items introduced variability in sentence acceptability and reduced the predictability of anomalous trials. Additionally, recent evidence suggests that filler composition itself may shape participants' expectations and processing behavior (Arehalli and Wittenberg, 2021). This consideration is particularly relevant in paradigms investigating predictive processing, where the statistical structure of the experimental environment may directly influence anticipatory behavior. Each stimulus was followed by either a comprehension question or a grammaticality judgement. This combination served to maintain attention throughout the task and encouraged participants to engage both with sentence meaning and with linguistic well-formedness during continuous reading.

Each stimulus was presented in a sans-serif font (Arial, 21.25 pt) on a 10.1-inch Samsung tablet, with a 14.9 cm × 24.5 cm touchscreen at a resolution of 1,920 × 1,200 pixels. This setup allowed participants to trace each sentence with their finger while reading, providing a fine-grained, temporally precise measure of reading behavior.

2.3 Procedures

Participants were tested individually in a quiet room to minimize external distractions. After a brief familiarization phase consisting of a practice trial, they were instructed to read each stimulus silently while tracing the text with their index finger on the touchscreen. They were encouraged to move their finger smoothly and at their natural reading pace.

After reading each stimulus, a comprehension question appeared on screen, requiring a yes/no response (e.g., “Is this sentence correct?” or “Did the agent x make the action?”). Responses were given by tapping on the corresponding option, and both response accuracy and reaction times were recorded. These questions were included to ensure attentive reading but were not the primary dependent measure of our analyses.

The entire experiment lasted approximately 30 min per participant, including instructions and the main reading task.

2.4 Data acquisition and alignment

Reading sessions were conducted on a tablet in landscape orientation. The *Readlet* application (Marzi et al., 2025; Nadalini et al., 2023; Taxitari et al., 2021; Marzi et al., 2020; Ferro et al., 2018) recorded finger movements directly via the tablet's touchscreen interface. Finger trajectories are captured with an effective sampling rate ranging from approximately 60–120 Hz, producing a continuous stream of spatially and temporally resolved touch coordinates, each encoded with its position on the screen and its timestamp. This allowed reconstruction of raw finger movement trajectories during reading. Text-to-finger alignment was performed using a convolution-based mapping procedure that identified the optimal correspondence between continuous touch-event sequences and textual layout. For each uninterrupted sequence of touch events within a letter-specific spatial region, tracking time was computed as the temporal difference between the

first and last event. Letter-level tracking times were then aggregated into higher-level linguistic units (e.g., word tokens) by summation across constituent letters, yielding unit-level measures aligned with the stimulus structure.

This alignment procedure enabled a fine-grained reconstruction of the temporal evolution of finger position relative to the linguistic input, closely paralleling how a sequence of eye fixations is mapped onto a sequence of words (Crepaldi et al., 2022; Nadalini et al., 2023). This enables a fine-grained reconstruction of where the reader's finger was directed at any given point during reading and, indirectly, the locus of ongoing visual-motor engagement with the text.

2.5 Data analysis

Finger-tracking data were processed to derive two complementary measures at two levels of granularity: (i) the token level, corresponding to whole words, and (ii) the symbol level, corresponding to within-word positions.

At the token-level, total tracing time was computed as the duration between the first and last touch event associated with each given token. This measure provides a coarse-grained estimate of reading and lexical processing time.

At the symbol-level, finger trajectories were further resampled to produce fine-grained time series aligned with each orthographic character within a word. This measure captures subtle, within-word variations in finger movement dynamics that may reflect moment-to-moment processing difficulty.

The combination of token- and symbol-level analyses provides insight into how linguistic violations are detected and integrated over time in first and second language processing.

All statistical analyses and data visualizations were performed using R (version R-4.5.1, R Core Team, 2025).

Before statistical analyses, finger-tracking data were preprocessed to remove extreme outliers. Extreme values were trimmed using an empirically motivated cutoff of 1.3 seconds per token. This value lies well beyond the 99th percentile of the distribution and isolates the long-tail observations associated with pauses or accidental touches. This procedure removed 0.97% of data points.

Trials with a tracking time of 0 milliseconds (indicating no touch event) were inspected separately prior to exclusion, to check for possible systematic differences across groups or sentence positions. The overall proportion of zero-duration events did not differ significantly between L1 and L2 readers (Fisher's exact test, $p = 0.62$). However, their distribution across sentence positions did. A non-parametric Kolmogorov-Smirnov test comparing the positional distributions of zero-duration events revealed a significant difference between groups when position was expressed relative to the sentence end ($D = 0.18$, $p = 0.002$). Specifically, L1 readers showed a marked concentration of zero-duration events near sentence-final positions, consistent with anticipatory processing of the final tokens, whereas L2 readers displayed a more uniform distribution across positions. On the basis of this preliminary inspection, zero-duration events were

excluded from subsequent analyses, corresponding to 2.90% of data points removed during preprocessing.

3 Results

3.1 Overall reading performance

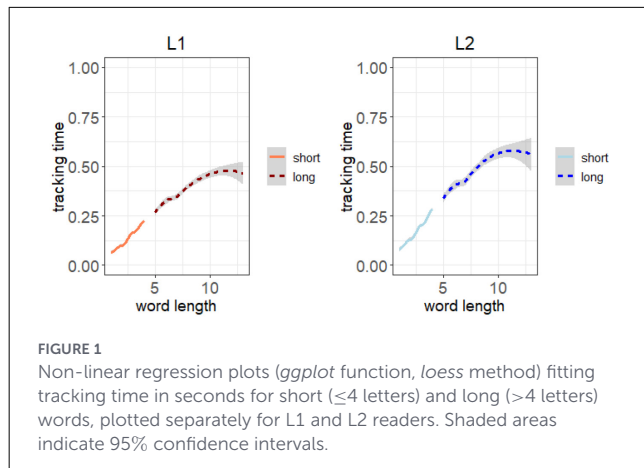
Preliminary results indicated significantly longer tracking times for L2 than L1 readers overall, together with lower comprehension accuracy. A non-parametric Kruskal-Wallis test on tracking time by tokens confirmed a significant main effect for group [$\chi^2(1) = 98.69$, $p < 2.2e - 16$], with longer times for L2 (mean = 0.27, median = 0.20) than for L1 readers (mean = 0.22, median = 0.17). Similarly, accuracy on comprehension and grammaticality questions differed significantly between groups [$\chi^2(1) = 5094.8$, $p < 2.2e - 16$], with L2 readers showing lower participant-level proportion of correct responses (mean = 0.54, median = 0.55) compared to L1 readers (mean = 0.75, median = 0.82).

