frontiers | Research Topics

Language acquisition in diverse linguistic, social and cognitive circumstances, volume II

Edited by
Maria Garraffa, Gary Morgan, Theodoros Marinis
and Maria Teresa Guasti

Published in
Frontiers in Psychology
Frontiers in Communication

ISSN 1664-8714
ISBN 978-2-8325-3573-8
DOI 10.3389/978-2-8325-3573-8

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

Language acquisition in diverse linguistic, social and cognitive circumstances, volume II

Topic editors
Maria Garraffa — University of East Anglia, United Kingdom
Gary Morgan — Fundació per a la Universitat Oberta de Catalunya, Spain
Theodoros Marinis — University of Konstanz, Germany
Maria Teresa Guasti — University of Milano-Bicocca, Italy

Citation
Garraffa, M., Morgan, G., Marinis, T., Guasti, M. T., eds. (2023). *Language acquisition in diverse linguistic, social and cognitive circumstances, volume II*.
Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-3573-8

Table of contents

TYPE Editorial
PUBLISHED 13 September 2023
DOI 10.3389/fpsyg.2023.1282163

Check for updates

OPEN ACCESS

EDITED AND REVIEWED BY
Xiaolin Zhou,
Peking University, China

*CORRESPONDENCE
Maria Garraffa
m.garraffa@uea.ac.uk

RECEIVED 23 August 2023
ACCEPTED 04 September 2023
PUBLISHED 13 September 2023

CITATION
Garraffa M, Guasti MT, Marinis T and Morgan G (2023) Editorial: Language acquisition in diverse linguistic, social and cognitive circumstances, volume II. *Front. Psychol.* 14:1282163.
doi: 10.3389/fpsyg.2023.1282163

Editorial: Language acquisition in diverse linguistic, social and cognitive circumstances, volume II

Maria Garraffa[1]*, Maria Teresa Guasti[2], Theodoros Marinis[3] and Gary Morgan[4]

[1]Faculty of Medicine and Health Sciences, School of Health Sciences, University of East Anglia, Norwich, United Kingdom, [2]Department of Psychology, University of Milano-Bicocca, Milan, Italy, [3]Department of Linguistics, University of Konstanz, Konstanz, Germany, [4]School of Health and Psychological Sciences, City University of London, London, United Kingdom

KEYWORDS

language acquisition, attention deficit and hyperactivity disorder (ADHD), multilingualism, pragmatics, narrative abilities, developmental language disorder (DLD)

Editorial on the Research Topic
Language acquisition in diverse linguistic, social and cognitive circumstances, volume II

This volume is the second of the Research Topic "Language acquisition in diverse linguistic, social and cognitive circumstances," (Volume 1—Garraffa et al., 2018). The new volume presents research on multilingual speakers, including underrepresented languages in a set of different linguistics abilities in both typical and atypical populations. It presents nine contributions, divided into two sections: one section on language competence in children with neurodevelopmental disorders and a second one on language learning in a multilingual context.

Section 1: language competence in children with neurodevelopmental disorders

Studies on child language acquisition requires evidence that is representative of the typological language diversity present in the ∼7,000 or so languages spoken in the world. This is often not the case in research: a recent review of language acquisition studies found that we have research evidence on around 103 languages, representing ∼1.5% of the world's languages (Kidd and Garcia, 2022). More investigations in underrepresented languages are of particular importance in atypical language development, where data are scarce and there is a lack of resources for research. The relevance of testing underrepresented languages for atypical language development was addressed in Abu Bakar et al. on comprehension and production of Wh-sentences in children's speakers of Malay with and without developmental language disorder (DLD). The Malay speakers with DLD performed similarly to younger children with typical language development (TLD) and their language use reflected colloquial Malay with selective omissions of particles unlike age-matched children with TLD. Specific grammatical strategies are adopted by children with DLD in line with an immature acquisition of syntax, opening the debate of the need of qualitative and theoretical driven analyses of the production in children with DLD.

Research on bilingual speakers with a neurodevelopmental disorder is in great demand due to the rise of bilingualism worldwide and growing awareness of the communication needs in different populations (Garraffa et al., 2023). Phonological development in a group of bilingual speakers with Williams syndrome (WS) were investigated in a study on trajectories in late phonological development (Pérez et al.). Individual differences with a tendency for omissions, e.g., of final consonants, have been reported as specific to WS at all ages. The study brings together evidence for an appropriate phonological assessment and treatment for people with WS across the lifespan.

In a scoping review on possible effects of bilingualism on cognition and behavior in speakers with attention deficit hyperactivity disorder (ADHD), Köder et al. reported a lack of research and no consistent evidence in support of the hypothesis that bilingualism can mitigate or act as a barrier in speakers with ADHD. Two main outcomes are reported in the review: first, there should not be any concern about individuals with ADHD learning additional languages; second, researchers should be cautious in investigating this topic, as studies are very diverse in the methodology and their scope.

Section 2: language learning in a multilingual context

Learning new words is a common experience for a multilingual speaker that must develop two vocabularies. By studying the acquisition of cognates in trilingual speakers, Xue et al. reported a facilitatory effect from the first language (L1) on novel word learning and a strong interference of the second language (L2), suggesting that the multilingual experience can lead to different outcomes in word learning.

Pragmatic abilities in the speech of L2 speakers can be affected by contextual variables which may affect their oral production. By investigating oral speech act production in Chinese English as a foreign language (EFL) learners, Huang and Lu showed that some speech acts are more difficult than others and contextual variables may have distinct effects on different speech acts. These findings have important implications for the L2 classroom and can inform pedagogical approaches to teaching L2 pragmatics.

Focusing on reading in bilingual children and adults, Friesen et al. examined reading comprehension and reading strategies use. Bilingual adults reach comparable reading comprehension performance to monolingual children despite their lower vocabulary by using several strategies, e.g., inferencing. The study puts an emphasis on the role of effective strategies for reading comprehension in addition to language instruction.

In a study on narrative abilities in bilingual children's speakers of Urdu-Cantonese, Chan et al. investigated the relationship between macrostructure and microstructure between languages. Although the number of different words was consistently a positive predictor of the macrostructure's components in both languages, other microstructure measures showed mixed results, suggesting that the relationship between macrostructure and specific microstructural abilities can manifest similarly and differently between the L1 and the L2.

The grammatical skills of Dutch children with 22q11.2 Deletion Syndrome were investigated by Boerma et al. in comparison with children with DLD using spontaneous speech and standardized assessments, showing that although children with 22q11.2 had more deficits in their receptive language skills compared to children with DLD, there was an overlap between the two groups in their expressive language skills. This suggests that there are multiple routes to deficits in grammatical skills despite differences in etiology.

Farsi-English Heritage language speakers were investigated in the study by Komeili et al. that profiled vocabulary, morphosyntax, and narrative microstructure skills in both languages. The study demonstrated that on the vocabulary and narrative tasks participants were more dominant in English than in Farsi, while on grammar there were no significant differences between the two languages, supporting the importance of measuring language across multiple domains in studies of bilingual children.

Author contributions

MG: Writing—original draft, Writing—review and editing. MTG: Writing—review and editing. TM: Writing—original draft, Writing—review and editing. GM: Writing—review and editing.

Acknowledgments

We would like to thank all 37 authors and 22 reviewers and editors who offered their manuscripts and their constructive comments for this Research Topic.

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.

The author(s) 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.

Publisher's note

References

Garraffa, M., Guasti, M. T., Marinis, T., and Morgan, G. (2018). Editorial: Language acquisition in diverse linguistic, social and cognitive circumstances. *Front. Psychol.* 9, 1807. doi: 10.3389/fpsyg.2018.01807

Garraffa, M., Sorace, A., and Vender, M. (2023). *Bilingualism Matters: Language Learning Across the Lifespan*. Cambridge: Cambridge University Press. ISBN: 9781009333375.

Kidd, E., and Garcia, R. (2022). How diverse is child language acquisition research? First *Lang.* 42, 703–735. doi: 10.1177/01427237211066405

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 28 October 2022
DOI 10.3389/fpsyg.2022.948992

Check for updates

OPEN ACCESS

EDITED BY
Richard Wiese,
University of Marburg, Germany

REVIEWED BY
Likan Zhan,
Beijing Language and Culture University, China
Anna Cardinaletti,
Ca' Foscari University of Venice, Italy

*CORRESPONDENCE
Norsofiah Abu Bakar
abnorsofiah@usm.my

SPECIALTY SECTION
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

RECEIVED 20 May 2022
ACCEPTED 29 September 2022
PUBLISHED 28 October 2022

CITATION
Abu Bakar N, Smith G, Razak RA and Garraffa M (2022) The comprehension and production of Wh- questions among Malay children with developmental language disorders: Climbing the syntactic tree. *Front. Psychol.* 13:948992.
doi: 10.3389/fpsyg.2022.948992

The comprehension and production of Wh- questions among Malay children with developmental language disorders: Climbing the syntactic tree

Norsofiah Abu Bakar[1]*, Giuditta Smith[2], Rogayah A. Razak[3] and Maria Garraffa[2]

[1]School of Humanities, Universiti Sains Malaysia, Pulau Pinang, Malaysia, [2]Faculty of Medicine and Health Sciences, University of East Anglia, Norwich, United Kingdom, [3]Department of Education, Faculty of Social Science & Liberal Arts, UCSI University, Kuala Lumpur, Malaysia

This study is an investigation of both comprehension and production of Wh- questions in Malay-speaking children with a developmental language disorder (DLD). A total of 15 Malay children with DLD (ages 7;0–9;11 years) were tested on a set of Wh- questions (*who* subject and object, *which* subject and object), comparing their performance with two control groups [15 age-matched typically developing (TD) children and 15 younger TD language-matched children]. Malay children with DLD showed a clear asymmetry in comprehension of Wh- questions, with a selective impairment for which NP questions compared with who questions. Age-matched controls performed at ceiling in all Wh- questions, while the language-matched group reported a subject/object asymmetry selective for the which NP, as reported in other languages. In production, both children with DLD and younger children showed a preference for questions with *in situ* Wh- elements, a structure that is allowed in colloquial Malay, but which is not produced by the age-matched TD group. Several non-adult-like strategies were adopted particularly by the children with DLD to avoid complex sentences, including substitution with yes/no echo questions, production of the wrong Wh- question, and use of a generic Wh- element. The study provides an insight on the mastery of Wh- questions in both typical Malay children and children with DLD. Implications for the definition of a clinical marker for DLD in a free word order language with Wh- *in situ* option will be discussed.

KEYWORDS

Malay, DLD, Wh- questions, clinical markers, which-questions, subject/object asymmetry

Introduction

Children with developmental language disorder (DLD) can be defined as children with impairment in acquiring language components, often selective to a specific linguistic domain, such as syntax or phonology (Stark and Tallal, 1981; Leonard, 1998; Novogrodsky and Friedmann, 2009). As was proposed in the last consensus paper aiming at agreeing on the definition of DLD (Bishop et al., 2017), the language ability of children with DLD is not consistent with their age-group, and this is not attributed to factors external to the language system, such as hearing impairments, cognitive delays, or oral motor and neurological impairments, but rather to a specific impairment in the language system (Bishop, 2006, 2017). In the present paper, we present a study exploring the nature of the impairment in Malay children, particularly focusing on Wh-questions. The study was funded by the Research University Grant (GUP) (code UKM-GUP-2011-134) of the Universiti Kebangsaan Malaysia and the research grant PHUMANITI6315272 of the Universiti Sains Malaysia.

One of the issues in the study of the specific language profiles in children with DLD is that theories on language acquisition are often derived from considerations based on adult language and do not take into account the process of development, adopting a strict definition of correctness based on adult competence (Wexler, 1998, 2003). For example, in the domain of syntactic abilities, the debate on how grammar is acquired and which factors contribute to reaching grammatical competence is still ongoing in developmental linguistics, with few collaborations between linguists and developmental psychologists, indicating a need for more evidence across languages. Although the milestones of language acquisition have been uncovered to a good level of detail, thanks to crosslinguistic studies in different populations and the refinement of the underlying linguistic theory, it is still unclear whether there are interdependencies between acquisition of specific aspects of grammar and predicted stages of development and/or cognitive capacities ancillary to language (e.g., working memory or attention), with few studies tracking the development of a specific grammatical structure in both typical and atypical development (see Guasti, 2017 for an overview of the growth of grammar).

A phenomenon has been studied extensively in the acquisition of syntax is the acquisition of interrogative sentences (see Thornton, 2016 for an overview). Languages can vary on some specific syntactic properties, for example, allowing the presence of a Wh- element in its base position or fronting the element emphasizing its discourse relevance. For instance, the author reported non-adult-like productions of Wh- questions with referential which NP Wh- phrases in children aiming to avoid more complex configurations. Since Thornton's seminal paper on production of Wh- in children (Thornton, 1995), more subtle distinctions have been reported in the acquisition of interrogatives.

In a relevant cross-sectional study, De Vincenzi et al. (1999) looked at the development of Wh- questions in 3- to 11-year-old children focusing in particular on the comprehension of which NP questions in Italian. 352 typically developing (TD) children were presented with *who* and *which* subject/object reversible questions in a thematic assignment picture selection task where children were asked to answer questions of the *who/which did what to whom?* kind. While comprehension of subject questions was above chance already in the youngest group, comprehension of reversible object questions appeared to be delayed. Furthermore, a clear distinction between *who* and *which* questions emerged, with the latter being more delayed and systematically lower in performance than who object questions across all age-groups. This asymmetry was used as proof to rule out the hypothesis that the main problem in the acquisition of Wh- sentences with non-canonical word order, such as object questions, was an overall delay in any sentence type with non-canonical word order. Rather, a more fine-grained distinction of the syntactic factors at play was needed. Similar results were reported with English-speaking children, who showed above chance performance on *who* questions (both subject and object) but lower accuracy in both production and comprehension of *which* questions, and in particular object *which* questions (Yoshinaga, 1996; Avrutin, 2000; Hirsch and Hartman, 2006). The overall picture for English-speaking children is a clear subject/object asymmetry in *which* questions also reported for Hebrew children at the age of 4 years (Friedmann et al., 2009). Finally, in a recent study on comprehension of *who/which* questions in a group of 47 English children (mean age 5;2 years), it was reported that *which* questions do not correlate with general grammatical knowledge measured with a standard grammatical test (Bishop Dorothy, 2003; Bishop, 2006). These results advocate for a distinct grammatical process for *which* questions that needs to be acquired and is independent from other grammatical processes (Riches and Garraffa, 2017).

To better explore the delay reported in many languages on both *who* and *which* object questions, a study was conducted to explore the role of similarity between the arguments during thematic role assignment (Guasti et al., 2012). In the task, the number features of the two nouns were mismatched in a transitive reversible sentence; if the agent was plural, the patient was singular, and *vice versa*, as in the sentence "chi legano gli orsi?," *Who do the bears tie?*. The authors find that mismatch improves accuracy due to a transparent distinction between elements and the consequent lack of similarity effects.

Overall, the main finding across languages is a selective problem with both the subject and object *which* questions in acquisition, with several substitutions of the Wh- element and the emergence of strategies to avoid the full movement of the complex which NP element (see Belletti and Guasti, 2015 for comprehensive accounts; Thornton, 2016). What emerges from these studies is that children are following a trajectory in the acquisition of Wh- questions—and syntactic dependencies in general—starting from simpler subject questions (*who* subject) and then moving to more complex object extraction with NP restriction (*which* object). This trajectory should be considered in continuity with the development of the grammatical system and investigated within a set of predictions for the acquisition of each grammatical structure.

A model which fits the developmental stages in the acquisition of syntactic dependencies was recently proposed based on the notion of minimality between arguments (Friedmann et al., 2009). The model, originally developed to address the canonicity pattern reported in adults with language disorders (Garraffa and Grillo, 2008), suggested that young children have an immature grammatical system with poor production and comprehension of sentences with a moved object and intervening material, in particular when there is structural similarity between the two (Garraffa, 2017), as schematized in (1). According to the authors, children are more sensitive to effects of similarity between arguments, which leads them to adopt a restricted version of minimality: the model makes clear predictions that unlike adults, any representation like the one in (1a) and (1b) is equally perceived as a violation in children, with no full disjunction in the specification of the grammatical features of the arguments.

(1)
1a. + A... + A... < + A > (identity)
–UNGRAMMATICAL
1b. + A, + B... + A... < + A, + B > (inclusion)
–UNGRAMMATICAL FOR CHILDREN
1c. + A... + B... < + A > (disjunction)
–GRAMMATICAL FOR CHILDREN

The generalization that emerges is that if the target of the movement and the intervening subject argument are sufficiently different in their internal featural composition, the configuration is unproblematic (e.g., a Wh- question with one animate and one inanimate argument). The defining factor appears to be the presence or absence of a lexical NP restriction. This model assumes that the source of difficulties in children's grammatical development is based on a partial encoding of the grammatical information, not sufficient to parse sentences similar to (1b). Children therefore adhere to a stricter version of the locality principle, requiring distinct feature specifications for the target and for its intervener, and imposing a disjoint specification.

It is interesting to note that in these immature grammatical systems, an internal grammatical pressure of coping with the next level of the configuration can determine the production of sentences that are severely dispreferred in adults and not attested in the standard varieties. This is the case, for example, for children's production of passives in Italian, which are substituted with a set of forms that are not attested in adult varieties (see Belletti, 2017 for details) or the case of the non-adult-like *which* questions with extra copies of the movement of the Wh- element in English (as in Thornton, 1995).

Wh- questions among children with developmental language disorder

The comprehension and production of Wh- questions have been extensively investigated in typical language acquisition (Thornton, 2016). Research shows that the ability of children with DLD to produce interrogatives is not consistent with that of their age-matched peers (Levy and Friedmann, 2009). A few structural aspects of Wh- questions were examined in the research on DLD, including word order (canonical vs. non-canonical), the difference between moved Wh- and *in situ* Wh-, and the difference between Wh- argument and Wh- adjunct. The inclination of many researchers is to study comprehension of Wh- questions in terms of subject Wh- vs. object Wh- questions, exploring the canonicity pattern and the factors underlying the discrepancy in comprehension between the two structures (Ebbels and van der Lely, 2001; Deevy and Leonard, 2004; Friedmann and Novogrodsky, 2011).

Based on the word order of sentences, it was found that children with DLD found acquiring specific Wh- questions particularly challenging. Ebbels and van der Lely (2001) studied four English-speaking children with DLD (aged 11–13 years) that showed that even after a language intervention program, they were still unable to comprehend questions *which* (object) compared with the structure of Wh- subject questions *who*, *what*, and *which* (subject). Interestingly, studies by Wong et al. (2004) and Friedmann and Novogrodsky (2011) on Hebrew- and Mandarin-speaking children with DLD, respectively, supported this finding. This seemed to suggest that the structure of *which* (object) is a structure that is difficult across languages, an intuition which is theoretically supported by the concept of movement, as was proposed in some relevant accounts of the phenomenon. One such account is provided in the study by Friedmann and Novogrodsky (2011) who suggested that movement is responsible for the difficulties faced by children with DLD. The feature checking requirements of Wh- questions initiate the movement of an element (the Wh- element itself) crossing over the subject position to reach a higher position, as sketched in (1). This operation, which creates a dependency between the moved element and its trace that is interrupted by the subject, is understood to be difficult in children in intervention accounts.

The syntactic difficulty faced by children with DLD has been described, for example, by the deficit in computational grammatical complexity (DCGC) theory (Deevy and Leonard, 2004; Marinis and van der Lely, 2007). This theory presupposes that this difficulty is caused by the generation of highly complex sentences that involve movement utilizing various cycles of derivations (van der Lely and Stollwerck, 1997). The fact that object questions have been identified as complex is supported by studies such as van der Lely and Battell (2003) on English-speaking children, who found children with DLD to consider movement application as an optional phenomenon. In terms of produced structures, questions produced by children with DLD are not grammatical as verified by van der Lely and Battell's (2003) study. Hamann's (2006) study on Wh- questions among French-speaking children with DLD showed that these children did not produce Wh- fronting questions. Instead, they produced Wh- *in situ* questions. These studies describe an atypical development of Wh- dependencies, potentially due to an immature system. The pattern of produced sentences by children with DLD is of sentences that are more derivationally economical, avoiding any movement of elements on the left positions and echoing the order of the declarative sentence. Hamann's (2006) findings were supported by Hansson and Nettelbladt's (2006) findings who also found that Swedish-speaking children with DLD produced sentences that can be described as more economical due to the avoiding of fronting the Wh- element. Similarly, Jakubowicz (2011), eliciting different Wh- dependencies in French, shows that long-distance dependencies are avoided by both children with DLD and TD children, but children with DLD and younger TD children in particular resort to ungrammatical structures when a long Wh- dependency is elicited.

As discussed in this session, the structure of non-canonical Wh- questions derived through movement is a difficult structure for children with DLD, and a potential explanation for this difficulty is based on the presence of an intervener (the subject), which causes an effect of similarity between arguments. In fact, this difficulty toward which object structures is universal in nature as it is found across many languages.

Wh- questions in Malay

Malay has both Wh- *in situ* questions and Wh- questions with movement (Kader, 1981; Salleh, 1989; Razak, 2003). The moved Wh- form is the grammatical form used in the standard Malay (SM) variety, particularly in the written form. In this variety, the Wh- word is fronted, and the interrogative affix -*kah* is attached to the questioned constituent, and the relative particle *yang* is present. The colloquial Malay (CM) variant has both the *in situ* and the moved Wh- forms, with the former being the most common. In the moved form, the interrogative affix -*kah* is absent, but the relative particle *yang* is present. **Table 1** provides a full list of examples for each condition in the two varieties.

Wh- *in situ* questions are questions in which the Wh- word constituent appears in the base position in the sentence, and the sentence conforms to the SVO word order. In the Wh- *in situ*, the Wh- word is in the base position and a raising intonation marks the sentence as an interrogative. In CM, for both *in situ* and moved options, the grammatical interrogative particle *kah* is not required. On the other hand, in SM, the moved Wh- questions comprise sentences that are generated from the base position and then undergo movement to the specifier position of the CP, and the specification of the relative particle *kah* on the Wh- element is obligatory (Razak, 2003). Affix *kah* is an overt morphological marker of interrogation in the specifier position, and it agrees with the particle *yang*. The *yang* construction in Malay is found in both SM and CM. Generally, it functions as a *yang*-type restrictive relative clause headed by *yang* (REL). Its function is to modify the head noun in a complex NP construction. Other functions of *yang* include *yang* as a deictic marker in a focused construction ("Yang tu kuat"/That one is strong) and *yang* as a complementizer with a [+ Q] feature (Malay also having *bahawa* [-Q], and the null complementizer [C ø]) (Wong, 2008). In Wh- questions as exemplified in **Table 1**, *yang* can be interpreted as a [+ Q] complementizer.

According to Wong (2008), in Wh- question formation, any argument in a position lower than that of a subject has to be passivized to become a derived subject before the extraction can occur. The Wh- phrase moves to the specifier position of an obligatory interrogative *yang*. The specifier, being an argument position, does not allow extraction from a position other than the highest subject position, a derived subject. Subjects in embedded or subordinate clauses can be questioned, provided the matrix verb is passivized as in (2) (example from Wong, 2008).

(2)	Siapakah	yang	dikatakan (oleh)	
	Who-PRT-Q	that	PASS-say (by)	
	John	akan	membeli	buku itu?
	John	will	ACT-buy	book the

"*Who did John say will buy the book?*"

In the context of the Malay language, studies on the ability of children with DLD with Wh- questions are limited compared with studies on Wh- questions of TD children (Aman, 2007, 2014; Kader and Tan, 2022). Studies on the language acquisition of TD Malay children confirm that there exist specific stages in the linguistic development of children. Long (1993) discovered that the majority of Malay children aged 5 and 6 years can understand and use verb and noun affixes in their school and home settings. Affixation is a pervasive morphological process in Malay and a prerequisite to produce standard Malay Wh- sentences. Importantly, an acquisitional trajectory was identified in the process, where older children master affixation more than younger children. In a longitudinal study of the spontaneous speech of five Malay children between the ages of 18 and 48 months collected weekly over a period of 2 years and 6 months (Tan, 1999; see Razak, 2013 for

TABLE 1 Examples of Wh- interrogative sentences for standard Malay and colloquial Malay.

Structures	Standard Malay	Colloquial Malay	
	Fronted	**Fronted**	***In situ***
Who subject	Siapakah yang menangis? Who-PRT-Q that ACT-cry "Who was crying?"	Siapa yang menangis? Who that ACT-cry "Who cried?"	Siapa menangis? Who ACT-cry "Who cried?"
Who object	Siapakah yang kanak-kanak tarik? Who-PRT-Q that child pull "Who was the child pulling?"	Siapa yang kanak-kanak tarik? Who that child pull "Who did the child pull?"	Kanak-kanak tarik siapa? Child pull who "Who did the child pull?"
Which subject	Budak lelaki yang manakah menangis? Child male that which-PRT-Q ACT-cry "Which boy was crying?"	Budak lelaki yang mana menangis? Child male that which ACT-cry "Which boy cried?"	Budak lelaki mana menangis? Child boy which ACT-cry "Which boy cried?"
Which object	Budak manakah yang dia pilih? Child which-PRT-Q that he choose "Which boy did he choose?"	Budak mana yang dia pilih? Child which that he choose "Which boy did he choose?"	Dia pilih budak yang mana? He choose child that which "Which boy did he choose?"

an overview), it was determined that the first Wh- words to appear are *mana*/where, *apa*/what, and *siapa*/who for children between 26 and 30 months. These are followed by *kenapa*/why (at 31 months) and *macam mana*/how (at 34 months). The last Wh- word to be acquired and rarely used is *berapa*/how much (35–36 months). Later, combined Wh- words such as *preposition + wh- word* and *wh- word + particle* appear, as in *dekat siapa*/near whom, *dengan siapa*/with whom, *dengan mana*/which one, *untuk apa*/for what, *macam ap*a/like what, *kat mana*/where at, and *macam mana*/how.

Another study collected naturalistic data from two Malay-speaking children around the age of 3 over a period of 3 months (Aman, 2007). The main finding of the study was the presence of both moved and *in situ* questions in the speech of both parents and children. However, an asymmetry between arguments and adjuncts was reported, with a preference for the *in situ* structure for the arguments in both parents and children. For particular adjunct questions (*how* and *why*), there was a strong tendency to select the moved question structure. This asymmetry between *in situ* arguments and moved adjuncts was reported in both short and long questions. A proposal to explain the asymmetry in Malay children is that it relies on *in situ*, rather than displaced, constructions to produce questions as a strategy to avoid any non-local dependencies (Cole and Hermon, 1998). A follow-up step in their grammatical development will be to attach all obligatory elements to the verb, thus licensing Wh- elements in the left periphery of the clause.

According to the model proposed by Aman (2007), TD children acquiring Malay will first make use of the *in situ* strategy and subsequently acquire a new grammatical operation. This operation is the generation of a gap without the need to reconstruct the Wh- element. It is possible that acquiring this mechanism is hard for children with DLD, who will prefer to stick with a simpler available version, compatible with CM. If this is the case, namely, if children with DLD do not fully acquire gap constructs for Wh- questions, then they will struggle to understand and produce certain Wh- questions, particularly NP restricted Wh- questions (e.g., *Siapakah yang membaca buku itu*? Who reads the book?) due to the necessary specification of an operator in the CP domain. In this account, TD children will go on to acquire more complex operations that they will be able help in the comprehension and production of Wh- constructions.

An interesting matter to explore is the reason behind this lack of progress in the grammatical development of children with DLD, assuming similar language exposure between DLD and TD children and the impact of the educational system on the grammar. The standard variant of Malay is part of the curriculum taught in primary schools, including the introduction of more complex sentences such as focused questions, as in *Buku yang Mary beli* ("It was a book that Mary bought").

Current study

The present study aims to expand on previous findings on the acquisition of grammar in Malay children with DLD looking at syntactic abilities in both comprehension and production of Wh- questions. It investigates the abilities of children on different Wh- questions, aiming to explore whether there are differences among them and to record the strategies in place to overcome more complex structures.

Materials and methods

Participants

The study sample comprised three groups of children speakers of Malay as a dominant language and attending public government schools in Malaysia: one experimental group and two control groups. The DLD group comprised 15 children with a diagnosis of DLD (12 boys and three girls; age range 7;0–9;11 years); the control group matched by chronological age (CA) comprised 15 children (12 boys and three girls) with typical language; and the second control group was matched by language abilities (LA) and consisted of 15 children (age range 4;0–6;11 years). CA participants were matched by age to DLD on a one-to-one basis (± 2 months), and LA participants were matched on performance on a linguistic assessment. Subjects from both control groups had normal hearing, as reported by their parents.

The 15 children with DLD were recruited from a pool of students who obtained C, D, and E grades in their Malay language subject in the year-end school examinations. They failed the national LINUS examination, which screens students in year 1 for the 3Rs—reading, writing, and arithmetic in addition to reasoning, and were placed in remedial classes (Luyee et al., 2015). There was an initial total of 26 subjects recruited; however, four students did not meet the normal score of Raven's Colored Matrices Test, and seven students failed to obtain consent from their parents/caregivers to participate in this study. All children were clinically diagnosed with language impairment and were receiving treatment at the time of testing. Their status was confirmed using a battery of baseline tasks that assessed the children's non-verbal and verbal abilities. Raven's Colored Progressive Matrices Test (Raven et al., 1998) was used to measure the children's non-verbal abilities, which were within the norm. The subjects were screened by an audiologist, and they had normal hearing (not exceeding 25 dB), and from an SLT through an oro-motor assessment that determined there were no articulatory conditions interfering with language. The screening for language included the Malay Preschool Language Assessment Tool (MPLAT, Razak et al., 2018), the sentence repetition task, and school grades in the Malay language subject. The MPLAT assessment is a standardized tool that has normative data of 510 Malay children aged 4;0–6;11 years. It tests both receptive and expressive language and early literacy skills of Malay preschool children. **Table 2** is a summary of the linguistics components included in the MPLAT.

TABLE 2 Linguistics components included in the MPLAT screening test (Razak et al., 2018).

Dimension	Modality	Task
Morphology	Comprehension	Picture vocabulary Sentence comprehension
Syntax	Repetition	Sentence repetition
Semantics	Comprehension/ production	Referential meaning Relational meaning
Early literacy skills	Reading/writing	Awareness of alphabets, alphabet-sound correspondence, copying, spelling skills

Results from the MPLAT were used to determine the language-matched group. Children with DLD obtained a score of −2 SD below the average standard score for their age-group, and they were thus matched with children belonging to the age-group whose scores were similar to those of the participants with DLD, as shown in **Table 3**. A *t*-test for independent sample confirmed there is no significant difference between the children with DLD and the language-matched control group ($p = 1.81$). These results showed that the communicative ability of children with DLD lies within the ability range of preschool children (LA group).

Task materials

A total of two tasks were adapted to Malay in order to assess comprehension and production of Wh- questions, one from Friedmann and Novogrodsky (2011) and one from Jakubowicz (2011).

Sentence comprehension

Sentence comprehension was explored with a sentence to picture matching task, targeting arguments in subject and object positions. The task was composed of 40 items: 10 *Siapa/who* subject questions, 10 *yang mana/which one* subject questions, 10 *Siapa/who* object questions, and 10 *yang mana/which one* object questions. Examples of the four structures are provided in (3)–(6).

(3) *Siapa* in subject:
Siapa cium adik?
Who kiss little sister/brother
"Who kissed little sister/brother?"

(4) *Yang mana* in subject:
Nenek yang mana cium adik?
Grandmother which kiss little sister/brother
"Which grandmother kissed little sister/brother?"

(5) *Siapa* in object:
Siapa yang adik cium?
Who that little sister/brother kiss
"Who did little sister/brother kiss?"

(6) *Yang mana* in object:
Nenek yang mana adik cium?
Grandmother which little sister/brother kiss
"Which grandmother that little sister/brother kissed?"

All items in the comprehension tasks were moved Wh-sentences in line with the grammar of SM, but with the absence

TABLE 3 Demographic details and language scores on the MPLAT components and school grades for Malay language for the three groups.

	DLD (SD)	Age-matched (SD)	Language–matched (SD)
Age (SD)	9;7 (1.79)	9;67 (1.79)	5;73 (1.75)
MPLAT scores			
MPLAT overall scores	72.53 (11.60)	100.00 (0.00)	83.69 (29.4)
MPLAT receptive language	45.3 (7.5)	100.00 (0.00)	50.2 (14.5)
MPLAT expressive language	27.2 (8.05)	100.00 (0.00)	33.49 (19.4)
School language scores			
Grammar score (School grade)	29.43 (13.93)	78.21 (8.61)	61.7 (16.78)
Composition score (School grade)	26 (9.08)	71.36 (6.41)	NA

of the *-kah* interrogative particle. In the sentence–picture matching task, the children listened to the recorded sentence and were asked to point to the two pictures that matched the sentence. The stimuli for the tasks included a picture set and an audio recording (see **Figure 1**).

Elicitation of Wh- questions

In this task, children are instructed to ask questions to a puppet with the appearance of a cartoon character. A total of 40 items were provided to prompt the child to ask Wh- questions with four different configurations, namely, *who subject* (10 items), *who object* (10 items), *which subject* (10 items), and *which object* (10 objects) questions.

To elicit the production of a Wh- element, part of the picture stimulus was hidden. The child is instructed to ask the puppet "Angry Bird" about the hidden information, as exemplified in (7) for *Siapa*/who subject question.

(7) Elicitation of a *Siapa*/who subject question:

Preamble: Itik sedang makan. Kita tak tahu nama orang yang beri itik makan. Cuba adik tanya *angry bird.*

The duck is eating. We do not know the name of the person who is feeding the duck. Please ask Angry Bird who.

Expected answers:

(7a) Moved Wh:

Siapa	yang	beri	itik	makan?
Who	*that*	*give*	*duck*	*eat*

"*Who is feeding the duck?*"

(7b) *In situ* Wh:

Yang	beri	itik	makan	siapa
That	*give*	*duck*	*eat*	*who*

"*Who is that (person) who gave the duck food?*"

An example of the elicitation material is offered in **Figure 2**.

Scoring

The scoring procedures followed the scoring method used by Aman (2007). In the comprehension task, a score of 1 was given if the children's answers matched the target pictures and a score of 0 if the children's answers did not match the target pictures. For the production task, a score of 1 was given if the children's responses matched the situations given. Substitution of nouns/personal pronouns (e.g., *ibu/mother* is replaced by *kakak/older sister*), use of contracted forms (e.g., *tidak*/NEG to *tak*), and deletion of the open syllable (determiner *ini/this* to *ni*) were still considered correct if the structure of the question matched the elicited question. A score of 0 was given if the children's answers resulted in a change in the original structure to another syntactic structure, sentences that change the target sentence's meaning. Because elicited contexts were felicitous with both a moved Wh- element and an *in situ* construction, all felicitous answers were further analyzed for the type of answer provided. These were "movement" and "*in situ.*"

In terms of the qualitative analysis, errors committed by children in the production task were transcribed and divided into structural and lexical errors. Structural errors encoded errors in the omission of Wh- questions, order of sentences that differed from the target sentences, incorrect usage of the Wh- elements, and ungrammatical sentences. Lexical errors encoded errors in the addition, omission, or substitution of lexical items.

Reliability

A second speech and language therapist native speaker of Malay transcribed productions from two children. The reliability of the transcription was measured by using a formula that calculates the percentage agreement for verbatim transcriptions. The results showed that the reliability between assessors was around 93%. Scoring reliability was also enforced using the test–retest method on five children from the entire subject population.

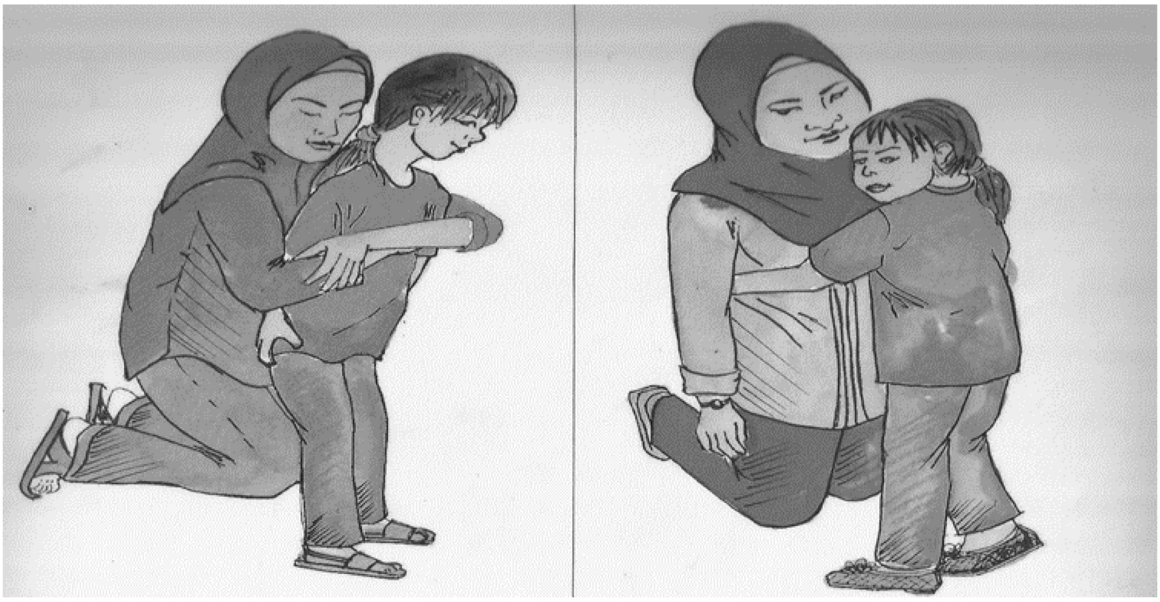


FIGURE 1
Picture pair used for the sentence "*Emak mana yang adik peluk?*" (*Which mother is little daughter hugging?*).

Results

Table 4 presents accuracy results across tasks (comprehension and production). Inferential statistics were run on R Studio (R Studio Team, 2022) and Jamovi (The Jamovi Project, 2021) and repeated measures ANOVA were implemented.

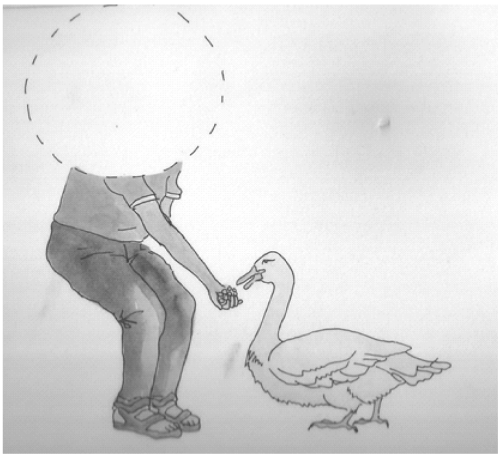


FIGURE 2
An example for elicitation of the Wh- sentences *Siapa yang beri itik makan?* (*Who is feading the duck?*).

Comprehension

Table 5 presents accuracy in the comprehension of the four Wh- questions tested (who subject/who object and which subject/which object) in the three groups.

A repeated measures ANOVA was run to compare the effect of group and condition (who subject, who object, which subject, which object). There was a significant difference in score between groups [$F_{(2,42)} = 17.1$, $p < 0.001$], a significant difference between conditions [$F_{(3,103)} = 26.61$, $p < 0.001$], and a significant interaction between groups and conditions [$F_{(6,103)} = 4.77$, $p < 0.001$]. Tukey's *post-hoc* comparisons revealed there was a significant difference between the group of children with DLD and age-matched controls ($p < 0.001$) but not between children with DLD and language-matched group ($p = 0.78$). The only operator to be significantly different from all others is *yang mana* (which object) ($p < 0.001$). In terms of the interactions between groups and conditions, *post-hoc* comparisons reveal that children with DLD are significantly worse than age-matched controls only in the *yang mana* (which object) condition ($p < 0.001$) and not on all others, while they are not statistically different from the language-matched group in any of the conditions.

Production

A repeated measures ANOVA was run to determine whether there was a significant effect of group and condition (viz., type of operator: why/who/where/what) on accuracy in production. There was a significant effect of group [$F_{(2,42)} = 26.4$, $p < 0.001$], but not of condition [$F_{(3,126)} = 1.71$, $p = 0.16$], or their interaction [$F_{(6,126)} = 1.61$, $p = 0.15$]. Tukey's *post-hoc* comparisons run for group determined that significant differences appear between the DLD group and the age-matched group ($p < 0.001$), but not the language-matched group ($p = 0.82$).

A second analysis was run on the target answers produced in the elicitation task to check whether group determined differences in the type of answer selected across all conditions, namely, *in situ* or movement. As described in the introduction, both options are grammatical in CM, although only movement structures are grammatical in SM.

Table 6 shows results across all Wh- elements in the three groups. A repeated measures ANOVA was run to determine the effects of group on the type of answer selected. The results of the ANOVA show a significant effect of the interaction between groups and types of answer [$F_{(2,1984)} = 82.2$, $p < 0.001$]. Tukey's *post-hoc* comparisons revealed that children with DLD are significantly different from CA both in the selection of *in situ* ($p < 0.001$), which is selected 1.7% of the time by CA and 27% of the time by DLD, and movement ($p < 0.001$), which is selected 96% of the time by CA and 18% of the time by DLD. No differences are reported between children with DLD and LA children.