To evaluate whether finger-tracking captures well-established lexical determinants of reading time, we examined the influence of word length and frequency on token-level tracking times using Generalized Additive Models (GAMs). We firstly fitted a GAM with group (L1 vs. L2), word length (expressed as number of orthographic letters), and their interaction as fixed effects, and random intercepts for participants. Crucially, a significant group $\times$ length interaction emerged ($p < 2e - 16$), with L2 readers showing a steeper increase in tracking time with word length compared to L1 readers. This pattern suggests that non-native readers can be more sensitive to surface-level orthographic demands as compared with native ones. The model explained approximately 50 % of the variance (adjusted $R^2 = 0.498$). Full model specifications are provided in Table A1 in Appendix.

Next, we examined the influence of token frequency (log-transformed and derived from the Subtlex corpus, Van Heuven et al. 2014) on finger-tracking times fitting a GAM with the interaction of group and frequency, and random intercepts for participants. Results confirmed a strong facilitatory effect of frequency. L2 readers showed overall longer tracking times than L1 readers ($p < 0.001$), with a significantly stronger effect of frequency for L2 than for L1 ($p < 8e - 14$), suggesting that L2 readers may depend more heavily on word-level statistical information during reading. The model explained approximately 41 % of the variance (adjusted $R^2 = 0.406$). See Table A2 in Appendix for full model specifications.

Together, these preliminary results confirm that finger-tracking reliably captures canonical lexical influences on reading behavior, while also revealing systematic group differences in the strength of such effects.

To further explore the effect of word length, tokens were divided into two groups: short (≤ 4 characters) and long ($> 4 \leq 13$ characters). Tracking times in seconds were analyzed fitting a GAM with fixed effects for group (L1 vs. L2), length group (short vs. long), their interaction, and a random intercept for participants. The model revealed a significant main effect of length group ($p < 2e - 16$), with longer words associated with increased tracking times. While the baseline tracking time for short



words did not differ significantly between L1 and L2 readers ($p > 0.05$), the group $\times$ length interaction was highly significant ($p < 7e - 16$), indicating that L2 readers show a steeper increase in tracking time for long words compared to L1 readers. Full model specifications are provided in Table A3 in [Appendix](#). [Figure 1](#) shows mean tracking times separately for L1 and L2 readers. The plot highlights that both groups exhibit longer tracking times for long words, but the effect is clearly amplified in L2 readers, suggesting greater surface-level orthographic demands.

Overall, these evidence suggest that L2 readers are slower and more sensitive to word-level orthographic properties, consistent with a more effortful, locally driven reading strategy.

3.2 Sensitivity to linguistic violations

Within-group comparisons revealed that tracking times were modulated by sentence well-formedness (correct vs. violation) in L1, but less so in L2. Specifically, for L1 readers, stimuli containing linguistic violations elicited significantly longer tracking times across all tokens compared to well-formed control stimuli [*Kruskal-Wallis* $\chi^2(1) = 5.68$, $p = 0.017$]. In contrast, no effect of sentence well-formedness was observed for L2 readers, whose tracking times did not significantly differ between violated and correct stimuli [*Kruskal-Wallis* $\chi^2(1) = 0.02$, $p > 0.05$].

A parallel within-group analysis was conducted on response accuracy. For L1 readers, sentence well-formedness significantly affected overall accuracy, with lower accuracy for sentences containing violations compared to correct sentences [*Kruskal-Wallis* $\chi^2(1) = 7.16$, $p < 0.01$]. For L2 readers, the effect of sentence well-formedness did not reach conventional levels of statistical significance [*Kruskal-Wallis* $\chi^2(1) = 3.63$, $p = 0.057$], indicating only a marginal trend toward reduced accuracy for violated sentences. Thus, response accuracy mirrored the pattern observed in tracking times, with clear sensitivity to violations in L1 readers and a reduced or absent effect in the L2 group, in line with the overall lower accuracy previously observed in L2 readers.

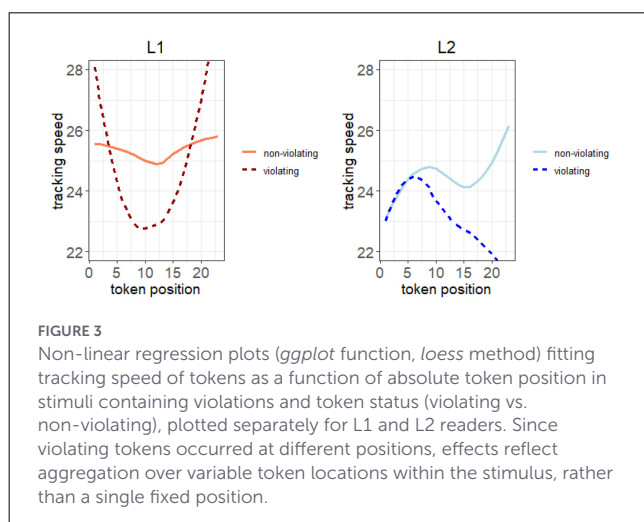
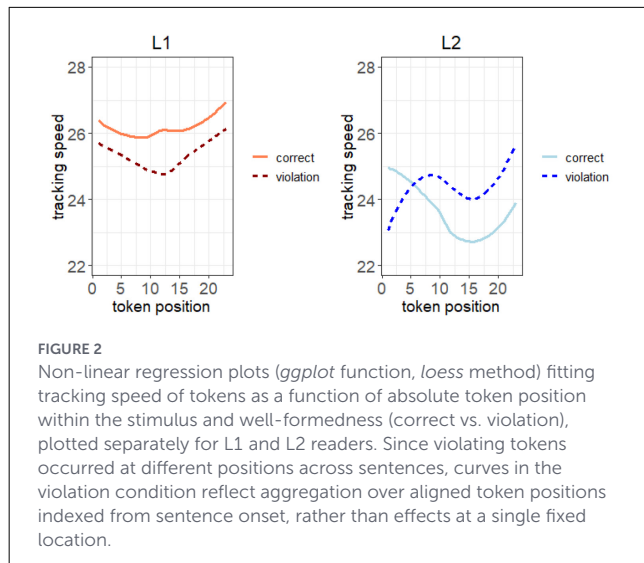
While tracking time provides a straightforward estimate of processing cost, it is inherently influenced by token

length, as shown in [Figure 1](#). To better capture moment-to-moment reading efficiency independently of word length, we complemented time-based analyses with a measure of tracking speed. Tracking speed was computed as the ratio between each token length (in number of characters) and its tracking time, yielding a continuous estimate of reading speed as characters per second. This transformation allows processing dynamics to be compared across tokens of different lengths and provides a more direct index of online reading efficiency. [Figure 2](#) shows mean tracking speed as a function of token absolute position for correct and violation sentences, plotted separately for L1 and L2 readers. Since violating tokens occurred at different positions across sentences, position-based effects reflect aggregation over token positions indexed from sentence onset rather than a fixed critical location. Related GAM specifications are provided in Tables A4, A5 in [Appendix](#). Tracking speed profiles revealed distinct temporal dynamics across the two groups, with stronger modulation by sentence well-formedness in L1 readers than in L2 readers. Descriptive statistics further indicated a reduction in median tracking speed for L1 readers (22.97 vs. 21.85 char/s for correct and violation stimuli, respectively), which was statistically significant (*Mann-Whitney* $W = 6.18 \times 10^6$, $p < 0.001$). By contrast, L2 readers showed highly similar median tracking speeds across conditions (18.27 vs. 18.47 char/s), with no significant difference (*Mann-Whitney* $W = 5.66 \times 10^6$, $p = 0.65$).

Thus, L1 readers exhibited a generally higher tracking speed across stimuli, with a clear overall reduction in speed for violation sentences relative to correct ones. In violation sentences, the speed profile for L1 readers displayed a characteristic *V-shaped* pattern, with a transient slowdown followed by a speed recovery toward the end of the stimulus. This pattern is consistent with a temporally circumscribed disruption in processing associated with the presence of a linguistic violation, potentially reflecting predictive and globally coordinated reading mechanisms.

In contrast, L2 readers showed overall slower tracking speeds and a markedly different speed profile as compared to L1 readers. Rather than exhibiting a coherent, focal modulation, their tracking speed displayed more distributed and variable fluctuations across the stimuli. Although the smoothed trajectories for correct and violation sentences appeared to diverge in the initial portion of the stimulus, they subsequently evolved in a largely parallel manner. Given that inferential analyses did not reveal a reliable effect of sentence well-formedness in L2 readers, these descriptive differences should be interpreted cautiously. Overall, the L2 profiles suggest greater variability in tracking dynamics across sentence positions, without clear evidence for robust modulation by linguistic violations.

Since the analysis reported in [Figure 2](#) provides a global description of tracking dynamics across sentence positions, to further disentangle local vs. global effects of linguistic violations, we conducted an additional analysis restricted to stimuli containing a violation. Within these ones, tracking speed at the violated token was contrasted with tracking speed at semantically/grammatically correct tokens. This approach allows us to assess whether processing costs are tightly localized at the anomalous input or instead distributed across the sentence. Correct tokens were defined as all tokens in the same sentence excluding the anomalous token.



Importantly, effects reflect aggregation across tokens differing in their position relative to sentence onset.

As shown in Figure 3, L1 readers exhibited a pronounced slowdown specifically at the violating token, producing a marked “U-shaped” profile, whereas non-violating tokens displayed only a shallow “V-shaped” decrease in speed. Importantly, this effect is observed despite the fact that violating tokens were distributed across sentence positions (range = 1–23). Nevertheless, the minimum tracking speed consistently occurred around the median violation position across stimuli (median = 10, $M = 10.43$, $SD = 6.15$), suggesting that the observed pattern cannot be attributed to a fixed positional confound. In contrast, L2 readers showed a more variable pattern, with tokens carrying the violation that show a progressively decreasing speed toward the end of the stimuli. This suggests that non-native readers integrate the violation more gradually, relying less on anticipatory mechanisms and more on cumulative, *post-hoc* processing across the sentence.

Overall, these evidence highlight a qualitative difference in the reading strategies of L1 and L2 participants. Native readers show a selective, token-specific slowdown when encountering violations,

while non-native readers exhibit a slower, more distributed adjustment, consistent with delayed integrative processing.

3.3 Effects of morphosyntactic vs. semantic violations

We next examined whether tracking speed differed as a function of violation type (morphosyntactic vs. semantic) within each group. For L1 readers, no significant difference was observed between morphosyntactic and semantic violations (Kruskal–Wallis $\chi^2 = 2.43$, $p = 0.12$), indicating that native readers slow down similarly in response to both types of anomalies. In contrast, L2 readers exhibited a significant difference (Kruskal–Wallis $\chi^2 = 4.64$, $p = 0.031$), with greater slowdown for morphosyntactic than for semantic violations, suggesting that non-native readers are particularly sensitive to formal grammatical discrepancies.