Error analysis

Children's errors in producing the utterances were analyzed and are reported in Table 7. Errors committed by the subjects were grouped into two categories, namely, structural and lexical errors. The total number of errors committed by children with DLD and the language-matched group is comparable for all sentence structures. The two groups also share the main error types, that is, substitution with echo questions (declaratives with interrogative intonation) and wrong use of Wh- elements, whereas this type is not reported in the age-matched group.

Children with DLD and language-matched children tended to substitute Wh- questions with echo questions, declarative sentences with no Wh- element, and the insertion of the NP. An example of substitution with an echo question is reported in (8).

(8) Target sentence
Abang makan nasi kat mana?
Brother eat rice at where
"Where did brother eat rice?"

Subject's Response
Abang makan nasi kat dapur?
Brother eat rice at kitchen
"Brother ate rice in the kitchen?"
(DLD SB: K, 5;2).

TABLE 4 Comprehension and production overall performance across the three groups.

Group	Comprehension %	Production %
DLD	85.5	46
Age-matched	99.2	98.3
Language-matched	87.2	50.7

The second most frequent error produced by both language-matched children and children with DLD was the use of the wrong Wh- word, considering the context given and the expected targeted Wh- word and adopting a generic mana ("where") element.

(9) Target response
Kat mana ayah pasang khemah?
At where father set uptent
"Where did father set up the tent?"

Subject's Response
Kenapa ayah pasang khemah?
Why father set up tent
"Why father set up the tent?"
(DLD: MI, 8;7).

A second error classification on lexical errors is proposed in Table 8. These are errors are not apparently targeting a grammatical property and mainly targeting the knowledge of the verbs. It is interesting to note that the age-matched group did not produce any lexical error, showing a fully-fledged mastery of the verbal domain.

The most frequent type of lexical errors was the substitution of verbs with another semantically related form. An example of a lexical error of verb substitution is seen in (10). The targeted *belajar*, "to study" was substituted with the verb *mengajar*, "to teach."

(10) Target response
Abang belajar dengan siapa?
Elder brother study with whom
"*With whom did elder brother study?*"

Subject's response
Abang mengajar dengan siapa?
Big brother Aff-teach with who
"*Who did elder brother teach?*"
(DLD: 8:8 years old).

TABLE 5 Comprehension of Wh- questions (standard deviation) for the three groups.

Sentence type	DLD (n = 15)	Age-matched (n = 15)	Language-matched (n = 15)
Siapa (Who subject)	94 (0.88)	100 (0.00)	96.67 (1.02)
Yang mana (Which subject)	90 (0.89)	99.3 (0.25)	94 (1.09)
Siapa (Who object)	88 (1.09)	99.3 (0.25)	90 (1.26)
Yang mana (Which object)	70 (1.98)	97.3 (0.57)	74 (1.96)

TABLE 6 Answer types provided across all conditions in Wh- question elicitation for the three groups.

Sentence type	DLD (n = 15)	Age-matched (n = 15)	Language-matched (n = 15)
In situ	27 (6.6)	1.7 (0.94)	25 (7.63)
Movement	18 (6.19)	96.5 (1.4)	27.5 (9.93)
Non-target	54.3 (8.34)	1.8 (1.06)	47.3 (12.12)

TABLE 7 Types and occurrences of structural errors in Wh- question productions for the three groups.

Types of errors	DLD	Age-matched	Language-matched
Substitution with Eco Questions	17	1	11
Wrong use of Wh-	9	1	15
Incorrect Wh- movement	2	0	0
Verb omission	3	0	2
Insertion of *yang*	1	0	2

TABLE 8 Lexical errors across the structure of Wh- questions.

Types of errors	*In situ*			Movement		
	DLD	Age-matched	Language-matched	DLD	Age-matched	Language-matched
Verb substitution	7	0	0	1	0	2
Preposition substitution	2	0	1	3	0	8
Lexical additions	1	0	2	0	0	2

TABLE 9 Sensitivity, specificity, and accuracy values for Wh- structures.

Sentence structure	Sensitivity	Specificity	Accuracy
Comprehension *siapa*/ Who (subject)	53% (8/15)	100% (15/15)	77% (23/30)
Comprehension *yang mana*/Which one (subject)	67% (10/15)	87% (13/15)	77% (23/30)
Comprehension *yang mana*/ Which one (object)	93% (14/15)	100% (15/15)	7% (29/30)
Production *-wh*	87% (13/15)	93% (14/15)	90% (27/30)

Malay Wh- questions as clinical markers

To explore the sensitivity and specificity for Wh- questions, a comparison between the performance of children with DLD and their peers, the CA group, was conducted. A value of 2 standard deviations below the mean value of the CA group was used as suggested in the literature (Conti-Ramsden et al., 2001; Paul et al., 2012). The calculation of the sensitivity and specificity values used 80% as minimum value and 90% and above as good/excellent for clinical markers (see Bortolini et al., 2006 for more information). A set of proposed clinical markers for Wh- questions, defined as the elements that could characterized the DLD profile, are presented in **Table 9** in terms of the sensitivity, specificity, and accuracy values for difficult Wh- structures for Malay children with DLD.

The results showed that for the comprehension of Wh- sentences, the Wh-*yang mana*/which one (object), has the best sensitivity value (93%) and excellent specificity value (100%). This supports the finding that Wh- sentences with *yang mana/which* one are the most difficult for children with DLD and a potential candidate to be investigated in further studies.

Discussion

The study reported a set of data on Wh- sentence production and comprehension in a group of children native speakers of Malay. In this article, two main findings have been reported when comparing the performance of typical children and children with DLD, namely, a selective deficit for comprehension of which questions and a clear asymmetry in the production for children with DLD.

Where they produced *in situ* questions, age-matched children prefer to move the Wh- element at the root of the sentences and create a filler–gap relation as required in standard Malay. More interesting from a developmental point of view is the convergence of results between children with DLD and the younger group of language-matched children, making a strong case for a delay in the language development of DLD compatible with a pre-stage of language development. Malay younger children and children with DLD seemed to adhere to a similar timetable, and they have not developed structures that can be described as dependent or late acquired, for example, long-distance dependencies with lexical restricted items.

From a theoretical point of view, the data on interrogatives discussed in this article can be interpreted as an instance of grammatical reduction of the formal features necessary to activate the upper part of the syntactic tree. The outcome of this specific reduction could be a structure at play in both younger children and children with DLD, truncated in Rizzi's sense as shown in **Figure 3** (Rizzi, 1993/1994). This reduced structure allows the activation of the left periphery of the clause with base-generated placeholders, but it is not rich enough to license the movement on which NP restricted elements, favoring a lower *in situ* position for these elements.

The results are evidence collected from the literature for operators, such as which NPs, and need to be licensed as DPs to be permitted in the upper part of the syntactic tree. This can be formally represented, for example, as pied-piper features that have to be included in the derivation of the structure during its numeration and not later (see Watanabe, 2006 for a more detailed explanation). If the pied-piper features are not placed as a result, for example, of an underrepresented structure, the computation cannot proceed further, and effect like minimality but based on a minimal logical form (LF) representation can be

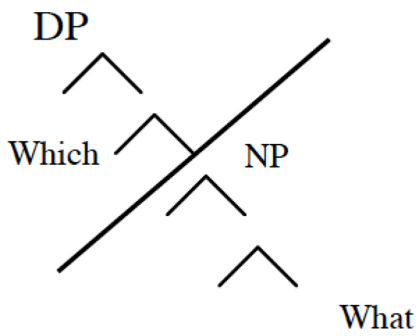


FIGURE 3
A truncated structure for acquisition of Wh- questions.

the cause of an underspecified/reduced representation. In the case of children acquiring Malay, it is possible that a principle like "avoid pronoun" or a similar principle of structural economy is a play with the effect of minimal pied piping at LF representation, placing the reduction of the syntactic tree at the interface between syntax and semantic. This is also supported by the qualitative analyses of the errors with a preference for a selection of generic operators, echo questions, and, in general, a selective impairment of critical features to allow which-X operator to be represented as DPs[1].

Overall, the results of our research show an asymmetry between comprehension and production in both atypical and early typical language acquisition. However, a detailed analysis of the linguistic strategies adopted to carry out the production task, which allows to generate alternatives to express information in a way that is more in line with the person's grammatical knowledge, shows differences in the language systems of the participants. All the strategies adopted aim at avoiding the more complex syntactic computation involved in long syntactic dependencies of more complex operators. Crucially, resorting to this tactic means that the syntactic structure present in SM is not a strong option in the language systems of DLD and young children, as exemplified by the difficulty found in the production of any Wh- moved question. We would like to argue that the data reported support the idea that the movement strategy, while present to some extent, is

1 In natural languages interesting linguistic evidence related to the syntax-semantic interface have been presented in support of a positional distinction of Wh operators based on semantic properties of definiteness restriction (Heim, 1987). In her investigation on Wh-traces and definite variables Heim presents data related to "definiteness" of the moved wh-phrase, (example below). (11) a. ?? Which one of the two men was there in the room/*drunk? b. ?? Which actors were there in the room/*laughing? c. ? Who was there in the room when you got home? d. What is there in Austin? A sort of definiteness hierarchy seems related to the possibility of extraction of a Wh operator in "there is" constructions, with "What" the less specified operator and a licit extractable element. This fact could allow us to describe the non-standard linguistic competence of our children, with all extractions are treated as the case in d, the one with the less specified operator.

not mastered in both children with DLD and younger kids. An asymmetry was found where both control children and participants with DLD comprehend all question types with movement, suggesting that they have acquired the computation involved in movement of restricted elements. However, a difference emerges between conditions, where one type of moved question is selectively impaired, namely, which object questions. Because movement as an operation is present in the abstract representations of these children, the issue with this specific occurrence of it is finer grained. A suggested interpretation will be given shortly.

If comprehension of Wh- questions with movement is generally mastered, younger controls and participants with DLD struggle with their production. In fact, when asked to produce questions, both groups produce non-target answers about half of the time, and they do not show a preference for either *in situ* of moved structures when they produce target answers, regardless of the elicited operator (which/who/where/when), in clear contrast with older control children who overwhelmingly prefer the production of moved structures. This point is crucial to underline the importance of testing children in more than one modality.

Considering all these pieces of evidence, the discrepancy between comprehension and production in a parallel testing ground is a fruitful method for evaluating grammatical knowledge. Where linguistic development is grammatically consistent, linear heuristics are not adopted, and implementation of a syntactic algorithm is preferred to the use of a "good enough" extralinguistic strategy. Findings indicate that both younger children and children with DLD face difficulties in production compared with comprehension, as previously reported for other languages (Contemori and Garraffa, 2010). But a modality explanation does not cover the more detailed pattern of errors visible in production, with children with DLD facing a delay in the acquisition of discourse-linked questions. This result is consistent with the findings of Ebbels and van der Lely (2001) and Friedmann and Novogrodsky (2011), supporting the idea of a delay in the acquisition of selective instances of Wh- movement. This matter can be proven for the more complex extractions of *which* questions as in *kakak yang mana* which functions as object in (12).

(12) [kakak yang mana]$_i$ adik kejar t_i?
↑___________________|

There are two options to explain the selective deficit with which object questions for this group: it can be described as a consequence of a reduction of the featural representation, leading to more intervention errors due to the lexical restrictions on which questions, or of a more structural reduction related to a truncated structure (as in **Figure 3**), where the edge of the syntactic tree is omitted in younger children and children with DLD. The production data in the present study support the second model, showing the relative absence of movement of Wh- elements to the left periphery in DLD and younger children but not older children, who learn it as a by-product of education. It seems to be the case that in the case of poor language learners such as individuals with DLD, these are not able to move toward the next step of the syntactic structure. This statement was supported by Aman's (2007) study who reported an effect of exposure to the variation of a standard language on children's syntactic abilities.

With regard to sentence production, the results of this study show that a significant difference occurs between the performance of the children with DLD and the age-matched group. Children with DLD have acquired and prefer the Wh- *in situ* structure compared with TD children of their same age, who use moved structures of different kinds, for example, Wh- fronting with subject–auxiliary inversion. In the context of the Malay language, it is possible that the use of productive Wh- movement questions among age-matched children has a connection with their exposure to a formal learning of grammar that is the prescribed grammar of standard Malay in schools. Both groups of children aged 7–9 years received exposure to the Malay language, which follows the rules set by Karim et al. (2009:15), a prescribed grammar book that provides an explanation of Malay grammar and is used as reference grammar by schoolteachers. In standard Malay grammar rules, the Wh- movement question is the only option allowed for questions. It is possible that exposure through formal learning has an influence on the differences between the two subject groups.

This study also examined aspects of errors committed by the three subject groups. Results show that there is an inclination for children with DLD and language-matched children to omit the Wh- element and produce an Echo question, namely, a declarative question with an interrogative intonation. The omission of Wh- elements is not surprising in the language acquisition process (Gerken, 1994; Schmerse et al., 2013). An interesting error reported in our DLD group is the substitution of the Wh- element with a generic *mana* (where). Such a strategy shows that children produce Wh- questions, but they express them in an arbitrary manner without considering the context of the sentence and, more importantly, making use of an element that does not require any interference with the subject position. This assumption is consistent with results reported in Long's (1993) study for affixation, with Malay children using affix forms arbitrarily during acquisition, following a grammatical underspecification strategy. The use of a generic Wh- placeholder is also supported by the pervasive error of verb omission reported in children with DLD in this study, who tend to produce phrasal utterances instead. According to Aman (2007), the omission of verbs is one of the sentence simplification strategies adopted by young children speakers of Malay.

Regarding lexical level errors, this study findings reveal that children tend to substitute verbs with other verbs with same interference in the semantic relation, for example, an antonymous relationship (push for pull). A similar error was recently reported in the interpretation of active reversible sentences in a group of Malay speakers with aphasia (Aziz et al., 2020) with lexical substitutions in favor of semantically related verbs. These results were explained as an underspecification of the grammatical affixes in transitive verb forms (e.g., agentive markers and voice), often reported and theorized in adults with acquired language impairments as well as in children with DLD (Garraffa and Grillo, 2008; Adani et al., 2010).

Clinical implications

Examining sentence structures that might potentially be a clinical marker for Malay children with DLD, the threshold score of the CA group was used as the cutoff point to measure sensitivity and specificity for all constructions, in support of a syntactic structure-based approach to clinical markers. This was implemented as there were significant differences on the performance of children with DLD compared with the TD Malay children of the same chronological age. The results obtained in this study strongly suggest that Which-questions, and in particular the comprehension of Which-object questions, are possible candidates to be linguistic clinical markers in Malay. Future studies are required to further corroborate the results on a larger population and to further investigate acquisition of *which* questions in children with DLD. A follow-up study making use of a syntactic priming paradigm specific for *which* questions could better explore whether children with DLD can acquire any Wh- dependencies under a controlled setting and with more exposure to the structure (see Garraffa et al., 2018 for a study on the acquisition of relative clauses in DLD *via* syntactic priming).

Conclusion

This study represents a major contribution to the investigation of language development in children speakers of Malay and provides finer details on information regarding the ability and language development of children with DLD. Overall, the study reveals that Malay children with DLD at this stage (mean age 9;7 years) master comprehension of most Wh- questions, but not production, thus confirming a modality-driven component, which has been reported in several other studies for both TD children and children with DLD. However, in terms of the description of syntactic abilities of children with DLD, a modality-driven approach cannot explain the variation of both structural and lexical errors reported in the atypical group, as well as the selective difficulty with which object questions. In terms of quality, an analysis of errors shows that although quantitatively similar, the language make-up of children with DLD has some differences with that of younger, age-matched children.

One aspect of the late acquisition of Wh- questions in children with DLD is linked to the extraction of the Wh- from its argument position, supporting studies on Wh- questions across languages which show difficulties in understanding which questions. In the case of Malay as reported here, children with DLD adopt a series of strategies that appear to be related to an immature or truncated syntactic tree. This reduced tree allows for non-adult-like optional constructions, including Wh- *in situ* questions, use of a generic Wh- element, substitution with yes/no echo questions, and, at lexical level, incorrect use of verbs.

Factors such as the formal education of Malay were reported to have an influence on the usage pattern of Wh- questions in the older children, but not for children with DLD. Children with DLD at the ages of 7–9 years are still unable to use the particle –*kah* compared with age-matched children. The implications of formal education on the acquisition of grammatical properties and the need for extra support for the children with atypical language development need to be explored further.

Data availability statement

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

Ethics statement

The studies involving human participants were reviewed and approved by the Research Ethics Committee, Universiti Sains Malaysia (USM). Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Author contributions

NA conceptualized the research, collected and analyzed the data, drafted the work for intellectual content, gave final approval of the version to be published, and was responsible for the accuracy and integrity of data and that the data have been investigated and resolved thoroughly. GS and MG analyzed the data, drafted the work for intellectual content, gave final approval of the version to be published, and was responsible for the accuracy and integrity of data and that the data have been investigated and resolved thoroughly. RR conceptualized the research, analyzed the data, drafted the work for intellectual content, gave final approval of the version to be published, and was responsible for the accuracy and integrity of data and that the data have been investigated and resolved thoroughly. All authors contributed to the article and approved the submitted version.

Funding

This study was funded by two grants: Research University Grant (GUP) (code: UKM-GUP-2011-134) from Universiti Kebangsaan Malaysia and research grant PHUMANITI 6315272 from Universiti Sains Malaysia.

Acknowledgments

We thank each of the children participants and their parents for their involvement in this study. A special note to Norhaida Aman, National Institute of Education, Singapore, for providing us a copy of her work on Wh- Q among Malay children in Singapore.

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.

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

Adani, F., van der Lely, H., Forgiarini, M., and Guasti, M. T. (2010). Grammatical feature dissimilarities make relative clauses easier: A comprehension study with Italian children. *Lingua* 120, 2148–2166. doi: 10.1016/j.lingua.2010.03.018

Aman, N. (2007). *The Acquisition of Malay Wh-Questions*. Muenchen: LINCOM.

Aman, N. (2014). "An Overview of The Acquisition of Malay Wh-Questions," in *South and Southeast Asian Psycholinguistics*, eds H. Winskel and P. Padakannaya (Cambridge: Cambridge University Press), 89–98. doi: 10.1017/CBO9781139084642.011

Avrutin, S. (2000). "Comprehension of Wh-questions by children and Brocas aphasics," in *Language and The Brain: Representation and Processing*, eds Y. Grodzinsky, L. P. Shapiro, and D. A. Swinney (San Diego: Academic Press), 295–313. doi: 10.1016/B978-012304260-6/50017-7

Aziz, M. A., Hassan, M., Razak, R., and Garraffa, M. (2020). Syntactic abilities in Malay adult speakers with aphasia: A study on passive sentences and argument structures. *Aphasiology* 34, 886–904. doi: 10.1080/02687038.2020.1742283

Belletti, A. (2017). "(Past) Participle Agreement," in *The Wiley Blackwell Companion to Syntax,* (2 Edition), eds M. Everaert, C. Henk, and V. Riemsdijk (New York, NY: John Wiley and Sons, Inc), 1–29. doi: 10.1002/9781118358733.wbsyncom081

Belletti, A., and Guasti, M. T. (2015). *The Acquisition of Italian Morphosyntax and Its Interfaces in Different Modes of Acquisition*. Amsterdam: John Benjamins. doi: 10.1075/lald.57

Bishop, D. V. M. (2006). What causes specific language impairment in children? *Curr. Dir. Psychol. Sci.* 15, 217–221. doi: 10.1111/j.1467-8721.2006.00439.x

Bishop, D. V. M. (2017). Why is it so hard to reach agreement on terminology? The case of developmental language disorder (DLD). *Int. J. Lang. Commun. Disord.* 52, 671–680. doi: 10.1111/1460-6984.12335

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T., and The Catalise-2 consortium . (2017). 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 Dorothy, V. M. (2003). *Test of Reception of Grammar (Trog-2)*. London: Harcourt Assessment.

Bortolini, U., Arfé, B., Caselli, M. C., Degasperi, L., Deevy, P., and Leonard, L. B. (2006). Clinical markers for specific language impairment in Italian: The contribution of clitics and non-word repetition. *Int. J. Lang. Commun. Disord.* 41, 695–712.

Cole, P., and Hermon, G. (1998). The typology of wh-movement: Wh-questions in Malay. *Syntax* 1, 221–258. doi: 10.1111/1467-9612.00009

Contemori, C., and Garraffa, M. (2010). Comparison of modalities in SLI syntax: A study on the comprehension and production of non-canonical sentences. *Lingua* 120, 1950–1965. doi: 10.1016/j.lingua.2010.02.011

Conti-Ramsden, G., Botting, N., and Faragher, B. (2001). Psycholinguistic markers for specific language impairment (SLI). *J. Child Psychol. Psychiatry* 42, 741–748.