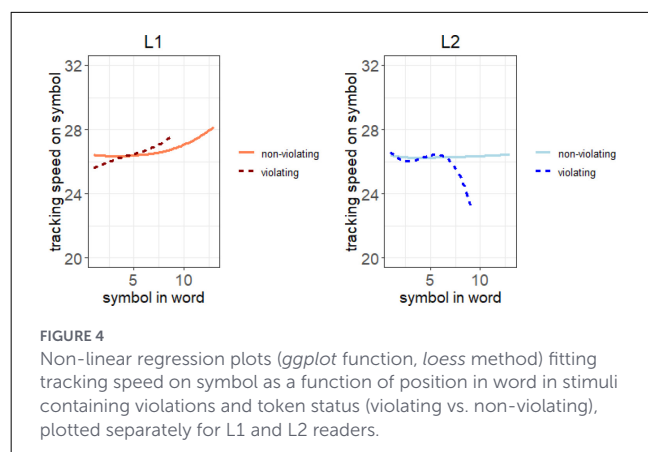
We then compared L1 and L2 readers separately for each violation type. For morphosyntactic violations, L1 readers were substantially faster than L2 readers ($\chi^2 = 67.41$, $p < 2.2e - 16$), whereas for semantic violations, L1 readers also outperformed L2 readers, although the group difference was smaller ($\chi^2 = 18.09$, $p < 2.2e - 5$).

Additionally, *post-hoc* Dunn tests (with Holm correction) conducted separately for L1 and L2 readers revealed markedly different patterns on morpho-syntactic violations. While L1 readers showed selective slowdowns limited to highly marked constructions, such as Saxon genitives (e.g., *the teacher's book*, see Rosenbach, 2005) and, to a lesser extent, double negation—a construction that may increase processing cost due to its sensitivity to contextual predictability and increased integration demands (Rück et al., 2021)—L2 readers exhibited a redistribution of processing costs across violation subtypes, with even canonical structures such as SVO word order (Subject-Verb-Object; e.g., *The boy eats the apple*) eliciting substantial slowdowns. Full results of the *post-hoc* comparisons are reported in Table A6 in Appendix.

Together these findings suggest that L2 readers are generally slower and that this lag is especially pronounced for morphosyntactic violations, consistent with a more reactive, locally driven processing strategy. More generally, these results suggest a more distributed processing cost for morphosyntactic violations in L2 readers, whereas semantic violations tend to elicit more circumscribed effects. Overall, these findings are consistent with a more selective and predictive reading strategy in L1 readers, and a more reactive and distributed processing profile in L2 readers.

3.4 Within-word temporal dynamics

To further characterize L1–L2 differences in real-time processing, we complemented the token-level analyses with the temporal dynamics of tracking speed within words. By analyzing successive segments of each token, this approach provides a fine-grained view of how reading behavior unfolds in response to linguistic violations over the course of a word. A within-word perspective is, in fact, particularly informative for distinguishing



predictive from reactive processing, as early modulation of tracking speed is expected when violations are anticipated or immediately recognized, whereas delayed effects point to more incremental, locally driven integration.

Overall, L2 readers were significantly slower than L1 readers even at the symbol level (Kruskal–Wallis $\chi^2 = 278.66$, $p < 2.2e - 16$), confirming that group differences in reading speed are present at both token- and symbol-level resolutions. When considering all symbols in word token, tracking speed differed reliably between symbols belonging to correct vs. violation stimuli in L1 readers ($\chi^2 = 64.08$, $p < 0.001$), indicating a robust global effect of sentence-level violations. The same contrast did not reach significance in L2 readers ($\chi^2 = 3.38$, $p = 0.066$).

We then restricted the analysis to stimuli containing a violation and contrasted symbols belonging to the violating token with symbols from non-violating tokens within the same stimulus. For L1 readers, symbols carrying the violation were tracked significantly more slowly than non-violating symbols ($\chi^2 = 15.86$, $p < 0.001$). No such difference was observed in L2 readers ($\chi^2 = 0.61$, $p = 0.44$), with any local divergence being restricted to word-final segments of relatively long words (exceeding mean word length of 4 characters, median 4.14). To visually characterize these effects, [Figure 4](#) shows non-linear regression plots of tracking speed across symbol positions within words, separately for symbols belonging to violating vs. non-violating tokens, in sentences containing a violation. Related GAM specifications are provided in [Tables A7, A8](#) in [Appendix](#).

Together, these results indicate that in L1 readers sensitivity to violations operates both globally across the sentence and locally at the level of the violating symbol, whereas in L2 readers it emerged only at later positions within longer words, suggesting a more gradual and incremental integration process.

3.5 Summary of main results

Across analyses, finger-tracking measures reliably differentiated L1 and L2 reading behavior. Considering sentence-wide dynamics derived from token-level analyses, L1 readers showed systematic modulations of tracking speed in response to violations, with violating tokens eliciting sharper and localized slowdowns, whereas L2 readers exhibited more variable and less systematic profiles. At

the symbol level, violating tokens showed early slowdown in L1 but delayed effects in L2.

Together, these results indicate both quantitative and qualitative differences in the temporal dynamics of reading between native and non-native speakers, observable across token- and symbol-level measures.

Overall, the present analyses show that (i) finger-tracking yields systematic and group-sensitive measures of reading speed, (ii) these measures are modulated by the presence and type of linguistic violations, and (iii) they reveal distinct temporal profiles across groups in the unfolding of sentence processing.

4 Discussion

4.1 General overview of findings

Our results provide novel insights into the online dynamics of sentence processing in L1 and L2 readers, as captured through fine-grained finger-tracking measures.

Across analyses, finger-tracking measures reliably differentiated L1 and L2 reading behavior. Native readers were consistently faster than L2 readers and showed higher comprehension accuracy across conditions, confirming robust group differences in overall reading efficiency.

These global differences were observed across multiple levels of analysis, including sentence-wide dynamics derived from token position, as well as token- and symbol-level measures.

Importantly, the finger-tracking paradigm proved to be sensitive to linguistic well-formedness, capturing both overall slowdowns and fine-grained temporal modulations associated with processing surprisal and difficulty.

4.2 Predictive vs. reactive processing

When focusing on individual tokens within violation-containing sentences, the effect of the critical token differed markedly between groups. In L1 readers, the token introducing the violation elicited an immediate and pronounced slowdown, whereas in L2 readers, the effect was more gradual, with tracking speed declining progressively toward the sentence end.

These patterns are consistent with the overall token-level analyses: sensitivity to violations emerged early and robustly in L1 readers, whereas in L2 readers no uniform token-level contrast was observed. This suggests that anomaly-related effects in non-native reading does not operate continuously across the sentence, but rather unfold more gradually across the processing stream.

Collectively, these findings are broadly consistent with previous evidence suggesting that native speakers engage fast, predictive-driven mechanisms during sentence comprehension, including during the processing of syntactic or semantic anomalies ([Staub and Goddard, 2019](#); [Rayner, 1998](#)), with neural responses to words modulated by their informational content and effort required to update the reader's mental model ([Wang et al., 2026](#); [Frank et al., 2015](#)). Importantly, this literature typically does not involve continuous, time-resolved tracking of visual-motor behavior. In contrast, evidence from L2 reading suggests more

incremental and integration-based processing profiles, which have been described in terms of reduced predictive processing and greater reliance on bottom-up information (Hopp, 2016, 2015b; Clahsen and Felser, 2006).

4.3 Morphosyntactic vs. semantic violations

Analyses examining the effect of violation type revealed clear differences between L1 and L2 readers. For L1 readers, no significant difference in tracking speed was observed between morphosyntactic and semantic violations, indicating that native speakers slow down similarly in response to both types of anomalies. This suggests that predictive processing in L1 is engaged continuously, and the detection of a violation—be it morphosyntactic or semantic—elicits a comparable adjustment in reading speed. In contrast, L2 readers exhibited greater sensitivity to morphosyntactic violations than to semantic ones, suggesting that non-native readers are particularly affected by formal structural discrepancies. Morphosyntactic violations may engage more distributed processing demands, whereas semantic violations appear to produce more circumscribed effects.

The pattern observed in L2 readers may reflect a more locally driven and incremental processing profile, in which formal grammatical violations require more effortful processing, whereas native readers rely on continuous predictive monitoring that responds similarly to both types of anomalies, revealing slowdowns limited to highly marked constructions (Hagoort, 2003). Notably, this interpretation should be considered in light of differences in language proficiency between the two groups. Although the L2 participants in the present study represented relatively advanced learners, differences in cumulative language experience and processing automaticity relative to native speakers may still have influenced the efficiency of online linguistic integration.

More generally, this interpretation aligns with evidence from ERP studies showing that the processing of individual words during reading is shaped by interactions between lexical-level information and message-level representations, and that these integrative processes are modulated by reading proficiency (Wang et al., 2026; Hopp, 2016; Stafura and Perfetti, 2014).

4.4 Temporal dynamics within words

Analyzing reading at the symbol level within each word provided a more fine-grained perspective on how linguistic violations are detected and processed in real time. This approach allowed us to distinguish between early, predictive adjustments and later, reactive integration, complementing the token-level analysis.

Our symbol-level results revealed clear differences between L1 and L2 readers. In L1 readers, symbols within violating tokens elicited an early slowdown, observable at the word onset. This effect suggests that native readers engage in rapid, anticipatory processing, dynamically adjusting reading speed as soon as a potential violation is encountered. The early emergence of this slowdown aligns with continuous predictive monitoring, consistent with prior evidence from eye-tracking and ERP studies

showing rapid detection of syntactic and semantic anomalies (Staub and Goddard, 2019; Rayner, 1998).

In contrast, L2 readers did not show an immediate symbol-level effect for violating tokens. Sensitivity to violations appears to be limited to later positions within the word, particularly in longer tokens. This pattern may be consistent with a more gradual and incremental integration process, whereby anomaly-related effects become observable only after sufficient lexical and syntactic information has been accumulated. Such a pattern is broadly consistent with accounts proposing that non-native readers rely more heavily on reactive, bottom-up mechanisms, rather than continuous prediction (Hopp, 2016, 2015b; Clahsen and Felser, 2006).

Taken together, the symbol-level analyses reinforce and extend the sentence- and token-level findings. They suggest that the temporal unfolding of reading behavior differs both quantitatively and qualitatively between native and non-native readers: while L1 reading is characterized by earlier and more broadly distributed sensitivity to anomalies, L2 reading appears to exhibit more gradual and context-dependent responses to unexpected input.

Crucially, these patterns are compatible with hierarchical accounts of linguistic prediction, which posit that predictive processes operate at multiple representational levels and unfold over different time scales during word recognition (Heilbron et al., 2022). From this perspective, the early within-word slowdowns observed in L1 readers may reflect rapid prediction error signals at sublexical or lexical levels, whereas the delayed effects in L2 readers suggest that predictive mechanisms are either weaker or shifted to later stages of integration.

Our results have important theoretical implications. First, they provide behavioral evidence that predictive mechanisms in L1 operate across multiple levels of granularity, from sentence structure down to individual symbols within words. Second, they highlight that L2 reading, even in upper-intermediate readers, may be constrained by the temporal dynamics of processing, resulting in delayed responsiveness to anomalies.

4.5 Methodological and theoretical implications

Our study suggests that finger-tracking provides a continuous, fine-grained behavioral measure for real-time reading dynamics, complementing traditional eye-tracking and ERP measures (Kessler et al., 2021). By capturing dynamic adjustments in tracking speed, this method reveals how readers respond to linguistic anomalies at multiple levels—sentence, token, and symbol.