De Vincenzi, M., Arduino, Ciccarelli, and Job, R. (1999). "Parsing strategies in children comprehension of interrogative sentences," in *Proceedings of Long Papers of European Conference of Cognitive Science*, ed. S. Bagnara (Rome: Istituto di Psicologia del CNR), 301–308.

Deevy, P., and Leonard, L. B. (2004). The comprehension of wh-questions in children with Specific Language Impairment. *J. Speech Lang. Hear. Res.* 47, 802–816. doi: 10.1044/1092-4388(2004/060)

Ebbels, S., and van der Lely, H. (2001). Meta-syntactic therapy using visual coding for children with severe persistent SLI. *Int. J. Lang. Commun. Disord.* 36, 345–350. doi: 10.3109/13682820109177909

Friedmann, N., Belletti, A., and Rizzi, L. (2009). Relativized relatives: Types of intervention in the acquisition of A-bar dependencies. *Lingua* 119, 67–88. doi: 10.1016/j.lingua.2008.09.002

Friedmann, N., and Novogrodsky, R. (2011). Which questions are most difficult to understand? *Lingua* 121, 367–382. doi: 10.1016/j.lingua.2010.10.004

Garraffa, M. (2017). Grammar as a maturational controlled behavior: The case of grammatical based intervention structures. *Biolinguistics* 11, 353–366. doi: 10.5964/bioling.9097

Garraffa, M., Coco, M., and Branigan, H. (2018). Impaired implicit learning of syntactic structure in children with language impairment: Evidence from structural priming. *Autism Dev. Lang. Impairment* 3, 1–15. doi: 10.1177/2396941518779939

Garraffa, M., and Grillo, N. (2008). Canonicity effects as a grammatical phenomenon. *J. Neurolinguistics* 21, 177–197.

Gerken, L. (1994). Young children's representation of prosodic phonology: Evidence from English-speakers' weak syllable productions. *J. Mem. Lang.* 33, 19–38. doi: 10.1006/jmla.1994.1002

Guasti, M. T. (2017). *Language Acquisition: The Growth of Grammar*. Cambridge, MA: MIT Press. doi: 10.1093/oxfordhb/9780199573776.013.12

Guasti, M. T., Branchini, C., and Arosio, F. (2012). Interference in the production of Italian subject and object wh-questions. *Appl. Psycholinguist.* 33, 185–223. doi: 10.1017/S0142716411000324

Hamann, C. (2006). Speculations About Early Syntax : The Production of Wh-questions by Normally Developing French Children and French Children with SLI. *Catalan J. Linguistics*. 5, 143–189. doi: 10.5565/rev/catjl.82

Hansson, K., and Nettelbladt, U. (2006). Wh-questions in Swedish children with SLI. *Int. J. Speech Lang. Pathol.* 8, 376–383. doi: 10.1080/14417040600880722

Heim, I. (1987). "Where does the definiteness restriction apply?," in *The respesentation of (In) Definites*, ed. E. Reuland (Cambridge, Mass: MIT Press).

Hirsch, C., and Hartman, J. (2006). "Some (wh-) questions concerning passive interactions in. language acquisition and development," in *The Proceedings of Gala 2005, Belletti, Bennati, Chesi*, eds Di Domenico and Ferrari (Newcastle upon Tyne: Cambridge Scholar Press), 256–268.

Jakubowicz, C. (2011). Measuring derivational complexity: New evidence from typically-develop- ing and SLI learners of L1-French. *Lingua* 121, 339–351. doi: 10.1016/j.lingua.2010.10.006

Kader, M. (1981). *The Syntax of Malay Interrogatives*. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Kader, M., and Tan, H. T. (2002). Pemerolehan morfologi dan sintaksis kanak- kanak Melayu. [The acquisition of morphology and syntax among malay children]. *J. Bahasa* 2, 352–373.

Karim, N. S., Onn, F. M., Musa, H., and Mahmood, A. H. (2009). *Tatabahasa Dewan*. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Leonard, L. B. (1998). *Children with Specific Language Impairment*. Cambridge: MIT Press.

Levy, H., and Friedmann, N. (2009). Treatment of syntactic movement in syntactic SLI: A case study. *First Lang.* 29, 15–49. doi: 10.1177/0142723708097815

Long, J. (1993). *Pemerolehan Imbuhan Pada Peringkat Prasekolah dan Implikasinya Terhadap Pendidikan Bahasa [The Acquisition of Affixes at the Preschool and its Implication Toward Language Education]*. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Luyee, E. O., Roselan Nor, F. I., Anwardeen, N. H., and Mohd Mustafa, F. H. (2015). Suitability of the Literacy and Numeracy Screening (LINUS) 2.0 Program in assessing children's early literacy. *Malays. Online J. Educ. Sci.* 3, 36–44.

Marinis, T., and van der Lely, H. K. J. (2007). On-line processing of wh-questions in children with G-SLI and typically developing children. *Int. J. Lang. Commun. Disord*. 42, 557–582. doi: 10.1080/13682820601058190

Novogrodsky, R., and Friedmann, N. (2009). The production of relative clauses in syntactic SLI: A window to the nature of the impairment. *Advan. Speech Lang. Pathol.* 8, 364–375. doi: 10.1080/14417040600919496

Paul, R., Norbury, C., and Gosse, C. (2012). *Language disorders from infancy through adolescence*. Amsterdam: Elsevier.

Raven, J. C., Court, J. H., and Raven, J. (1998). *Colored Progressive Matrices*. Paris: Les Editions du Centre de Psychologie Appliquée.

Razak, R. A. (2003). *The Syntax and Semantics of Quantification in Malay*. Kuala Lumpur: Akademi Pengajian Melayu.

Razak, R. A. (2013). "Studies on the acquisition of morphology and syntax Malay children in Malaysia: Issues, challenges and needs," in *South and Southest Asian Psycholinguistics*, eds Heather Winskel and Prakash Padakannaya (Cambridge: Cambridge University Press), 133–144. doi: 10.1017/CBO97811390846 42.015

Razak, R. A., Neelagandan, A. I, Md Yusuf, N., Lim, H. W., Ahmad, K., and Madison, C. (2018). Validation of the Malay Preschool Language Assessment Tool (MPLAT): The Screening and Diagnostic Versions. *Malays. J. Public Health Med.* 1, 191–115.

Riches, N. G., and Garraffa, M. (2017). A discourse account of intervention phenomena: An investigation of interrogatives. *Glossa J. General Linguist.* 2, 1–20. doi: 10.5334/gjgl.100

Rizzi, L. (1993/1994). Some Notes on Linguistic Theory and Language Development: The Case of Root Infinitives. *Lang. Acquis*. 3, 371–393. doi: 10.1207/s15327817la0304_2

R Studio Team (2022). *RStudio: Integrated Development Environment for R. RStudio*. Boston, MA: RStudio Team.

Salleh, R. (1989). *Fronted Constituents in Malay-Base Structures and Move Alpha in a Configurational Non-Indo-European Language*. Kuala Lumpur: Dewan Bahasa Dan Pustaka.

Schmerse, D., Lieven, E., and Tomasello, M. (2013). Error patterns in young German children's wh-questions. *J. Child Lang*. 40, 656–671. doi: 10.1017/S0305000912000104

Stark, R. E., and Tallal, P. (1981). Selection of Children with Specific Language Deficits. *J. Speech Hear. Disord*. 46, 114–122. doi: 10.1044/jshd.4602.114

Tan, H. T. (1999). *Pemerolehan bahasa melayu: Kajian ke atas perkembangan ayat interogatif oleh kanak-kanak melayu berumur 18-48 bulan [Malay language acquisition: A study of the development interrogative sentences by Malay children between 18-48 months]*. Unpublished Ph.D. Dissertation. Penang: Universiti Sains Malaysia.

The Jamovi Project, (2021). *Jamovi. (Version 1.6) [Computer Software]*. Available online at https://www.jamovi.org (accessed July 1, 2022).

Thornton, R. (1995). Referentiality and Wh-Movement in Child English: Juvenile D-Linkuency. *Lang. Acquis.* 4, 139–175. doi: 10.1080/10489223.1995.9671662

Thornton, R. (2016). "Acquisition of questions," in *The Oxford Handbook of Developmental Linguistics*, eds J. Lidz, W. Snyder, and J. Pater (Oxford, UK: Oxford University Press), 310–340. doi: 10.1093/oxfordhb/9780199601264.013.14

van der Lely, H. K., and Stollwerck, L. (1997). Binding theory and grammatical specific language impairment in children. *Cognition* 62, 245–290. doi: 10.1016/S0010-0277(96)00783-4

van der Lely, H. K. J., and Battell, J. (2003). Wh-movement in children with grammatical SLI: A test of the RDDR hypothesis. *Language* 79, 153–181. doi: 10.1353/lan.2003.0089

Watanabe, A. (2006). *In WH movement: Moving on Edit Lisa Cheng and Norbert Cover*. Cambridge: MIT Press.

Wexler, K. (1998). Very early parameter setting and the unique checking constraint: A new explanation of the optional infinitive stage. *Lingua* 106, 23–79. doi: 10.1016/S0024-3841(98)00029-1

Wexler, K. (2003). "Lenneberg's dream: Learning, normal language development, and specific language impairment," in *The Language Competence Across Populations: Toward a Definition of Specific Language Impairment*, eds Yonata Levy, C. Jeannette, and Schaeffer (Mahwah, NJ: Lawrence Earlbaum), 11–61.

Wong, A. M. Y., Leonard, L. B., Fletcher, P., and Stokes, S. F. (2004). Questions without movement: A study of Cantonese-speaking children with and without specific language impairment. *J. Speech Lang. Hear. Res.* 47, 1440–1453. doi: 10.1044/1092-4388(2004/107)

Wong, B. E. (2008). Wh-Questions in malay: An explanation for the restriction of extraction to subject position with yang. *3L Southeast Asian J. Engl. Lang. Stud.* 14, 109–126.

Yoshinaga, N. (1996). *Wh-questions: A comparative Study of their Form and Acquisition in English and Japanese*, Ph.D thesis, Hawaii: University of Hawaii.

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 16 November 2022
DOI 10.3389/fpsyg.2022.992512

Check for updates

OPEN ACCESS

EDITED BY
Theodoros Marinis,
University of Konstanz, Germany

REVIEWED BY
Maria Martzoukou,
University of Ioannina, Greece
Eirini Sanoudaki,
Bangor University, United Kingdom

*CORRESPONDENCE
Eliseo Diez-Itza
ditza@uniovi.es
Verónica Martínez
martinezveronica@uniovi.es

SPECIALTY SECTION
This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Psychology

RECEIVED 12 July 2022
ACCEPTED 01 November 2022
PUBLISHED 16 November 2022

CITATION
Pérez V, Martínez V and Diez-Itza E
(2022) Late phonological development
in Williams syndrome.
Front. Psychol. 13:992512.
doi: 10.3389/fpsyg.2022.992512

Late phonological development in Williams syndrome

Vanesa Pérez[1,2], Verónica Martínez[1]* and Eliseo Diez-Itza[1]*

[1]LOGIN Research Group, University of Oviedo, Oviedo, Spain, [2]Escuelas Universitarias Gimbernat, University of Cantabria, Torrelavega, Spain

Williams syndrome is a neurodevelopmental genetic disorder characterized by a unique phenotype, including mild to moderate intellectual disability and an uneven neuropsychological profile of relative strengths and weaknesses. Language structure components (i.e., phonology, morphosyntax, and vocabulary) have been considered an area of specific ability compared to pragmatic language use. However, research on phonological development in Williams syndrome is very scarce, and it suggests atypical patterns. Therefore, the aim of the present study was to explore the profiles of late phonological development in Spanish-speaking children, adolescents, and adults with Williams syndrome, based on the analysis of five classes of processes (Syllable Structure, Substitution, Omission, Assimilation, and Addition) in spontaneous speech. The phonological profiles of seven children (aged 3–8 years), and seven adolescents and young adults (aged 14–25 years) with Williams syndrome were compared with two normative groups of typically developing (TD) children at different stages of late phonological development (aged 3 and 5 years). The frequency of phonological processes in the group of children with Williams syndrome was similar to that of 3-year-old TD children, which suggests that they would be in the first stage of late phonological development (expansion stage). The group of older individuals with Williams syndrome showed a much lower frequency of processes, similar to that of 5-year-old TD children in the last stage of phonological development (resolution stage). However, their phonological processes appeared to be persistent and independent of chronological age. Furthermore, asynchronies in quantitative and qualitative profiles (relative frequency) indicated atypical and complex trajectories in late phonological development, which cannot be described as simply delayed or protracted. Remarkable individual differences were observed, especially in the group of adolescents and adults with Williams syndrome, although the majority of cases conformed to the modal profiles of their groups. A major tendency for Omission, including final consonant deletion, may be considered atypical and specific to Williams syndrome at all ages. The results of the present study raise the need for continued and appropriate phonological assessment and treatment for people with Williams syndrome across the lifespan.

KEYWORDS

Williams syndrome, phonological development, intellectual disability, spontaneous speech assessment, phonological processes, atypical language development, neurodevelopmental genetic disorders

Introduction

Williams syndrome (WS) is a neurodevelopmental disorder caused by a heterozygous deletion of between 26 and 28 genes on chromosome 7q11.23 (Pérez Jurado, 2003). The WS physical phenotype includes a distinctive facial appearance, hoarse voice, and sound sensitivities (hyperacusis, odynacusis, auditory allodynia, and auditory fascinations) (Kozel et al., 2021). Individuals with WS may show mild-to-moderate intellectual disability in conjunction with a distinct neurocognitive profile of relative strengths and weakness (Bellugi et al., 2000). Several studies have identified specific deficits in executive functioning (working memory, attentional abilities, and inhibition), problem-solving, and visuospatial skills (Camp et al., 2016; Heiz and Barisnikov, 2016; D'Souza et al., 2020). In contrast, auditory processing and face recognition are strengths in the WS profile (D'Souza et al., 2015; Miezah et al., 2020). Akin to the uneven cognitive profile, they also appear to show relative strengths and weaknesses in the motor profile, in the context of persisting fine and gross motor difficulties into childhood and adulthood (Mayall et al., 2021). Behavioral and emotional problems (attention, anxiety, and a range of social problems) have been also reported, together with a unique prosocial personality characterized by overfriendliness, a strong drive to approach strangers, gregariousness, bias toward positive affect, and heightened social engagement yet difficult peer interactions (Järvinen et al., 2013; Pérez-García et al., 2017). Special difficulties in adaptive behavior related to personal autonomy have also been described (Kirchner et al., 2016).

Language was first described as being selectively preserved and dissociated from other cognitive functions (Bellugi et al., 1988), although further research noted that language skills in individuals with WS were not intact and had complex interrelations with cognitive abilities (Mervis et al., 2004; Mervis and Becerra, 2007). Superior verbal skills reported in individuals with WS may be explained in terms of asynchronous trajectories of development with verbal ability progressing at a faster rate than non-verbal ability (Jarrold et al., 2001). In the same vein, language also shows asymmetrical development across different levels with varying outcomes in respect to what is expected for chronological and mental age (Brock, 2007). Pragmatic ability is an area of relative weakness, both in narrative and conversational settings (Stojanovik et al., 2001; Reilly et al., 2004; Stojanovik, 2006; Diez-Itza et al., 2018, 2022). In contrast, structural aspects of language have been described as relative strengths in the WS linguistic profile. Morphosyntactic abilities had been considered selectively spared (Clahsen et al., 2004), although this assumption was challenged in several studies indicating some degree of atypical morphological processing (Thomas et al., 2001; Boloh and Ibernon, 2010; Benítez-Burraco et al., 2017; Diez-Itza et al., 2017). Receptive vocabulary is also an area of relative strength in people with Down syndrome, but only for concrete vocabulary (Mervis and John, 2008; Garayzábal et al., 2014; Moraleda and López, 2020). Regarding lexical production, a tendency to use rare words and an atypical pattern of semantic categorization has been reported (Bellugi et al., 1994; Purser et al., 2010).

The phonological level is often considered another area of strength in the WS linguistic profile, although very few studies have directly assessed it. Most previous research focuses on phonological fluency, short-term memory (STM), phonological perception, and phonological awareness and processing (Vicari et al., 1996a,b; Volterra et al., 1996; Majerus et al., 2003; Majerus, 2004). Different studies have also been conducted on prosodic skills and their specific characteristics in the WS profile (Stojanovik, 2010; Martínez-Castilla et al., 2012). Only a few more recent studies have addressed phonological production in individuals with WS, although spontaneous speech was not analyzed but rather, words elicited from articulation tests (Hidalgo, 2019; Huffman, 2019). In general, both direct studies of production and those of phonological processing or prosody show that these skills are not fully preserved and that difficulties persist into adolescence and adulthood. However, in late phonological development, individuals with WS reach more advanced stages than other neuroevolutionary genetic syndromes, such as WS duplication syndrome, Smith Magenis syndrome, Down syndrome, and Fragile X syndrome (Mervis et al., 2015; Huelmo et al., 2017; Hidalgo and Garayzábal, 2019; Diez-Itza et al., 2021).

The existence of within-domain dissociations within the linguistic domain in WS, as well as specific developmental trajectories and atypical features, especially in the case of morphology, has been widely discussed (Karmiloff-Smith et al., 1997; Karmiloff-Smith, 1998; Diez-Itza et al., 2017). Phonological development provides a better example of emergent complexity, i.e., the changing nature of a complex system over time, revealing principles and milestones across languages (Davis and Bedore, 2013; McLeod and Crowe, 2018). The study of late stages in phonological development also suggests that the underlying dynamics are complex, from system expansion at around 3 years of age to its resolution at 5 years of age, which does not directly correspond to lexical production (Diez-Itza et al., 2001; Diez-Itza and Martínez, 2004). In this context, it could be discussed whether the alterations respond to a mere quantitative delay compared to typical development or whether they present trajectories specific to each disorder or syndrome (Rose and Inkelas, 2011). In this sense, the existence of protracted phonological development has been suggested in those cases with developmental trajectories that tend to converge late with those of typical development (Bernhardt and Stemberger, 2017; Vergara et al., 2021).

Both quantitative and qualitative differences could also depend on the age of the WS individuals studied. This question was addressed in one of the few studies that directly assessed the consonant articulation accuracy in two groups of English-speaking WS individuals (younger children: aged 4–9 years; older children and adolescents: aged 10–17 years) administered a Test of Articulation (Huffman, 2019). Consonant production accuracy was below expectations in both groups, but it was significantly higher for older children and adolescents. Patterns of articulatory accuracy in the group of younger children with WS were similar to the patterns of typically developing (TD) children, which means that articulation was significantly more accurate for early-developing consonants, followed by middle-developing consonants, and less accurate for late-developing consonants. In the group of older children and adolescents, all the early-developing consonants were correct, but this was not the case for middle- and late-developing consonants, where a similar proportion of articulatory accuracy was found. Manner-of-production was one of the sources of variation in articulatory accuracy, with Nasal and Stop consonants being significantly more accurate than Fricative and Approximant consonants in both groups. Although the patterns were similar, the older individuals showed quantitative growths: Nasal and Stop consonants reached full accuracy, and Fricative and Approximant consonants increased their accuracy by 50% to almost 90% of correct production. Articulatory accuracy of consonant clusters was also assessed and showed a sharp increase of almost 100% in the group of older children and adolescents with WS, and quite different patterns concerning particular vocal tract planes of movement in the control for articulatory accuracy.

The phonological production of Spanish-speaking individuals with WS between 4 and 31 years of age, compared with that of other syndromes, was also investigated by Hidalgo (2019) from the perspective of the phonological processes of simplification described by Bosch (2004) in TD children aged 3–7 years and the late stages of phonological development (expansion, stabilization, and resolution) established by Diez-Itza and Martínez (2004). From an articulation test, she observed that beyond the age of 6 years, phonetic and phonological repertoires were acquired by children with WS, although in some adolescents and adults, processes related to rhotic consonants persisted. The most frequent syllabic structure processes were cluster reduction (attacks and complex nuclei) and metathesis, and in a lower percentage, unstressed syllable omission, and addition, while reduplication and final consonant deletion processes were absent. In the case of segmental processes, the most frequent were absence or backing of rhotics, and in a lower percentage backing and deaffrication of other consonants, as well as assimilation processes.