Differences between L1 and L2 readers suggest that non-native readers may compensate for less efficient predictive mechanisms by relying more on cumulative lexical and syntactic information across the sentence, resulting in slower and more gradual sensitivity to violations. This highlights quantitative and qualitative differences in processing strategies and reinforces the utility of finger-tracking for probing the temporal micro-dynamics of reading.

Notably, the temporal dynamics observed at the symbol level parallel ERP findings, which consistently report earlier neural responses to violations in L1 readers and delayed or attenuated

responses in L2 readers (Hopp, 2015b; Clahsen and Felser, 2006; Rayner, 1998). Similar tendencies have also been reported in eye-tracking studies of sentence reading, where L2 readers often show reduced anticipatory processing and greater reliance on later integrative mechanisms relative to native readers (Kim and Grüter, 2021; Staub and Goddard, 2019; Hopp, 2016). This convergence indicates that finger-tracking captures behavioral manifestations of predictive vs. reactive processing, providing a continuous, real-time index that complements established methodologies for investigating online language processing—such as neurophysiological or eye-tracking measures.

From a theoretical perspective, these results support models of sentence processing that emphasize gradient predictability and incremental integration, and are compatible with hierarchical accounts of linguistic prediction in which anticipatory processes operate across multiple temporal and representational scales (Heilbron et al., 2022). Crucially, our findings show that such distinctions are observable not only in neural signals, but also in continuous behavioral dynamics.

From an applied standpoint, understanding how readers adjust reading speed in response to anomalies can inform educational and assessment practices. Distinct temporal profiles may serve as behavioral markers of processing efficiency, helping identify learners who rely more heavily on reactive strategies, and potentially guiding interventions aimed at strengthening anticipatory processing in L2 readers, for example through guided sentence-prediction exercises or structured exposure to language-specific highly constraining contexts (e.g., non-canonical word order, morphosyntactic dependencies, or constructions associated with frequent learner difficulties).

Overall, our findings reinforce the value of finger-tracking as a scalable and sensitive method for investigating real-time reading strategies and comprehension, bridging theoretical models of predictive processing with practical applications in assessment and pedagogy.

Several limitations should be however acknowledged. Our sample size was relatively small, although participant samples of this scale are not uncommon in controlled psycholinguistic studies involving fine-grained behavioral measures of L2 processing (e.g., Soares et al., 2019). In addition, L2 proficiency was assessed via self-report rather than objective measures, which may constrain the interpretation of group differences. Future studies should examine whether these patterns hold with larger samples and broader proficiency ranges, and across different types of linguistic anomalies, to establish the robustness and generalizability of the present findings.

Data availability statement

Text-aligned data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by Committee for Research Ethics - Italian National Research Council. The studies were conducted in accordance with the local legislation and

institutional requirements. The participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

EP: Writing – original draft, Investigation, Resources, Methodology. CM: Methodology, Supervision, Writing – review & editing, Conceptualization, Validation, Writing – original draft, Formal analysis, Visualization.

Funding

The author(s) declared that financial support was not received for this work and/or its publication.

Acknowledgments

This work was made possible thanks to the PRIN project “Reading to understand: an ICT-driven, large-scale investigation of early grade children’s reading strategies” (2017W8HFRX) coordinated by the ComphysLab at the Institute for Computational Linguistics, Italian National Research Council (<http://www.comphyslab.it>), which supported the development of the finger-tracking application adopted in this study.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that Generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of

their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Arehalli, S., and Wittenberg, E. (2021). Experimental filler design influences error correction rates in a word restoration paradigm. *Linguist. Vanguard* 7:20200052. doi: 10.1515/lingvan-2020-0052
- Aurnhammer, C., Delogu, F., Schulz, M., Brouwer, H., and Crocker, M. W. (2021). Retrieval (N400) and integration (P600) in expectation-based comprehension. *PLoS ONE* 16:e0257430. doi: 10.1371/journal.pone.0257430
- Clahsen, H., and Felser, C. (2006). Grammatical processing in language learners. *Appl. Psycholinguist.* 27, 3–42. doi: 10.1017/S0142716406060024
- Cop, U., Keuleers, E., Drieghe, D., and Duyck, W. (2015). Frequency effects in monolingual and bilingual natural reading. *Psychon. Bull. Review* 22, 1216–1234. doi: 10.3758/s13423-015-0819-2
- Covey, L., and Gabriele, A. (2023). “Psycholinguistic methods in second language research,” in *Current Approaches in Second Language Acquisition Research: A Practical Guide*, eds. A. Mackey, and S. M. Gass (Chichester: John Wiley & Sons), (Hoboken: Wiley Blackwell), 170–194. doi: 10.1002/9781394259670.ch8
- Crepaldi, D., Ferro, M., Marzi, C., Nadalini, A., Pirrelli, V., and Taxitari, L. (2022). “Finger movements and eye movements during adults’ silent and oral reading,” in *Developing Language and Literacy: Studies in Honor of Dorit Diskin Ravid*, eds. R. Levie, A. Bar-On, O. Ashkenazi, E. Dattner, and G. Brandes (Cham: Springer International Publishing), 443–471. doi: 10.1007/978-3-030-99891-2_17
- Ferro, M., Cappa, C., Giulivi, S., Marzi, C., Nahli, O., Cardillo, F. A., et al. (2018). “Readlet: reading for understanding,” in *Proceedings of 5th IEEE Congress on Information Science Technology (IEEE CIST’18) (Marrakech)* (New York, NY: IEEE). doi: 10.1109/CIST.2018.8596412
- Frank, S. L., Otten, L. J., Galli, G., and Vigliocco, G. (2015). The ERP response to the amount of information conveyed by words in sentences. *Brain Lang.* 140, 1–11. doi: 10.1016/j.bandl.2014.10.006
- Freeman, J. B., Dale, R., and Farmer, T. A. (2011). Hand in motion reveals mind in motion. *Front. Psychol.* 2:59. doi: 10.3389/fpsyg.2011.00059
- Hagoort, P. (2003). Interplay between syntax and semantics during sentence comprehension: ERP effects of combining syntactic and semantic violations. *J. Cogn. Neurosci.* 15, 883–899. doi: 10.1162/08989290322370807
- Hagoort, P., Brown, C., and Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Lang. Cogn. Process.* 8, 439–483. doi: 10.1080/01690969308407585
- Hagoort, P., and Brown, C. M. (2000). ERP effects of listening to speech compared to reading: the P600/SPS to syntactic violations in spoken sentences and rapid serial visual presentation. *Neuropsychologia* 38, 1531–1549. doi: 10.1016/S0028-3932(00)00053-1
- Hahne, A. (2001). What’s different in second-language processing? Evidence from event-related brain potentials. *J. Psycholinguist. Res.* 30, 251–266. doi: 10.1023/A:1010490917575
- Heilbron, M., Armeni, K., Schoffelen, J.-M., Hagoort, P., and De Lange, F. P. (2022). A hierarchy of linguistic predictions during natural language comprehension. *Proc. Nat. Acad. Sci. U. S. A.* 119:e2201968119. doi: 10.1073/pnas.2201968119
- Hopp, H. (2015a). Individual differences in the second language processing of object-subject ambiguities. *Appl. Psycholinguist.* 36, 129–173. doi: 10.1017/S0142716413000180
- Hopp, H. (2015b). Semantics and morphosyntax in predictive L2 sentence processing. *Int. Rev. Appl. Linguist. Lang. Teach.* 53, 277–306. doi: 10.1515/iral-2015-0014
- Hopp, H. (2016). The timing of lexical and syntactic processes in second language sentence comprehension. *Appl. Psycholinguist.* 37, 1253–1280. doi: 10.1017/S0142716415000569
- Hubers, F., Redl, T., De Vos, H., Reinartz, L., and De Hoop, H. (2020). Processing prescriptively incorrect comparative particles: evidence from sentence-matching and eye-tracking. *Front. Psychol.* 11:186. doi: 10.3389/fpsyg.2020.00186
- Kaan, E. (2014). Predictive sentence processing in L2 and L1: What is different? *Linguist. Approaches Bilingual.* 4, 257–282. doi: 10.1075/lab.4.2.05kaa

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2026.1792437/full#supplementary-material>

- Kessler, R., Weber, A., and Friedrich, C. K. (2021). Activation of literal word meanings in idioms: evidence from eye-tracking and erp experiments. *Lang. Speech* 64, 594–624. doi: 10.1177/0023830920943625
- Kim, H., and Grüter, T. (2021). Predictive processing of implicit causality in a second language: a visual-world eye-tracking study. *Stud. Sec. Lang. Acquisit.* 43, 133–154. doi: 10.1017/S0272263120000443
- Kotz, S. A. (2009). A critical review of erp and fmri evidence on l2 syntactic processing. *Brain Lang.* 109, 68–74. doi: 10.1016/j.bandl.2008.06.002
- Kuiken, F., and Vedder, I. (2012). “Syntactic complexity, lexical variation and accuracy as a function of task complexity and proficiency level in L2 writing and speaking,” in *Dimensions of L2 Performance and Proficiency: Complexity, Accuracy and Fluency in SLA*, eds. H. Alex, V. Ineke, K. Folkert (Amsterdam: John Benjamins Publishing Company), 143–170. doi: 10.1075/llt.32.07kui
- Kuperberg, G. R., and Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Lang. Cogn. Neurosci.* 31, 32–59. doi: 10.1080/23273798.2015.1102299
- Kutas, M., and Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annu. Rev. Psychol.* 62, 621–647. doi: 10.1146/annurev.psych.093008.131123
- Kutas, M., and Hillyard, S. A. (1983). Event-related brain potentials to grammatical errors and semantic anomalies. *Mem. Cogn.* 11, 539–550. doi: 10.3758/BF03196991
- Li, N., Wang, S., Kornrumpf, F., Sommer, W., and Dimigen, O. (2024). Parafoveal and foveal n400 effects in natural reading: a timeline of semantic processing from fixation-related potentials. *Psychophysiology* 61:e14524. doi: 10.1111/psyp.14524
- Marzi, C., Boni, N., Todesco, A., Nadalini, A., Albertin, G., Dolciotti, C., et al. (2026). “Reading dynamics and comprehension in cognitive aging: a multimodal language resource,” in *Proceedings of the fifteenth international conference on language resources and evaluation (LREC 2026)* (Paris: ELRA), 2608–2618. doi: 10.63317/3wjy3a8cwnw8
- Marzi, C., Nadalini, A., Lento, A., Srivastava, M., Todesco, A., Pirrelli, V., et al. (2025). Oral text reading as a multi-sensory task. *Lingue Linguaggio* 24, 141–156. doi: 10.1418/117447
- Marzi, C., Rodella, A., Nadalini, A., Taxitari, L., and Pirrelli, V. (2020). “Does finger-tracking point to child reading strategies?,” in *Proceedings of 7th Italian Conference on Computational Linguistics*, Bologna, vol. 2769, eds. J. Monti, F. Dell’Orletta, and F. Tamburini (Aachen: CEUR proceedings). doi: 10.4000/books.aaccademia.8695
- Nadalini, A., Marzi, C., Ferro, M., Taxitari, L., Lento, A., Crepaldi, D., et al. (2023). Eye-voice and finger-voice spans in adults’ oral reading of connected texts: implications for reading research and assessment. *Ment. Lex.* 18, 366–400. doi: 10.1075/ml.00025.nad
- Nahatame, S. (2023). Predicting processing effort during L1 and L2 reading: the relationship between text linguistic features and eye movements. *Biling.: Lang. Cogn.* 26, 724–737. doi: 10.1017/S136672892200089X
- Osterhout, L. (1997). On the brain response to syntactic anomalies: manipulations of word position and word class reveal individual differences. *Brain Lang.* 59, 494–522. doi: 10.1006/brln.1997.1793
- Osterhout, L. (2015). “Event-related brain potentials as tools for comprehending language comprehension,” in *Perspectives on Sentence Processing*, (Hove: Psychology Press), eds. C. Jr. Clifton, L. Frazier, K. Rayner, C. Clifton (New York, NY: Psychology Press), 15–44.
- Pickering, M. J., and Clark, A. (2014). Getting ahead: forward models and their place in cognitive architecture. *Trends Cogn. Sci.* 18, 451–456. doi: 10.1016/j.tics.2014.05.006
- Pickering, M. J., and Gambi, C. (2018). Predicting while comprehending language: a theory and review. *Psychol. Bull.* 144:1002. doi: 10.1037/bul0000158
- Pickering, M. J., and Garrod, S. (2013). An integrated theory of language production and comprehension. *Behav. Brain Sci.* 36, 329–347. doi: 10.1017/S0140525X12001495
- R Core Team (2025). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available online at: <https://www.R-project.org/>
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychol. Bull.* 124:372. doi: 10.1037/0033-2909.124.3.372