Regarding phonological fluency, initial studies suggested that this is preserved in the WS linguistic profile, with children and adolescents with WS aged 4–15 years scoring better than their mental age-matched TD controls on a phonological fluency test without semantic involvement (Volterra et al., 1996). Based on these results, it was hypothesized that if only the phonological aspects of language develop at a normal rate while grammatical and lexical-semantic components remain impaired, it is because there is a dissociation between normal short-term and impaired long-term verbal memory in WS (Vicari et al., 1996b). Furthermore, performance in a word span task revealed comparable effects of phonological similarity and length to those observed in TD children, while the effect of frequency was significantly lower in WS participants, which was interpreted as the result of impaired access to lexical-semantic knowledge (Vicari et al., 1996a). Thus, a complex pattern of dissociation in linguistic processing and "atypical" development of WS children was revealed. It is important to note that the strength in phonology that these studies revealed is in any case relative since they compare individuals with WS with children of equal mental age but of much younger chronological age. Moreover, phonological development culminates in TD before the age of 9 years, whereas lexical development is open-ended.

The repetition of pseudowords has also contributed to the study of STM, showing that individuals with WS continue to rely strongly on phonological STM in the acquisition of new words, which is observed in 4-year-old but not in 5-year-old children (Grant et al., 1997). Phonological perception skills according to a nonsense syllables repetition test were comparable to those of TD participants with the same chronological age (range: 11–52 years) (Böhning et al., 2002). In a group of four children with WS who were administered both a word and pseudoword repetition test, their relative strength in STM was also confirmed to be comparable to that of children of the same chronological and verbal age in many respects, especially in the case of pseudowords where the support of phonological and lexico-semantic knowledge was minimized (Majerus et al., 2003).

In addition to word span and non-word repetition, phonological processing and phonological awareness skills were also studied in a group of children, adolescents, and adults with WS, which were compared with those of a group of TD children (mean age: 6.9), with differences emerging only in the phoneme deletion subtest (Laing et al., 2001). However, when the control groups were of the same chronological age or a verbal age closer to their chronological age, differences were observed in most measures of phonological awareness (Majerus et al., 2003). These results were explained by impairment at the level of the phonological representation (less finely grained) and the lexical-semantic representation (suggesting an abnormally structured network).

Phonological development is also often related in the early stages to motor aspects, as is the case with babbling. It has been claimed that the delay in the onset of canonical babbling and the first words observed in infants with WS is due to a delay in the acquisition of early motor milestones (Masataka, 2001). These findings are consistent with Velleman et al. (2006) who also observed delays in prelinguistic vocal development in six toddlers with WS. The postverbal onset of declarative gestures has also been linked with an atypical path of language development (Becerra and Mervis, 2019). An atypical accelerated trajectory of phonological development in two children with WS aged 5 was described by Martínez et al. (2014). At later stages, individuals with WS tend to present few phonological errors, which contrasts with the fact that difficulties in planning and coordinating oral-motor praxis in adolescents and adults with WS seem to persist (Krishnan et al., 2015).

Most studies, however, have not been conducted using developmental designs or naturalistic methodologies. Levy and Eilam (2013) analyzed extended spontaneous conversations in a mixed longitudinal study of two groups of children with WS and DS across five stages of morphophonological development. They concluded that there is a late-onset in both groups, determining atypical trajectories, which tend to show greater syndromic specificity at later stages of development. Capirci et al. (1996) and Diez-Itza et al. (1998), in longitudinal case studies of children with WS, found atypical phonological errors in conversational speech. The only recent study to our knowledge that addresses some aspects related to phonological production in spontaneous speech is that of Hargrove et al. (2012), who observed that adolescents with WS, although maintaining similar levels of intelligibility to their age peers, present a significantly lower rate of phonological accuracy, reaching more than 3% of incorrect words. They also found, like previous studies, a significantly slower speech rate in individuals with WS (Semel and Rosner, 2003; Setter et al., 2007; Crawford et al., 2008). However, their aims were not focused on the detailed analysis of phonology, nor did they offer a developmental perspective.

Several studies of late phonological development in TD Spanish-speaking children have been conducted based on cross-sectional designs. Aguilar and Serra (2003) and Bosch (2004) devised articulation tests and administered them to deliver normative data from children aged 3–7, including age of acquisition of the phonemic inventory and common processes at the different age stages. Diez-Itza et al. (2001), Diez-Itza and Martínez (2004), and Martínez (2010) registered and analyzed spontaneous speech corpora computing the frequency and the percentage distribution of phonological processes in children aged 3–5. An explicit aim of these analyses was to describe stages of phonological development as in previous studies by Ingram (1976) and Grunwell (1981). However, beyond a taxonomic description of processes at the different stages, the research by Diez-Itza and colleagues looked for quantitative and qualitative differences and non-linear trajectories of development. They found a reduction of the frequency of processes and changes in their relative distribution as age increased, suggesting three stages in late phonological development: expansion (age 3), stabilization (age 4), and resolution (age 5). Within the same theoretical and methodological framework, the present study aimed to further advance in a detailed description of late phonological development in children, adolescents and young adults with WS.

Objectives

The main objective of the present study was to explore the profiles of late phonological development of Spanish-speaking individuals with WS to determine change across developmental stages and whether specific features would be exhibited. The profiles were based on the analysis of five classes of processes (Syllable Structure, Substitution, Omission, Assimilation, Addition) in spontaneous speech. The frequency and percentage distribution of processes were calculated, and modal profiles and outliers were determined by cluster analysis. It was hypothesized that late phonological development in WS follows the stages of typical development (i.e., expansion, stabilization, and resolution) and that phonological patterns show not only quantitative but also qualitative differences. To assess these hypotheses, the phonological profiles of children (aged 3–8), and adolescents and young adults (aged 14–25) with WS were compared with normative groups of TD preschool children at two stages of late phonological development (aged 3 and 5 years).

Materials and methods

Participants

The participants were 14 monolingual Spanish-speaking individuals with WS divided into two age groups (see

Table 1): the first group (WS1) were children (chronological age: *M* = 5.8; *SD* = 1.6); the second group (WS2) were adolescents and young adults (chronological age: *M* = 19.6 years; *SD* = 3.7). They had been previously diagnosed by the molecular genetic test fluorescence *in situ* hybridization (FISH) and presented the characteristic phenotype. Parents and legal guardians provided informed consent for the participants to take part in the study.

To assess verbal lexical age and its relationship with phonological development, the participants were administered the Peabody Picture Vocabulary Test (Dunn et al., 2010): WS1 verbal age (*M* = 3.6; *SD* = 1.1) and WS2 verbal age (*M* = 10; *SD* = 2.4).

Normative data on the late phonological development of TD children were obtained from Martínez (2010), who established three stages in late phonological development (expansion, stabilization, and resolution) from 3.0 to 5.11. This study provides normative data in Spanish about phonological processes with the same methodology of spontaneous speech analysis as the present study. Thus, the WS1 and the WS2 groups were matched respectively with the group of younger children in the expansion stage (TD1) and the group of older children in the resolution stage (TD2) based on the frequency of processes. The TD1 normative group consisted of 40 children (20 girls and 20 boys; chronological age: *M* = 3.3 years; *SD* = 0.2); and the TD2 normative group also consisted of 40 children (20 girls and 20 boys) (chronological age: *M* = 5.8 years; *SD* = 0.3).

The participants with WS and TD children in the normative groups belonged to urban middle classes based on their district of residence within the Principality of Asturias and Cantabria (Spain), where a standard variant of Spanish (Castilian) is spoken.

TABLE 1 Gender, chronological and verbal age, and education of the participants with Williams syndrome.

Group	Case	Gender	CA	VA	Education
WS1	S1	Male	3.7	2.5	Regular school
	S2	Female	4.5	2.8	Regular school
	S3	Male	5.5	3.11	Regular school
	S4	Female	5.5	2.11	Regular school
	S5	Male	5.5	3.4	Regular school
	S6	Female	7.9	5.1	Regular school
	S7	Female	8.2	5.2	Regular school
WS2	S8	Male	14.4	10.1	Special school
	S9	Male	15.3	9.6	Regular school
	S10	Female	18.8	14.4	Vocational training
	S11	Female	19.11	8.6	Occupational center
	S12	Female	20.8	11.8	Occupational center
	S13	Female	23.3	8.8	Special school
	S14	Female	25.8	7.2	Occupational center

CA, chronological age; VA, verbal age.

Instruments and procedure

The RETAMHE methodology, short for Recording, Transcription, and Analysis of Spontaneous Speech Samples (Diez-Itza, 1992; Diez-Itza et al., 1999), was used to obtain the spontaneous speech samples. Speech samples were collected via audio-visual recordings of dyadic conversations between each participant and a researcher, with an estimated duration of 45 min in natural settings, and which are part of larger corpora within the Syndroling Project (Diez-Itza et al., 2014). The researcher, who was familiar with the participants, introduced some degree of standardization by proposing common themes to all participants, according to the procedures developed by Abbeduto et al. (1995). The topics included telling a story, a visit to the doctor, a birthday party, talking about friends and family, weekend and daily activities, trips, and hobbies with variations among participants, following the spontaneous flow of conversation.

These conversations were transcribed in CHAT (Codes for the Human Analysis of Transcripts) format and analyzed with the FREQ program, one of the CLAN (Computerized Language Analysis) software programs, both provided by the CHILDES Project (MacWhinney, 2000). Each transcription was completed by a trained researcher and reviewed by two other researchers independently. Difficulties detected were analyzed jointly by the three investigators and discrepancies were resolved by the principal investigator. A total of 40,634 word tokens, 9,934 word types, and 2,806 phonological processes were analyzed, while 38 words were considered unintelligible.

The categories system proposed by Ingram (1976) and adapted by Diez-Itza et al. (2001) was used to code the phonological processes (PHO). The phonological processes were analyzed and classified into one of the following classes: Syllable Structure (SYS), Substitution (SBT), Omission (OMI), Assimilation (ASM), and Addition (ADD). In turn, each of these classes was divided into different subclasses of processes. Thus, SYS processes included Consonant Cluster Reduction (CCR), Final Consonant Deletion (FCD), Vowel Cluster (diphthong) Reduction (VCR), Unstressed Syllable Deletion (SYD), Metathesis (MTT), and Infrequent Processes (IFQ; Reduplication + Dissimilation + Analogy). SBT and OMI processes included Liquid (LIQ), Vowel (VOW), Fricative (FRC), Voiced Stop (VOS), Voiceless Stop (VLS), and Nasal (NSL). The following example illustrates the transcription and coding procedure according to the minCHAT format of the CHILDES Project.

*CHI: fesa [*] [: strawberry].
%err: **fesa = fresa $PHO:SYS:CCR**;

Data analysis

Once the transcriptions were coded, the frequency of lexical variables was obtained using the FREQ program, that is, the total number of words produced (tokens) by each participant, as well as the count of different words (types) in each transcription. Next, the frequency of the classes and subclasses of phonological processes encoded was obtained with the same program.

Given the variability in the size of the spontaneous speech samples of each participant, the number of processes could not be directly used in the analyses. Therefore, to control differences introduced by the size of the samples, the frequency of processes was calculated through a Phonological Index (PI) (number of processes over 100 tokens).

In addition to the quantitative profile provided by the PI, qualitative distribution of the processes in each participant was analyzed. Therefore, the Relative Frequency (RF) was calculated, i.e., the percentage distribution of phonological processes by classes and subclasses. To calculate the RF, participants in each group who did not present phonological processes were not included in the analyses.

Between-group differences in PI and RF were analyzed using the non-parametric Mann–Whitney *U* test (expressed with the *Z* value) for independent samples, given that the distributions did not always approach normality according to the Shapiro–Wilk test. Additionally, the effect size was calculated by Cohen's *d* using G*Power 3.1 statistical software. The *d* values are typically quantified as small (0.2), medium (0.5), and large (0.8) (Cohen, 1988). Spearman correlation was used to analyze the bivariate relationships between chronological age, verbal age, and PI.

In addition, individual similarities, and differences in the RF profiles of the classes and subclasses of phonological processes were explored by means of hierarchical cluster analysis, determining the modal cluster with the participants most similar to each other and best representing the group profile, additional clusters with participants that resemble each other, and extreme outlying cases.

Statistical analysis of the data was performed using SPSS software (Statistical Product and Service Solutions IBM SPSS Statistics 25.0).

Results

Phonological index

Table 2 reports the PI for each study group, including means for total processes and each class of processes. In the WS1 group, a strong positive correlation was found between chronological age and verbal age ($rs = 0.94$; $p = 0.002$), whereas the PI was negatively correlated with chronological age ($rs = -0.78$; $p = 0.041$); negative correlation between PI and verbal age failed to reach significance ($rs = -0.64$; $p = 0.119$). In the WS2 group, non-significant coefficients were obtained for negative correlation between chronological age and verbal age ($rs = -0.54$; $p = 0.215$); negative correlation between PI and verbal age ($rs = -0.64$; $p = 0.119$); and positive correlation between PI and chronological age ($rs = 0.39$; $p = 0.383$).

Mann–Whitney *U* comparisons showed statistically significant differences between the WS groups in PI (total and in all classes of phonological process), with a large effect size except for ASM processes. The comparisons indicated that the WS1 group presented a higher frequency of all phonological processes except for ASM. No differences were observed between the WS1 and TD1 groups, or between the WS2 and TD2 groups, indicating that they were comparable in terms of the total frequency of processes and the frequency by class of processes, except for OMI. In the WS1 group, the PI for OMI processes was higher than in the TD1 group, and the Mann–Whitney *U* test yielded a statistically significant difference with a large effect size. In the WS2 group, the PI for OMI processes was higher than in the TD2 group, and the Mann–Whitney *U* test showed a statistically significant difference with a medium effect size.

Table 3 reports the PI for SYS subclasses of processes in each study group. Mann–Whitney *U* comparisons showed statistically significant differences between the WS groups, with a large effect size. Analyses showed significantly higher scores for the WS1 group for all SYS processes. No differences were observed between the WS1 and TD1 groups or between the WS2 and TD2 groups, indicating that they were comparable, except for MTT processes in the WS1 vs. TD1 group, and FCD processes in the WS2 vs. TD2 group. The PI for the MTT processes in the WS1 group was higher than in the TD1 group, and the PI for the FCD processes in the WS2 group was also higher than in the TD2 group. In both cases, the Mann–Whitney *U* test yielded statistically significant differences with a medium effect size. Additionally, statistically significant differences in IFQ were observed between the WS1 and TD1 groups with a medium effect size, and between the WS2 and TD2 groups with a small effect size.

Table 4 reports the PI for SBT subclasses of processes in each study group. Mann–Whitney *U* comparisons showed statistically significant differences between WS groups, with a large effect size, except for FRC. In the WS1 group, a higher frequency of SBT processes was observed in all subclasses except for FRC. No differences were observed between the WS1 and TD1 groups or between WS2 and TD2 groups, indicating that they were comparable, except for VOW and NSL substitutions. The PI for VOW substitution processes was much higher in the WS1 group than in the TD1 group, and in the WS2 group than in the TD2 group. In both cases, the Mann–Whitney *U* test yielded statistically significant differences with a large effect size. In addition, in the WS1 group, the PI for NSL substitution processes was higher than in the TD1 group, and the Mann–Whitney *U* test showed a statistically significant difference with a medium effect size.

Table 5 reports the PI for the OMI subclasses of processes in each study group. Mann–Whitney *U* comparisons showed

TABLE 2 Phonological processes index (total and by classes) means and standard deviations for groups, Mann–Whitney *U* test, and effect size.

	WS1	WS2	TD1	TD2	WS1 vs. WS2			WS1 vs. TD1			WS2 vs. TD2		
	PI-M (*SD*)	PI-M (*SD*)	PI-M (*SD*)	PI-M (*SD*)	*U* (*Z*)	*p*	*d*	*U* (*Z*)	*p*	*d*	*U* (*Z*)	*p*	*d*
TOT	18.4 (16.9)	1.8 (1.8)	13.3 (11.0)	1.4 (1.8)	2 (2.9)	0.01	1.4	125 (0.5)	0.65	0.4	110 (0.9)	0.37	0.2
SYS	10.3 (8.7)	1.1 (1.1)	7.7 (6.7)	0.8 (1.2)	2 (2.9)	0.01	1.5	120 (0.6)	0.55	0.3	99 (1.2)	0.22	0.3
SBT	3.9 (4.0)	0.4 (0.6)	3.9 (4.8)	0.3 (0.6)	4 (2.6)	0.01	1.2	138 (0.1)	0.95	0	105 (1.1)	0.29	0.2
OMI	2.7 (2.8)	0.2 (0.2)	0.9 (1.4)	0.1 (0.2)	5 (2.5)	0.01	1.3	74 (2.0)	0.05	0.8	70.5 (2.3)	0.02	0.5
ASM	0.7 (0.8)	0.1 (0.1)	0.5 (0.5)	0.1 (0.1)	10 (1.9)	0.06	1.1	125 (0.5)	0.65	0.3	114 (0.8)	0.42	0.1
ADD	0.4 (0.3)	0.04 (0.04)	0.3 (0.2)	0.1 (0.1)	0 (3.1)	0.01	1.7	100 (1.2)	0.23	0.4	119 (0.6)	0.52	0.8

PI-M, phonological index mean; TOT, total phonological processes index; SYS, syllable structure; SBT, substitution; OMI, omission; ASM, assimilation; ADD, addition; d, Cohen's effect size.

statistically significant differences between WS groups, with a large effect size, except for VOS omission processes. In the WS1 group, a higher frequency of OMI was observed in all subclasses, except for VOS consonants. Differences between the WS1 and TD1 groups were observed in all OMI subclasses, except for LIQ and VOS omissions, where both groups were comparable. The PI of the VOW, NSL, and FRC omission processes was much higher in the WS1 group than in the TD1 group. In all three subclasses, the Mann–Whitney *U* test yielded statistically significant differences with a large effect size. For VLS omission processes, the difference was also statistically significant, with a medium effect size. No differences were observed between the WS2 and TD2 groups, indicating that they were comparable, except for the LIQ and VOW omissions, where the PI in the WS2 group was higher. In both cases, the Mann–Whitney *U* test yielded statistically significant differences with a medium effect size.

Relative frequency

In Figure 1, the compared profiles of RF, i.e., the percentage distribution, for processes by classes are shown. Figure 1A represents the profiles of WS1 and WS2 groups, which were very similar in terms of the percentage of the most frequent classes of processes (SYS, SBT). In the classes of OMI and ASM processes, the profiles of both groups intersected since the WS2 group showed a relatively lower percentage of OMI and a correspondingly higher percentage of ASM. However, the Mann–Whitney *U* test did not yield statistically significant differences: SYS ($U = 22$; $Z = 0.32$; $p = 0.75$; $d = 0.1$); SBT ($U = 17$; $Z = 0.96$; $p = 0.34$; $d = 0.1$); OMI ($U = 12$; $Z = 1.60$; $p = 0.11$; $d = 0.8$); ASM ($U = 11$; $Z = 1.73$; $p = 0.09$; $d = 1.1$); ADD ($U = 24$; $Z = 0.06$; $p = 0.95$; $d = 0.3$).

Figure 1B represents the compared profiles of normative groups TD1 and TD2 ($n = 39$), which were similar in terms of the percentage of SYS processes. The profile of the TD2 group showed a relatively lower percentage of SBT and OMI processes. In both classes, the Mann–Whitney *U* test showed statistically significant differences: SBT ($U = 520$; $Z = 2.56$; $p = 0.01$; $d = 0.4$); OMI ($U = 577.5$; Z = 2.07; $p = 0.04$; $d = 0.1$). Inversely, the profile of the TD2 group showed a relatively higher percentage of ASM and ADD processes, although no further statistically significant differences were observed: SYS ($U = 773$; $Z = 0.07$; $p = 0.95$; $d = 0.1$); ASM ($U = 725$; $Z = 0.54$; $p = 0.59$; $d = 0.1$); ADD ($U = 695$; $Z = 0.84$; $p = 0.40$; $d = 0.6$).

Figure 1C represents the compared profiles of WS1 and TD1 groups, where the profile of the WS1 group showed a higher percentage of OMI processes, and the Mann–Whitney *U* test yielded statistically significant differences: OMI ($U = 69$; $Z = 2.13$; $p = 0.03$; $d = 0.9$). The most frequent processes in both groups were SYS with similar percentages, while the profile of the WS1 group showed a relatively lower percentage of SBT and ASM processes, although no statistically significant differences were observed: SYS ($U = 135$; $Z = 0.\ 15$; $p = 0.88$; $d = 0.1$); SBT ($U = 105$; $Z = 1.05$; $p = 0.30$; $d = 0.5$); ASM ($U = 134$; $Z = 0.18$; $p = 0.86$; $d = 0.4$); ADD ($U = 105$; $Z = 1.05$; $p = 0.30$; $d = 0.1$).

Figure 1D represents the compared profiles of WS2 and TD2 groups ($n = 39$), which were similar in terms of the percentage of SYS and SBT processes. The profile of the WS2 group showed a relatively lower percentage of ASM and ADD processes, and a relatively higher percentage of OMI processes. However, the Mann–Whitney *U* test did not yield statistically significant differences: SYS ($U = 124$; $Z = 0.37$; $p = 0.71$; $d = 0.\ 2$); SBT ($U = 104.5$; $Z = 0.99$; $p = 0.32$; $d = 0.03$); OMI ($U = 84$;

TABLE 3 Syllable structure phonological processes index means and standard deviations for groups, Mann–Whitney *U* test, and effect size.

	WS1	WS2	TD1	TD2	WS1 vs. WS2			WS1 vs. TD1			WS2 vs. TD2		
	PI-M (SD)	PI-M (SD)	PI-M (SD)	PI-M (SD)	*U* (*Z*)	*p*	*d*	*U* (*Z*)	*p*	*d*	*U* (*Z*)	*p*	*d*
CCR	3.9 (3.2)	0.4 (0.5)	4.5 (4.6)	0.5 (1.0)	3 (2.8)	0.01	1.5	134 (0.2)	0.86	0.2	100 (1.2)	0.23	0.1
FCD	3.7 (3.2)	0.3 (0.4)	1.9 (2.3)	0.1 (0.2)	2 (2.9)	0.01	1.5	77 (1.9)	0.06	0.6	64 (2.4)	0.02	0.6
VCR	1.2 (1.2)	0.2 (0.1)	0.7 (0.7)	0.1 (0.2)	4 (2.6)	0.01	1.2	99 (1.2)	0.22	0.5	86 (1.6)	0.10	0.6
SYD	1.1 (1.3)	0.1 (0.1)	0.4 (0.6)	0.1 (0.1)	4 (2.6)	0.01	1.1	90 (1.5)	0.13	0.7	108 (1.0)	0.30	0.2
MTT	0.3 (0.3)	0.02 (0.02)	0.1 (0.3)	0.02 (0.04)	0 (3.1)	0.01	1.3	40 (3.1)	0.01	0.7	100 (1.6)	0.12	0
1IFQ	0.2 (0.2)	0.04 (0.1)	0.1 (0.1)	0.01 (0.03)	8.5 (2.1)	0.04	1.0	62 (2.5)	0.01	0.6	87.5 (2.1)	0.04	0.4

PI-M, phonological index mean; CCR, consonant cluster reduction; FCD, final consonant deletion; VCR, vowel cluster reduction; SYD, unstressed syllable deletion; MTT, metathesis; IFQ, infrequent processes; d, Cohen's effect size.

TABLE 4 Substitution phonological processes index means and standard deviations for groups, Mann–Whitney U test, and effect size.

	WS1	WS2	TD1	TD2	WS1 vs. WS2			WS1 vs. TD1			WS2 vs. TD2		
	PI-M (SD)	PI-M (SD)	PI-M (SD)	PI-M (SD)	*U* (*Z*)	*p*	*d*	*U* (*Z*)	*p*	*d*	*U* (*Z*)	*p*	*d*
LIQ	1.2 (1.3)	0.1 (0.1)	1.2 (2.0)	0.2 (0.5)	0 (3.1)	0.01	1.2	117 (0.7)	0.49	0	110 (1.0)	0.31	0.3
VOW	0.9 (1.4)	0.1 (0.1)	0.1 (0.2)	0.02 (0.04)	5 (2.5)	0.01	0.8	66 (2.3)	0.03	0.8	74 (2.4)	0.02	1.1
FRC	0.7 (0.8)	0.3 (0.6)	1.8 (4.0)	0.1 (0.2)	10.5 (1.8)	0.07	0.6	128.5 (0.4)	0.73	0.4	113.5 (0.9)	0.36	0.4
VOS	0.6 (0.5)	0.03 (0.04)	0.4 (0.6)	0.02 (0.04)	2 (2.9)	0.01	1.6	102.5 (1.1)	0.26	0.4	120 (0.8)	0.45	0.3
VLS	0.4 (0.5)	0.01 (0.02)	0.3 (0.5)	0.01 (0.03)	3 (2.8)	0.01	1.1	105 (1.1)	0.29	0.2	121.5 (0.9)	0.37	0
NSL	0.2 (0.3)	0.01 (0.03)	0.1 (0.2)	0.02 (0.1)	4 (2.7)	0.01	0.9	71 (2.2)	0.03	0.4	134 (0.3)	0.78	0.1

LIQ, liquid; VOW, vowel; FRC, fricative; VOS, voiced stop; VLS, voiceless stop; NSL, nasal; d, Cohen's effect size.

$Z = 1.75$; $p = 0.08$; $d = 0.3$); ASM ($U = 99$; $Z = 1.19$; $p = 0.23$; $d = 0.2$); ADD ($U = 130$; $Z = 0.19$; $p = 0.85$; $d = 0.5$).

In Figure 2, the compared profiles of RF, i.e., the percentage distribution, for the SYS subclasses of processes are shown. Figure 2A represents the profiles of WS1 and WS2 groups, which were very similar in terms of the percentage of the most frequent processes (CCR). In the subclasses of FCD and VCR processes, the profiles of both groups intersected since the WS2 group showed a relatively lower percentage of FCD and a correspondingly higher percentage of VCR, although the Mann–Whitney *U* test did not yield statistically significant differences: CCR ($U = 19$; $Z = 0.70$; $p = 0.48$; $d = 0.3$); FCD ($U = 11$; $Z = 1.73$; $p = 0.09$; $d = 1.1$); VCR ($U = 12$; $Z = 1.6$; $p = 0.11$; $d = 0.9$); SYD ($U = 20$; $Z = 0.58$; $p = 0.57$; $d = 0.3$); MTT ($U = 17$; $Z = 0.96$; $p = 0.34$; $d = 0.3$).

Figure 2B represents the compared profiles of normative groups TD1 and TD2 ($n = 36$). The profile of the TD2 group showed a relatively lower percentage in the most frequent subclasses of SYS processes (CCR, FCD) and in the less frequent subclass (MTT). In the three subclasses, the Mann Whitney *U* test showed statistically significant differences: CCR ($U = 533.5$; $Z = 1.94$; $p = 0.05$; $d = 0.5$); FCD ($U = 478$; $Z = 2.55$; $p = 0.01$; $d = 0.4$); MTT ($U = 501$; $Z = 2.54$; $p = 0.01$; $d = 0.1$). In addition, the profile of the TD2 group showed a relatively higher percentage of VCR and SYD processes, although no further statistically significant differences were observed: VCR ($U = 586.5$; $Z = 1.40$; $p = 0.16$; $d = 0.6$); SYD ($U = 619$; $Z = 1.08$; $p = 0.28$; $d = 0.3$).

Figure 2C represents the compared profiles of the WS1 and TD1 groups, where the WS1 group profile showed a

TABLE 5 Omission phonological processes index means and standard deviations for groups, Mann–Whitney *U* test, and effect size.

	WS1	WS2	TD1	TD2	WS1 vs. WS2			WS1 vs. TD1			WS2 vs. TD2		
	PI-M (*SD*)	PI-M (*SD*)	PI-M (*SD*)	PI-M (*SD*)	*U* (*Z*)	*p*	*d*	*U* (*Z*)	*p*	*d*	*U* (*Z*)	*p*	*d*
LIQ	1.3 (1.5)	0.1 (0.1)	0.6 (1.0)	0.03 (0.1)	8 (2.1)	0.04	1.1	79 (1.9)	0.06	0.5	45 (3.7)	0.001	0.7
VOS	0.4 (0.5)	0.1 (0.1)	0.2 (0.4)	0.04 (0.1)	11 (1.8)	0.07	0.8	99 (1.3)	0.21	0.4	111 (1.2)	0.23	0.6
VLS	0.4 (0.6)	0.004 (0.01)	0.1 (0.1)	0.002 (0.01)	4 (2.8)	0.01	0.9	53.5 (3.1)	0.01	0.7	124 (1.4)	0.17	0.2
VOW	0.3 (0.3)	0.02 (0.03)	0.03 (0.1)	0 (0)	6 (2.4)	0.02	1.3	38.5 (3.7)	0.001	1.2	80 (4.2)	0.001	0.7
NSL	0.3 (0.3)	0.01 (0.02)	0.03 (0.1)	0.004 (0.02)	33 (2.7)	0.01	1.4	48.5 (3.3)	0.001	1.2	127 (0.9)	0.36	0.3
FRC	0.1 (0.1)	0 (0)	0.02 (0.04)	0.008 (0.03)	35 (2.6)	0.01	1	64 (3.1)	0.01	1.1	126 (0.9)	0.39	0.3

LIQ, liquid; VOS, voiced stop; VLS, voiceless stop; VOW, vowel; NSL, nasal; FRC, fricative; d, Cohen's effect size.

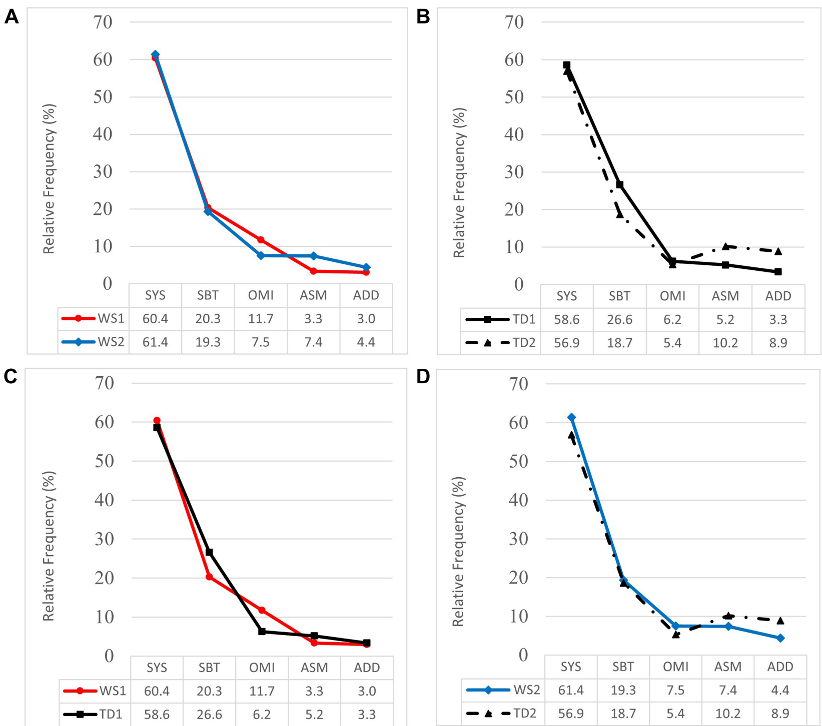


FIGURE 1
Profiles of relative frequency of processes by classes (SYS, syllable structure; SBT, substitution; OMI, omission; ASM, assimilation; ADD, addition) for WS groups and TD groups. **(A)** Profiles of WS1 and WS2 groups. **(B)** Profiles of TD1 and TD2 groups. **(C)** Profiles of WS1 and TD1groups. **(D)** Profiles of WS2 and TD2 groups.

relatively lower percentage of CCR processes and a relatively higher percentage of FCD and MTT processes. In the three subclasses, the Mann–Whitney *U* test yielded statistically significant differences: CCR ($U = 49$; $Z = 2.72$; $p = 0.01$; $d = 1.3$); FCD ($U = 60$; $Z = 2.39$; $p = 0.02$; $d = 1.2$); MTT ($U = 56$; $Z = 2.56$; $p = 0.01$; $d = 0.6$). Both profiles were similar in terms of the percentage of VCR and SYD processes: VCR ($U = 108$; $Z = 0.96$; $p = 0.34$; $d = 0.2$); SYD ($U = 108$; $Z = 0.96$; $p = 0.34$; $d = 0.1$).

Figure 2D represents the compared profiles of WS2 and TD2 groups ($n = 36$), where the profile of the WS2 group showed a relatively higher percentage of FCD processes and the Mann–Whitney *U* test yielded statistically significant difference:

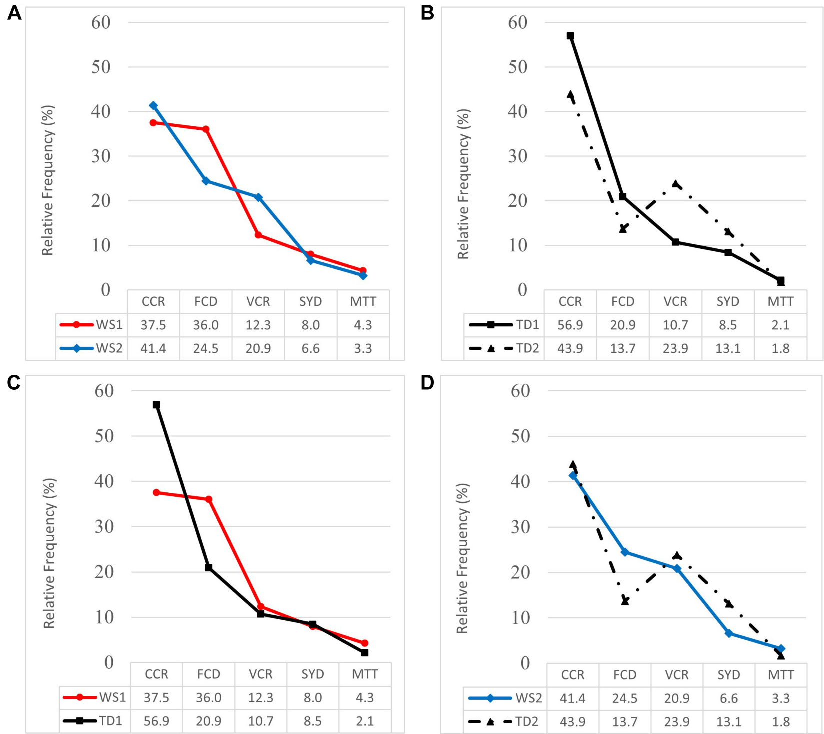


FIGURE 2
Profiles of relative frequency of syllable structure processes (CCR, consonant cluster reduction; FCD, final consonant deletion; VCR, vowel cluster reduction; SYD, unstressed syllable deletion; MTT, metathesis) for WS groups and TD groups. **(A)** Profiles of WS1 and WS2 groups. **(B)** Profiles of TD1 and TD2 groups. **(C)** Profiles of WS1 and TD1 groups. **(D)** Profiles of WS2 and TD2 groups.

FCD (U = 67.5; Z = 1.99; p = 0.05; d = 0.7). The most frequent subclasses of SYS processes in both groups were CCR and showed similar percentages, while the profile of the WS2 group presented relatively lower percentages of VCR, SYD, and a relatively higher percentage of MTT processes, although no further statistically significant differences were observed: CCR (U = 120.5; Z = 0.18; p = 0.86; d = 0.1); VCR (U = 110; Z = 0.53; p = 0.60; d = 0.1); SYD (U = 119.5; Z = 0.23; p = 0.82; d = 0.4); MTT (U = 83; Z = 1.79; p = 0.07; d = 0.3).

In **Figure 3**, the compared profiles of RF for SBT subclasses of processes are shown. **Figure 3A** represents the profiles of the WS1 and WS2 groups. In the WS2 group profile, a relatively lower percentage of VLS substitutions and a relatively higher percentage of NSL substitutions were observed, and the Mann–Whitney U test yielded statistically significant differences: VLS (U = 6; Z = 2.42; p = 0.02; d = 0.7); NSL (U = 7; Z = 2.33; p = 0.02; d = 0.2). Further intersections in the profiles of both groups were observed, since the WS2 group showed relatively lower percentages of LIQ and VOS substitutions, and correspondingly higher percentages of VOW and FRC substitutions. However, these differences were not statistically significant: LIQ (U = 22; Z = 0. 32; p = 0.75; d = 0.2); VOS (U = 12; Z = 1.62; p = 0.11; d = 0.8); VOW (U = 24; Z = 0.06; p = 0.95; d = 0.4); FRC (U = 24; Z = 0; p = 1.0; d = 0.3).

Figure 3B represents the compared profiles of normative groups TD1 (n = 37) and TD2 (n = 25), where the TD2 profile showed relatively lower percentages of FRC, VOS, and VLS substitutions. In the three subclasses, the Mann–Whitney U test showed statistically significant differences: FRC (U = 329.5; Z = 1.93; p = 0.05; d = 0.3); VOS (U = 324.5; Z = 2.03; p = 0.04; d = 0.1); VLS (U = 241.5; Z = 3.42; p = 0.001; d = 0.8). The most frequent processes in both groups were LIQ substitutions showing similar percentages, while in the profile of the TD2 group relatively higher percentages of VOW and NSL substitutions were observed. However, these differences were not statistically significant: LIQ (U = 383. 5; Z = 1.14; p = 0.25; d = 0.01); VOW (U = 389; Z = 1.11; p = 0.27; d = 0.2); NSL (U = 433; Z = 0.49; p = 0.63; d = 0.5).

Figure 3C represents the compared profiles of the WS1 and TD1 groups (n = 37), where the WS1 profile presented relatively higher percentages of VOW and NSL substitution processes. In both subclasses, the Mann–Whitney U test showed statistically significant differences: VOW (U = 51; Z = 2.55; p = 0.01; d = 0.5); NSL (U = 72; Z = 1.95; p = 0.05; d = 0.4). The most frequent processes in both groups were LIQ substitutions showing similar percentages, while in the profile of the WS1 group a relatively lower percentage of FRC substitutions was observed. However, no further statistically significant differences were observed: LIQ

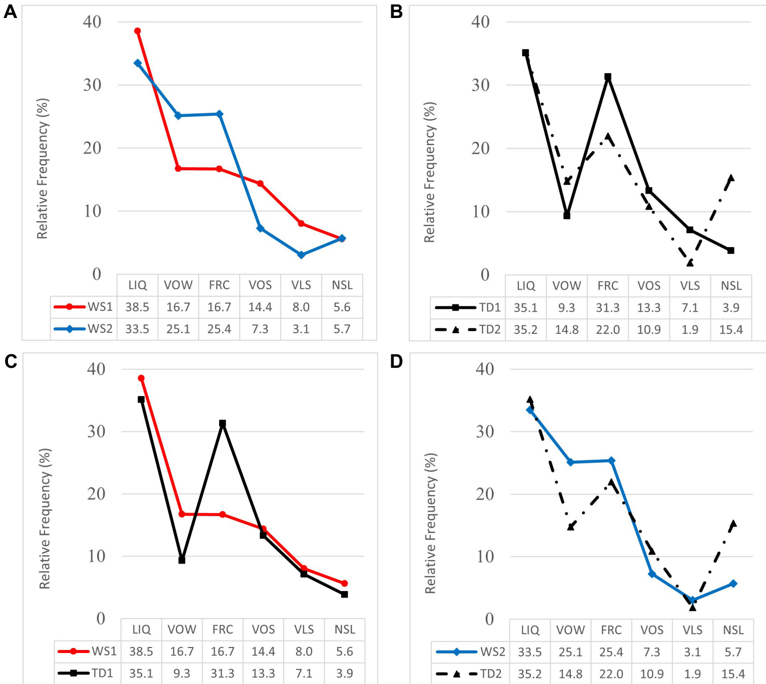


FIGURE 3
Profiles of relative frequency of substitution processes (LIQ, liquid; VOW, vowel; FRC, fricative; VOS, voiced stop; VLS, voiceless stop; NSL, nasal) for WS groups and TD groups. **(A)** Profiles of WS1 and WS2 groups. **(B)** Profiles of TD1 and TD2 groups. **(C)** Profiles of WS1 and TD1 groups. **(D)** Profiles of WS2 and TD2 groups.

(U = 120; Z = 0. 31; p = 0.76; d = 0.1); FRC (U = 87; Z = 1.37; p = 0.17; d = 0.7); VOS (U = 92; Z = 1.19; p = 0.23; d = 0.1); VLS (U = 107; Z = 0.73; p = 0.47; d = 0.1).

Figure 3D represents the compared profiles of the WS2 and TD2 groups (n = 25), which were similar in terms of the percentage of the most frequent processes (LIQ). The profile of the WS2 group showed a relatively higher percentage of VOW substitutions and a relatively lower percentage of NSL substitutions. However, the Mann–Whitney U test did not yield statistically significant differences: LIQ (U = 79; Z = 0.41; p = 0.69; d = 0.1); VOW (U = 60; Z = 1.38; p = 0.17; d = 0.4); FRC (U = 80; Z = 0.36; p = 0.72; d = 0. 2); VOS (U = 85.5; Z = 0.10; p = 0.92; d = 0.2); VLS (U = 80; Z = 0.47; p = 0.64; d = 0.2); NSL (U = 75; Z = 0.75; p = 0.45; d = 0.4).

In Figure 4, the compared profiles of RF for the OMI subclasses of processes are shown. Figure 4A represents the profiles of WS1 and WS2 groups (n = 6), where the WS2 profile presented a relatively lower percentage of the less frequent OMI processes (VLS, NSL, FRC). In these subclasses, the Mann–Whitney U test yielded statistically significant differences: VLS (U = 4.5; Z = 2.48; p = 0.01; d = 1.6); NSL (U = 8.5; Z = 1.88; p = 0.06; d = 0.9); FRC (U = 6; Z = 2.44; p = 0.02; d = 0.7). In contrast, the WS2 profile showed relatively higher percentages of the most frequent OMI processes (LIQ, VOS, VOW). However, in these subclasses, no statistically significant differences were found: LIQ (U = 18; Z = 0.43; p = 0.67; d = 0.4); VOS (U = 20; Z = 0.15; p = 0.88; d = 0.2); VOW (U = 14.5; Z = 0.94; p = 0.35; d = 0.1).

Figure 4B represents the compared profiles of normative groups TD1 (n = 30) and TD2 (n = 15). In the TD2 profile, a relatively lower percentage of VOW omissions was observed, where the Mann–Whitney U test yielded statistically significant differences: VOW (U = 157.5; Z = 2.33; p = 0.02; d = 0.4). Further intersections in the profiles of both groups were observed, since the TD2 group showed relatively lower percentages of LIQ omissions, and correspondingly higher percentages of VOS and FRC omissions. However, these differences were not statistically significant: LIQ (U = 155.5; Z = 1.71; p = 0.09; d = 0.5); VOS (U = 215; Z = 0.25; p = 0.81; d = 0.3); VLS (U = 176.5; Z = 1. 61; p = 0.11; d = 0.02); NSL (U = 192.5; Z = 1.04; p = 0.30; d = 0.1); FRC (U = 202.5; Z = 0.75; p = 0.46; d = 0.5).

Figure 4C represents the compared profiles of WS1 and TD1 groups (n = 30), where the WS1 profile showed relatively higher percentages of VOW, VLS, and NSL omission processes, and a relatively lower percentage of FRC omissions. In these subclasses, the Mann–Whitney U test yielded statistically significant differences: VOW (U = 49; Z = 2.44; p = 0.02; d = 0.3); VLS (U = 56.5; Z = 2.12; p = 0.03; d = 0.2); NSL (U = 49.5; Z = 2.42; p = 0.02; d = 0.4); FRC (U = 56; Z = 2.35; p = 0.02; d = 0.1). The most frequent processes in both groups

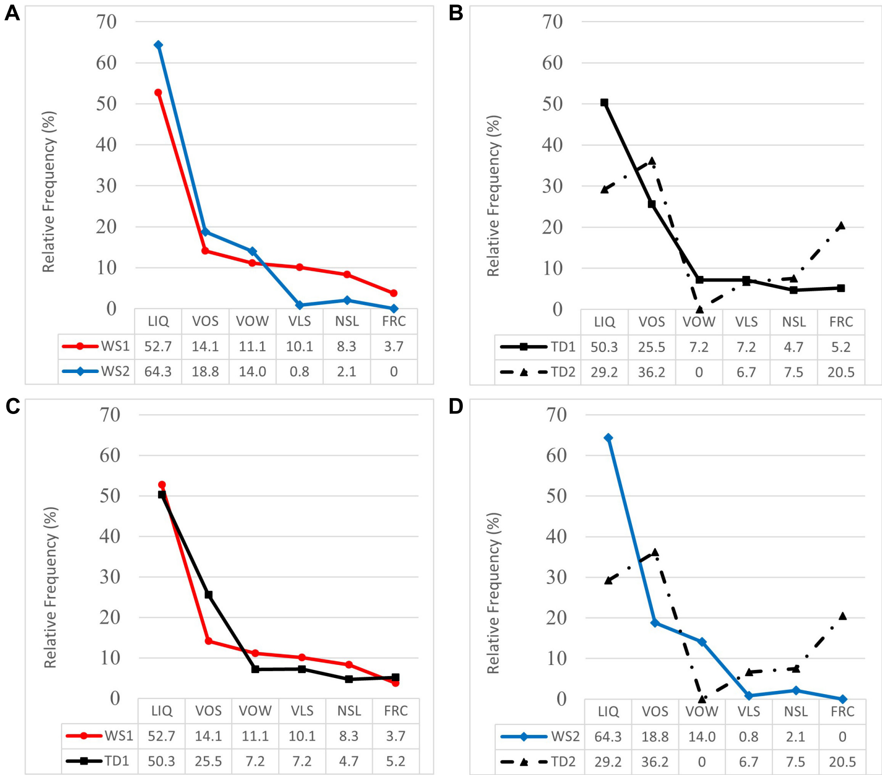


FIGURE 4
Profiles of relative frequency of omission processes (LIQ, liquid; VOW, vowel; FRC, fricative; VOS, voiced stop; VLS, voiceless stop; NSL, nasal) for WS groups and TD groups. **(A)** Profiles of WS1 and WS2 groups. **(B)** Profiles of TD1 and TD2 groups. **(C)** Profiles of WS1 and TD1 groups. **(D)** Profiles of WS2 and TD2 groups.

were LIQ omissions, with similar percentages, while the profile of the WS1 group presented a relatively lower percentage of VOS omissions, although no statistically significant differences were observed: LIQ ($U = 104$; $Z = 0.02$; $p = 0.98$; $d = 0.1$); VOS ($U = 86.5$; $Z = 0.72$; $p = 0.47$; $d = 0.6$).

Figure 4D represents the compared profiles of the WS2 ($n = 6$) and TD2 ($n = 15$) groups, where the profile of the WS2 group showed relatively higher percentages of LIQ and VOW omission processes. In both subclasses, the Mann–Whitney U test yielded statistically significant differences: LIQ ($U = 20.5$; $Z = 2.00$; $p = 0.05$; $d = 1.0$); VOW ($U = 22.5$; $Z = 2.88$; $p = 0.01$; $d = 0.5$). The WS2 profile showed relatively lower percentages of VOS, VLS, NSL, and FRC omissions, although no further statistically significant differences were observed: VOS ($U = 38.5$; $Z = 0.60$; $p = 0.55$; $d = 0.5$); VLS ($U = 41$; $Z = 0.61$; $p = 0.54$; $d = 0.3$); NSL ($U = 44$; $Z = 0.13$; $p = 0.90$; $d = 0.3$); FRC ($U = 33$; $Z = 1.37$; $p = 0.17$; $d = 0.5$).

Cluster analysis

In Figure 5, the clusters membership (solutions for 2, 3, and 4 clusters) indicate the individual similarities and differences in the RF profiles of classes of processes within the WS1 and WS2 groups. Figure 5A shows that, in the WS1 group, the profiles of cases 1, 2, 5, and 6 represent the modal profile, i.e., the predominant patterns. Cases 3, 4, and 7 diverge from that profile in two directions: they present a higher percentage of SYS processes diverging from the TD1 group; they also present a lower percentage of OMI processes converging with the TD1 group.

Figure 5B represents the clusters membership in the WS2 group, where it is observed that the profiles of cases 9, 10, 13, and 14 represent the modal profile. Cases 8, 11, and 12 diverge from that profile, diverging from the TD2 group, by presenting a higher percentage of SYS processes and a lower percentage of SBT processes.

In Figure 6, the clusters membership of the RF profiles of SYS subclasses of processes are shown. Figure 6A shows that the profiles of cases 2, 3, 4, and 6 represent the modal profile of the WS1 group, while case 7 is an extreme case because of its high percentage of FCD. Cases 1 and 5 diverge from the TD1 group by their lower percentage of CCR and a higher percentage of VCR processes.

Figure 6B represents the membership clusters in the WS2 group, where it is observed that the profiles of cases 11, 13,

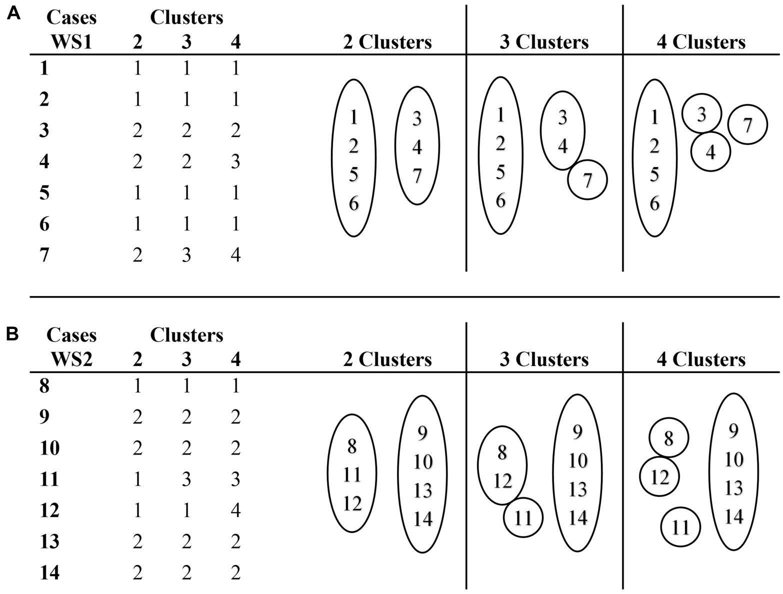


FIGURE 5
Cluster membership for a range of solutions (2, 3, and 4 clusters) for classes of processes. **(A)** Cluster membership of cases in WS1 group. **(B)** Cluster membership of cases in WS2 group.

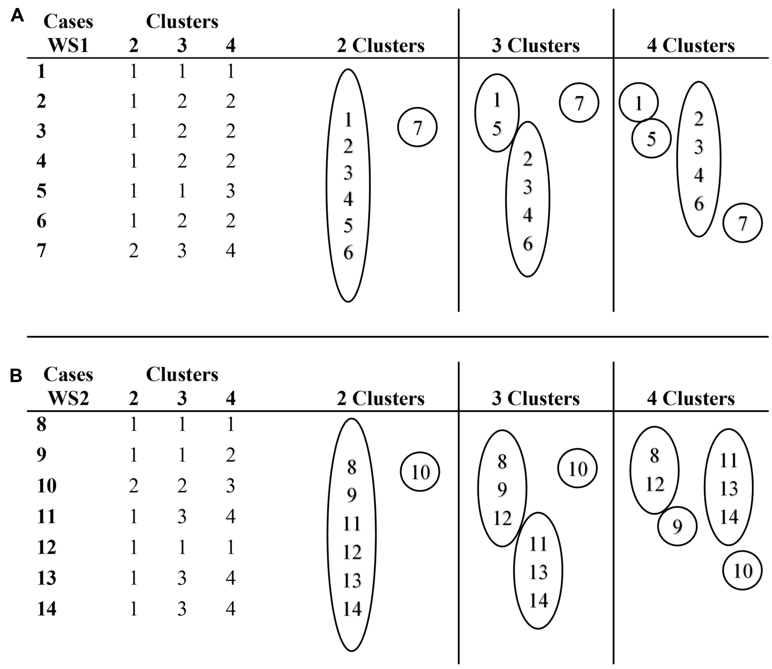


FIGURE 6
Cluster membership for a range of solutions (2, 3, and 4 clusters) for syllable structures processes. **(A)** Cluster membership of cases in WS1 group. **(B)** Cluster membership of cases in WS2 group.

and 14 represent the modal profile, while case 10 is an extreme case, due to its low percentage of CCR, and its high percentage of FCD and VCR. Cases 8, 9, and 12 are separated from this profile by their higher percentage of CCR, diverging from the TD2 group, and a lower percentage of FCD, converging with the TD2 group.

In **Figure 7**, the membership clusters of the RF profiles of SBT subclasses of processes are shown. **Figure 7A** shows that

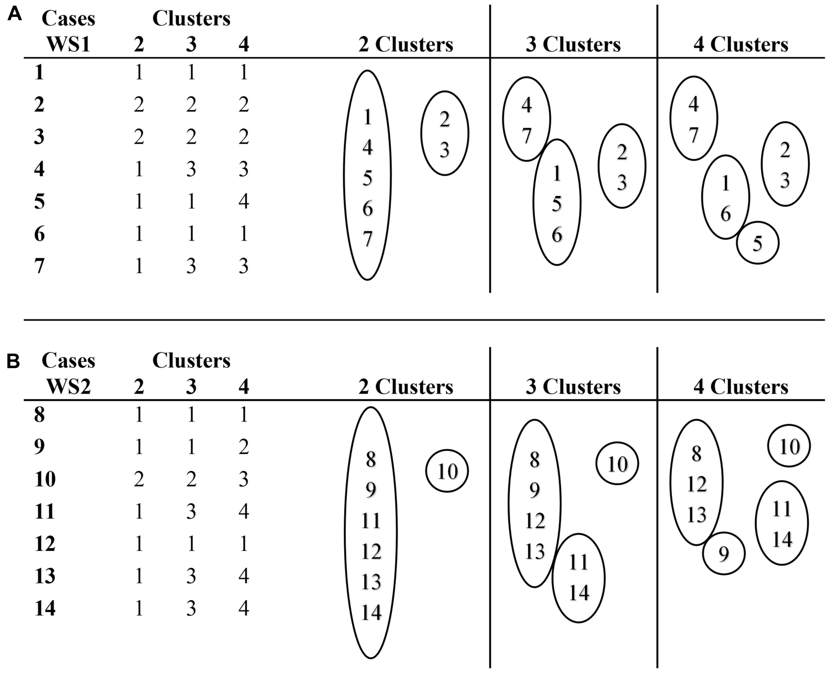


FIGURE 7
Cluster membership for a range of solutions (2, 3, and 4 clusters) for substitution processes. **(A)** Cluster membership of cases in WS1 group. **(B)** Cluster membership of cases in WS2 group.

the profiles of cases 1, 5, and 6 represent the modal profile of the WS1 group. Cases 2 and 3 present an additional profile, which diverges from the TD1 group by its lower percentage of LIQ substitutions and a higher percentage of VLS substitutions; and converges with the TD1 group by its higher percentage of FRC substitutions. Cases 4 and 7 present an additional profile, which diverges from the TD1 group due to its higher percentage of LIQ substitutions.

Figure 7B represents the membership clusters of the WS2 group participants, where it is observed that the profiles of cases 8, 12, and 13 represent the modal profile. Case 10 is an extreme case because of its high percentage of VOW substitutions processes, and case 9 is also an extreme case because it only presents LIQ substitutions. Cases 11 and 14 present an additional profile, which diverges from the TD2 group by the absence of LIQ substitutions and by its higher percentage of FRC substitutions.

In **Figure 8**, the membership clusters of the RF profiles of the OMI subclasses of processes are shown. **Figure 8A** shows that the profiles of cases 1, 2, 3, 4, 5, and 6 represent the modal profile of the WS1 group. Case 7 is an extreme outlying case because it only presents LIQ omissions processes. Case 3 is also an extreme outlying case, because of its low percentage of LIQ omissions, and its high percentage of VOS, VOW, and FRC omissions.

Figure 8B represents the membership clusters of the WS2 group participants, where it is observed that the profiles of cases 10, 12, and 13 represent the modal profile. Case 14 is an extreme outlying case, because of its high percentage of VOW omissions processes. Cases 8 and 11 present an additional profile, which converges with the TD2 group for its lower percentage of LIQ omissions and its higher percentage of VOS omissions.

Discussion

The purpose of this study was to explore late phonological development in individuals with WS by comparing the profiles of a group of children (aged 3–8 years) and a group of adolescents and adults (aged 14–25 years). To determine if they followed the stages of typical development and if they presented specific characteristics, they were also compared with the profiles of TD children in two phonological stages: expansion stage (aged 3 years) and resolution stage (aged 5 years). The profiles were based on the classes and subclasses of processes, calculating their PI (frequency of processes/100 words) and their RF (percentage distribution). Additionally, modal profiles and outliers were explored by cluster analysis.

Stages in late phonological development

The results of the cross-sectional comparison between the group of children and older individuals with WS suggest a late

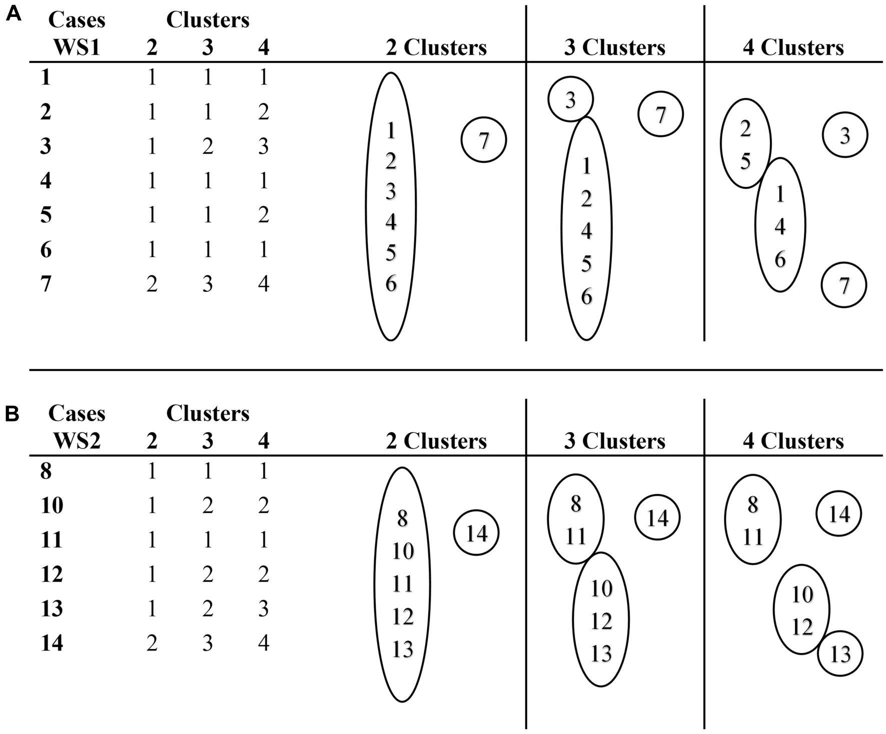


FIGURE 8
Cluster membership for a range of solutions (2, 3, and 4 clusters) for omission processes. **(A)** Cluster membership of cases in WS1 group. **(B)** Cluster membership of cases in WS2 group.

phonological developmental trajectory in which some processes persist into adolescence and adulthood. Children with WS presented a higher frequency of phonological processes in most classes and subclasses than older individuals, which is consistent with previous research (Huffman, 2019). The profiles of both groups were comparable respectively to those of 3- and 5-year-old children in the normative groups, so it could be interpreted that they were in different stages of late phonological development, i.e., the group of children with WS would be in the initial stage of expansion and the group of older individuals would be in the final stage of resolution, according to the chronology established for late phonological development in Spanish (Diez-Itza et al., 2001; Diez-Itza and Martínez, 2004).

The dynamics observed for phonological development also suggest that both groups are at different stages. The frequency of processes in the group of children with WS tended to decrease with chronological age, suggesting that phonological development occurs at a certain rate at this stage, which is not inconsistent with the findings previously reported by Martínez et al. (2014) in two children with WS. This rate of phonological development would compensate for the delay in language onset, which in turn has been related to delayed babbling (Masataka, 2001) and auditory-visual integration difficulties observed in infants and toddlers with WS and other neurodevelopmental syndromes (D'Souza et al., 2015). However, it remains unclear why syndromes follow quite different trajectories of phonological development (Huelmo et al., 2017; Hidalgo and Garayzábal, 2019; Diez-Itza et al., 2021).

In the case of WS, rapid outcomes during the stage of phonological expansion could be favored by an acceleration of lexical development, which initially presents an atypical trajectory where declarative gesture (pointing) is delayed about 6 months in relation to first words. Unlike in typical developing, it is not the onset of first words but the age of acquisition of pointing that best predicts the lexical development of children with WS at 4 years of age, and it also seems to mark the beginning of a necessarily accelerated reconvergence to the trajectory of typical development (Becerra and Mervis, 2019). The recovery of the rate of typical phonological development could be explained in the same way, given the close relationship between lexical and phonological development, and their interrelation with central cognitive processes, such as verbal working memory, reasoning ability, and verbal STM (Mervis et al., 2004; Stoel-Gammon, 2011).

Nevertheless, our results suggest that reconvergence during the expansion stage is not maintained over time, since the phonological profiles in the group of adolescents and adults with WS tended to be progressively divergent compared to those of TD children in the expansion and resolution stages of late phonological development. Moreover, the frequency of phonological processes in these older individuals with WS was not significantly correlated with chronological age and processes persisted in most classes, suggesting that the resolution stage is not completed during adolescence and adulthood in WS. At these ages, asynchronies might be more evident, since phonological production accuracy in older individuals with WS was below that expected for 6-year-old TD children, while their lexical verbal age was close to that expected for 10-year-old TD children, the age at which phonological acquisition can be considered complete.

In any case, the results of the present study indicate that phonology is not fully preserved in WS and should not be considered a relative strength compared to lexical development, as some initial studies had suggested (Udwin and Yule, 1990; Volterra et al., 1996). Similar results have been reported in the case of morphology, which also leads to question its intactness and typicality (Diez-Itza et al., 2017, 2019).

The persistence of phonological processes in adolescents and adults with WS could be related to the atypical phonological processing reported in previous studies (Majerus et al., 2003; Majerus, 2004). In this regard, Huffman (2019) also found that articulatory accuracy was closely associated with phonological processing, intellectual abilities, and lexical abilities. The strength in pseudoword repetition tests, which are at the level expected for chronological age, suggests that STM is not impaired in WS, unlike in Down syndrome (Jarrold and Baddeley, 2001). However, the persistence of processes might be consistent with the hypothesis of a dissociation between short-term and long-term memory in the verbal domain (Vicari et al., 1996b). Previous results, including also phonological awareness tasks, point to more complex cognitive, prosodic, and lexical factors, which determine less finely grained and abnormally structured phonological and lexical representations (Laing et al., 2001; Böhning et al., 2002; Majerus et al., 2003; Stojanovik, 2010). In TD children and adults, links between cognitive and linguistic processing demands and speech motor performance have been identified, whereby the phonological processes observed may also be related to oral-motor difficulties that adolescents and adults with WS still present (Nip et al., 2009; Krishnan et al., 2015).

Quantitative differences: Frequency of processes

Participants in the WS groups showed a higher frequency of omission than the children in the normative TD groups, including deletion of singletons both in onset syllable positions and in final word coda positions, which may have additional morphophonological developmental implications (Levy and Eilam, 2013). Spanish has a complex morphology where omissions or substitutions of final word sounds may have an impact on inflection in most word categories, especially in verbs. In fact, adolescents with WS presented a higher frequency of morphological omission errors than 5-year-old TD controls matched on verbal age (Diez-Itza et al., 2019).

The children with WS tended to omit all phoneme subclasses more frequently than 3-year-old TD children, although the differences were not statistically significant for voiced stop and liquid consonants, which was unexpected considering that unvoiced stops are less marked and cross-linguistically earlier acquired, i.e., less complex (McLeod and Crowe, 2018). In previous studies, mismatch patterns in tautosyllabic consonant clusters were more common when C1 was voiced (in Spanish, 13 tautosyllabic consonant clusters are possible: /p, t, k, b, d, g, f/ + /liquid/). Voiced stops are more marked and, from a sonority hierarchy approach, closer to C2 liquid consonants, therefore the cluster reduction patterns were considered to follow the principle of retaining the less sonorous consonant (Pérez et al., 2018; Vergara et al., 2021).

The higher frequency of vowel omission observed in the two groups with WS compared to their respective normative TD groups may be considered an atypical feature. In addition, a significantly higher frequency of liquid consonant omissions in the group of adolescents and adults with WS than in the 5-year-old TD normative group may suggest a deviant developmental trajectory. In this group, frequency of omissions of voiceless stop, nasal and fricative consonants may be interpreted as reconverging with the normative group, thus also following a non-linear trajectory of phonological development. The results of this study were partially consistent with those of Diez-Itza et al. (2021) who observed that children and adolescents with Down syndrome presented atypically more omission processes than their 3-year-old TD controls. A substantial portion of the segmental omissions corresponded to codas in medial and final position, which were significantly more frequent in participants with Down syndrome.

The frequency of metathesis was also atypically higher in the group of children with WS compared to the 3-year-old TD children, while in the group of older individuals it no longer differed from the 5-year-old TD children. Early case studies of children with WS have already referred to examples of metathesis as distinctive phonological errors of this syndrome, which was also documented when compared to other syndromes (Volterra et al., 1996; Diez-Itza et al., 1998; Hidalgo, 2019; Hidalgo and Garayzábal, 2019).

It is important to point out the possible effect of the elicitation method, as Diez-Itza et al. (2021) observed that children and adolescents with Down syndrome presented a higher tendency for the omission of segments in spontaneous speech than in articulation tests. Conversely, they found a lower tendency for substitutions in spontaneous speech, consistent with the findings of the present study, where participants with WS did not differ from TD children in consonant substitutions. Nevertheless, they presented a significantly higher frequency of vowel substitutions than TD children, which can be also considered an atypical feature in WS, since single vowels (i.e., simple syllable nuclei) usually appear to be already acquired in Spanish typical late phonological development (Diez-Itza et al., 2001). The striking fact that the study by Hidalgo (2019) did not observe final coda omissions (i.e., final consonant deletions) in any participant with WS also suggests greater facilitation in whole-word structure production when it is elicited through tests of articulation. The tendency for omissions observed in the participants with WS in the present study could therefore be related to the elicitation method, since spontaneous speech involves prosodic, articulatory, and linguistic planning factors quite different from picture naming. Nonetheless, production errors are much less frequent in WS than in Down syndrome, so speech intelligibility is rarely affected in this population, which usually shows a slowed speech rate (Semel and Rosner, 2003; Kumin, 2006; Setter et al., 2007; Crawford et al., 2008; Barnes et al., 2009; Hargrove et al., 2012). In any case, an early and continued speech therapy intervention that addresses specific problems in phonological production of people with WS and an improvement of the home literacy environment, also considering speech rate, should not be omitted (Mervis and Velleman, 2011; Ranzato et al., 2021).

Qualitative differences: Profiles of percentage distribution of processes

In addition to the quantitative differences observed, the study of relative frequencies further suggests that late phonological development in WS may not follow an entirely linear trajectory. Intersections between the profiles of relative frequency between the groups of children and older individuals with WS might suggest that the trajectories from the first stage to the final stage of late phonological development is toward reduction in the proportion of omissions and increase in the proportion of assimilation and addition processes. Such hypothesized trajectories are in line with the observed differences between TD normative groups and, therefore, the profiles of adolescents and adults with WS and 5-year-old TD children appeared to be quite close, suggesting that the trajectory observed in WS would correspond to the typical evolution from the expansion stage to the resolution stage in late phonological development. These results are consistent in part with those of Martínez and Diez-Itza (2012), who observed that assimilations tended to persist as errors of processing in the last stage of typical development.

However, the profiles of relative frequency presented an atypically higher percentage of omissions in the group of children with WS than in the normative group of 3-year-old TD children. Accordingly, the percentage of substitutions in the WS1 group was low, so that the profiles of both groups intersected at those points, suggesting that the children with WS may still be in an earlier stage, since in typical late phonological development an emergent process is observed in which substitutions tend to increase and omissions to decrease during the expansion stage (Vergara et al., 2021).

Differences became more apparent in the profiles of the subclasses of processes. Syllable structure subclasses showed intersecting profiles in the WS groups, with a relatively lower percentage of final consonant deletion and a relatively higher percentage of vowel cluster reduction (diphthongs) in the group of adolescents and adults with WS, which is in line with the profiles of the TD normative groups in the respective stages of expansion and resolution.

However, striking asymmetries were also found in the profiles of the WS groups when compared with the TD groups: the WS children presented a much lower percentage of consonant cluster reduction than the 3-year-old TD children, which contrasts with the high percentage of final consonant deletion. It is important to note that the present study as the previous one by Martínez (2010) included both tautosyllabic and heterosyllabic consonant clusters, which were also fully described in Diez-Itza and Martínez (2004). In contrast, a more recent study including non-linear analyses, where a brief description of the Spanish phonological system can be found, focused only in tautosyllabic clusters (Vergara et al., 2021). The observed profile of early acquisition of consonant clusters indicates an asynchronous development since the accurate production of consonant clusters is typically protracted in late phonological development and it is commonly impaired in speech disorders (McLeod et al., 2001; Pérez et al., 2018; Vergara et al., 2021). The profile of adolescents and adults with WS reconverges in this respect with that of the group of 5-year-old TD children, suggesting an atypical trajectory (Becerra and Mervis, 2019). However, the persistent deletion of final consonants remains a divergent feature in the profile of older individuals with WS, and this should be investigated in relation to the atypical morphophonological difficulties noted in some studies (Levy and Eilam, 2013; Diez-Itza et al., 2017).

There were also marked differences in the profiles when the relative frequencies of substitutions and omissions were analyzed. In the group of adolescents and adults with WS, a lower percentage was observed in the substitutions of voiceless stops, with a higher percentage of processes in voiced phonemes, in concordance with a typical trajectory also observed in the profiles of the TD controls (Vergara et al., 2021). The profiles of children with WS showed a higher percentage of vowel substitutions than those of 3-year-old TD children, which might be considered an atypical feature, as studies suggest that single vowels are acquired in the early stages of phonological development (Smith, 1973; Bosch, 2004). A relatively high percentage of vowel substitution processes is maintained in the WS2 group, although it also corresponds to a relative increase of vowel substitutions in the normative TD2 group. This observation is consistent with that of Donegan (2013), who suggests that there is greater vowel variability in children than is usually considered and this is explained by both phonetic and prosodic factors. It seems that vowels play a different role than consonants in language acquisition and they are related to prosody and the organization of syntactic constituents (Hochmann et al., 2011), so vowel substitution processes may be associated with the prosodic difficulties observed in WS (Stojanovik, 2010; Martínez-Castilla et al., 2012).

Regarding the subclasses of omission processes, the profiles of older individuals with WS showed significant reductions in the percentages of omission of voiceless stop, nasal, and fricative consonants, suggesting non-linear trajectories across stages of phonological development. In addition, the profiles of children with WS showed a lower percentage of fricative consonant omissions and higher percentages of omission of single vowels, voiceless stops, and nasals than those observed in the normative 3-year-old TD group, which again points to atypical features in WS late phonological development. The older individuals with WS presented a profile of relative frequency of omissions that also diverges from that of TD 5-year-old children, where higher percentages of liquid consonant and single vowel omissions were observed.

The results of the present study therefore reveal that, beyond the observed parallels, which suggested different stages and non-linear trajectories in late phonological development in both WS and TD, partially deviant profiles also appear when comparing the relative frequencies of the processes of the WS groups and their respective normative TD groups. These qualitative differences could be interpreted as atypical patterns in the profiles of individuals with WS with respect to what would be expected based on the stages of phonological development. Thus, phonological development across late stages might not be explained merely as a delay, i.e., only in terms of quantitative differences in frequency of processes based on chronological age. Atypical trajectories of development in individuals with WS and cross-syndrome differences have also been described at other levels of language such as morphology, prosody, lexical abilities, and pragmatics (Thomas and Karmiloff-Smith, 2003; Levy and Eilam, 2013; Diez-Itza et al., 2019, 2022).

Furthermore, adolescents and adults with WS, while they are in some respects at the same stage of resolution as 5-year-old TD children, exhibit an asynchronous and atypical persistence of certain processes suggesting that they have completed late phonological development without full mastery of phonology. This may be due to atypical phonological processing, inaccurate representations in long-term memory, or factors related to oral-motor development that require further investigation (Laing et al., 2001; Böhning et al., 2002; Majerus et al., 2003; Majerus, 2004, Nip et al., 2009; Krishnan et al., 2015).

Individual differences and modal profiles: Cluster analyses

The individual profiles based on relative frequencies of processes were compared by cluster analysis and it was observed that most individuals with WS presented modal profiles, i.e., adjusted to the mean of their group, for the different classes of processes. However, important individual differences also emerged, as previous studies had observed in conversations of people with WS (Stojanovik, 2006). These differences were in the direction of greater divergence from the profiles of the normative TD groups, expanding the atypical features of late phonological development in WS. Moreover, modal profiles were not always represented in the different classes of processes by the same participants, indicating a great complexity where individual differences interact with developmental trajectories.

In the more detailed analysis of the differences and similarities in the individual profiles of the subclasses of processes, it was observed that most of the children with WS presented a modal profile in syllable structure and omission processes. However, the case of the oldest participant in this group was an outlier in both classes of processes, diverging from the profiles of the normative TD group and the older participants with WS. She also diverged from the group in the profile of substitution processes, being the only case outside the modal profile in all classes and subclasses of processes. This may be interpreted as suggesting that because of her older chronological and verbal age she no longer represents the first stage of late phonological development in WS, but perhaps the intermediate stage of stabilization that was not captured in the present study.

The group of adolescents and adults with WS showed greater heterogeneity so that, in the profiles of all the subclasses of processes, the modal group did not include a majority of cases. This could be related to the fact that they are at a different stage of development and have a wider range of chronological age and verbal age. Among the extreme cases, the most outstanding were: the one with the highest verbal age who showed very atypical profiles of syllable structure and substitution processes; the one with the highest chronological age and the lowest verbal age, with a high proportion of vowel omission processes, typical of earlier stages; and the only case that was not included in any of the modal profiles and that also only presented substitution of liquid consonants. These results suggest that verbal age is a factor that may determine not only quantitative differences in phonological production but also greater complexity and qualitative differences in the classes and subclasses of processes. However, non-verbal abilities, gender, or schooling could also account for individual differences.

Limitations of the study

It is necessary to recognize several limitations in the present study. The sample size was small due to the difficulty of recruiting participants with this relatively rare syndrome and of applying a naturalistic methodology, more complex than the use of articulation tests, although it might be considered sufficient for a first exploratory study, taking into account that the word samples analyzed were large (more than 40,000 word tokens, and almost 10,000 word types). As shown by the analyses of individual differences, the chronological and verbal age ranges are too wide and not having separated groups for adolescents and adults is also a limitation. Future studies should better adjust the age of the groups and exclude atypical cases. The computation of the frequency of processes on the total number of tokens, instead of on the total number of word types, although providing control over the size of the individual samples analyzed, may in some cases overestimate the phonological index, computing the same error several times, or underestimate it, in those cases with more lexical diversity. Although the procedure followed in transcription and coding assured a high level of interrater reliability, the study lacked a numerical index to properly account for this potential source of error. Since the elicitation method could influence the results, it should be recommended to combine spontaneous speech assessment with articulation tests in future studies. Non-verbal abilities (e.g., short and long-term memory) and other factors such as word frequency and word length also may have had effects on phonological production that were not controlled for in this study, although it should be noted that the use of the words in spontaneous speech guarantees that they are part of the vocabulary of the participants.

Conclusion

The present exploratory study of late phonological development in WS suggested that children between 3 and 8 years of age and adolescents and adults between 14 and 25 years of age are at different stages of late phonological development. The frequency of phonological processes of the group of children with WS was comparable to that of 3-year-old TD children, implying that both would be at the same first stage of late phonological development (namely, the expansion stage). Older individuals with WS presented a much lower frequency of processes, similar to that of 5-year-olds in the last stage of phonological development (namely, the resolution stage). However, phonology no longer seems to be developing in the adolescents and adults with WS, whose phonological processes would therefore be persistent and independent of chronological age. Moreover, their marked age asynchrony of more than fourteen years with the TD normative group does not make it suitable to describe these persistent phonological difficulties in terms of delayed or protracted phonological development, nor the fact that they presented a frequency of phonological processes above that expected for their verbal lexical age. In contrast, children with WS showed a certain rate of phonological development that tends to bring them closer to the level expected for their verbal age, although with an asynchrony of almost 3 years below their chronological age.

These asynchronies are associated with atypical features in the phonology of individuals with WS that were revealed in both quantitative (frequency) and qualitative (proportion) assessments of phonological processes. Although the profiles were partially coincident with those of TD children, they also presented specific features, which were more evident when the subclasses of processes were analyzed in detail. The analysis of the underlying processes, especially in substitutions and omissions, revealed specific and complex phonological profiles in both groups of participants with WS. Individual differences tended to increase the divergence from the typical developmental profiles, being more salient in the group of adolescents and adults with WS, although participants who suited the average profile of the group predominated. The greater tendency to omissions in all syllable positions, including final codas, can be considered atypical and characteristic of WS at all ages, and may also be related to morphological processes. However, it is possible that this finding was in part influenced by the elicitation method based on spontaneous speech, as has been observed in Down syndrome.

The results of this study, although requiring further research, provide some new insight into atypical and dynamic phonological developmental trajectories in WS. Chronological and verbal age account for individual differences in phonological production, although other variables including short and long-term memory should be analyzed in future studies. The findings may also have clinical implications for speech intervention in this population requiring continued specific assessment and treatments adapted to the emerging characteristics of their phonological profiles throughout childhood, adolescence, and into adulthood.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by the Ethical Committee in Research of the University of Oviedo. Written informed consent to participate in this study was provided by the participants or their legal guardian/next of kin.

Author contributions

ED-I had a primary role in the conception and design of the study, development of the coding scheme, data analysis and discussion, and drafting of the manuscript. VP and VM helped with the design and conducted the research, carried out the transcription, coding, data analyses, and helped draft the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This research was supported by grant FFI2012-39325-C03-03 from the Spanish Ministry of Economy and Competitiveness (MINECO) to the SYNDROLING Project.

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.

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

Abbeduto, L., Benson, G., Short, K., and Dolish, J. (1995). Effects of sampling context on the expressive language of children and adolescents with mental retardation. *Ment. Retard.* 33, 279–288.

Aguilar, E., and Serra, M. (2003). *A-RE-HA: Análisis del retraso del habla: Protocolos para el análisis de la fonética y la fonología infantil.* Barcelona: Universitat de Barcelona.

Barnes, E., Roberts, J., Long, S. H., Martin, G. E., Berni, M. C., Mandulak, K. C., et al. (2009). Phonological accuracy and intelligibility in connected speech of boys with Fragile X syndrome or Down syndrome. *J. Speech Lang. Hear. Res.* 52, 1048–1061. doi: 10.1044/1092-4388(2009/08-0001)

Becerra, A. M., and Mervis, C. B. (2019). Age at onset of declarative gestures and 24-month expressive vocabulary predict later language and intellectual abilities in young children with Williams syndrome. *Front. Psychol.* 10:2648. doi: 10.3389/fpsyg.2019.02648

Bellugi, U., Lichtenberger, L., Jones, W., Lai, Z., and St. George, M. (2000). I. The neurocognitive profile of Williams syndrome: A complex pattern of strengths and weaknesses. *J. Cogn. Neurosci.* 12(Suppl. 1), 7–29. doi: 10.1162/089892900561959

Bellugi, U., Marks, S., Bihrle, A., and Sabo, H. (1988). "Dissociation between language and cognitive functions in Williams syndrome," in *Language development in exceptional circumstances*, eds D. Bishop and K. Mogford (London: Churchill Livingstone), 177–189.

Bellugi, U., Wang, P. P., and Jernigan, T. L. (1994). "Williams syndrome: An unusual neuropsychological profile," in *Atypical cognitive deficits in developmental disorders: Implications for brain function*, eds S. Broman and K. Grafman (Hillsdale, NJ: Lawrence Erlbaum Associates, Inc), 23–56.

Benítez-Burraco, A., Garayzábal, E., and Cuetos, F. (2017). Morphology in Spanish-speaking children with Williams syndrome. *Lang. Cogn.* 9, 728–740. doi: 10.1017/langcog.2017.6

Bernhardt, B. M., and Stemberger, J. P. (2017). "Investigating typical and protracted phonological development across languages," in *Crosslinguistic encounters in language acquisition*, eds E. Babatsouli, D. Ingram, and N. Müller (Bristol: Multilingual Matters), 71–108. doi: 10.21832/9781783099092-008

Böhning, M., Campbell, R., and Karmiloff-Smith, A. (2002). Audiovisual speech perception in Williams syndrome. *Neuropsychologia* 40, 1396–1406. doi: 10.1016/S0028-3932(01)00208-1

Boloh, Y., and Ibernon, L. (2010). Gender attribution and gender agreement in 4- to 10-year-old French children. *Cogn. Dev.* 25, 1–25. doi: 10.1016/j.cogdev.2009.09.011

Bosch, L. (2004). *Evaluación fonológica del habla infantil.* MASSON.

Brock, J. (2007). Language abilities in Williams syndrome: A critical review. *Dev. Psychopathol.* 19, 97–127. doi: 10.1017/S095457940707006X

Camp, J. S., Karmiloff-Smith, A., Thomas, M. S. C., and Farran, E. K. (2016). Cross-syndrome comparison of real-world executive functioning and problem solving using a new problem-solving questionnaire. *Res. Dev. Disabil.* 59, 80–92. doi: 10.1016/j.ridd.2016.07.006

Capirci, O., Sabbadini, L., and Volterra, V. (1996). Language development in Williams syndrome: A case study. *Cogn. Neuropsychol.* 13, 1017–1039.

Clahsen, H., Ring, M., and Temple, C. (2004). "). Lexical and morphological skills in English-speaking children with Williams syndrome," in *Language acquisition and language disorders*, Vol. 36, eds S. Bartke and J. Siegmüller (Amsterdam: John Benjamins Publishing Company), 221–244. doi: 10.1075/lald.36.14cla

Cohen, J. (1988). *Statistical Power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Crawford, N. A., Edelson, L. R., Skwerer, D. P., and Tager-Flusberg, H. (2008). Expressive language style among adolescents and adults with Williams syndrome. *Appl. Psycholinguist.* 29, 585–602. doi: 10.1017/S0142716408080259

Davis, B. L., and Bedore, L. M. (2013). *