- Rosenbach, A. (2005). Animacy versus weight as determinants of grammatical variation in English. *Language* 81, 613–644. doi: 10.1353/lan.2005.0149
- Rück, F., Dudschig, C., Mackenzie, I. G., Vogt, A., Leuthold, H., and Kaup, B. (2021). The role of predictability during negation processing in truth-value judgment tasks. *J. Psycholinguist. Res.* 50, 1437–1459. doi: 10.1007/s10936-021-09804-0
- Schlenter, J. (2023). Prediction in bilingual sentence processing: how prediction differs in a later learned language from a first language. *Biling.: Lang. Cogn.* 26, 253–267. doi: 10.1017/S1366728922000736
- Seyednozadi, Z., Pishghadam, R., and Pishghadam, M. (2021). Functional role of the N400 and P600 in language-related ERP studies with respect to semantic anomalies: an overview. *Arch. Neuropsychiatry* 58:249. doi: 10.29399/npa.27422
- Soares, A. P., Oliveira, H., Ferreira, M., Comesaña, M., Macedo, A. F., Ferrare, P., et al. (2019). Lexico-syntactic interactions during the processing of temporally ambiguous L2 relative clauses: an eye-tracking study with intermediate and advanced Portuguese-English bilinguals. *PLoS ONE* 14:e0216779. doi: 10.1371/journal.pone.0216779
- Song, J.-H., and Nakayama, K. (2009). Hidden cognitive states revealed in choice reaching tasks. *Trends Cogn. Sci.* 13, 360–366. doi: 10.1016/j.tics.2009.04.009
- Stafura, J. Z., and Perfetti, C. A. (2014). Word-to-text integration: Message level and lexical level influences in erps. *Neuropsychologia* 64, 41–53. doi: 10.1016/j.neuropsychologia.2014.09.012
- Staub, A. (2015). The effect of lexical predictability on eye movements in reading: critical review and theoretical interpretation. *Lang. Linguist. Compass* 9, 311–327. doi: 10.1111/lnc3.12151
- Staub, A., and Goddard, K. (2019). The role of preview validity in predictability and frequency effects on eye movements in reading. *J. Exp. Psychol.: Learn. Mem. Cogn.* 45:110. doi: 10.1037/xlm0000561
- Taxitari, L., Cappa, C., Ferro, M., Marzi, C., Nadalini, A., and Pirrelli, V. (2021). “Using mobile technology for reading assessment,” in *Proceedings of 6th IEEE Congress on Information Science Technology (IEEE CiST'20)* (Agadir) (New York, NY: IEEE). doi: 10.1109/CiST49399.2021.9357173
- Tiffin-Richards, S. P. (2024). Bilingual parafoveal processing: children and adults preprocess orthographic information of the upcoming word during sentence reading in their first and second language. *J. Exp. Psychol.: Learn. Mem. Cogn.* 50, 1844–1861. doi: 10.1037/xlm0001346
- Van Hell, J. G., and Tokowicz, N. (2010). Event-related brain potentials and second language learning: syntactic processing in late L2 learners at different L2 proficiency levels. *Sec. Lang. Res.* 26, 43–74. doi: 10.1177/0267658309337637
- Van Heuven, W. J., Mandera, P., Keuleers, E., and Brysbaert, M. (2014). SUBTLEX-UK: a new and improved word frequency database for British English. *Quart. J. Exp. Psychol.* 67, 1176–1190. doi: 10.1080/17470218.2013.850521
- Van Petten, C., and Luka, B. J. (2012). Prediction during language comprehension: benefits, costs, and ERP components. *Int. J. Psychophysiol.* 83, 176–190. doi: 10.1016/j.ijpsycho.2011.09.015
- Wang, W., Chen, L., and Perfetti, C. (2026). The late frontal positivity reflects incremental mental model updating: graded predictability effects during authentic text reading. *Brain Lang.* 272:105667. doi: 10.1016/j.bandl.2025.105667
- Zheng, X., and Lemhöfer, K. (2019). The “semantic p600” in second language processing: when syntax conflicts with semantics. *Neuropsychologia* 127, 131–147. doi: 10.1016/j.neuropsychologia.2019.02.010

Frontiers in Psychology

Paving the way for a greater understanding of human behavior

The most cited journal in its field, exploring psychological sciences - from clinical research to cognitive science, from imaging studies to human factors, and from animal cognition to social psychology.

Discover the latest Research Topics

[See more →](#)

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact

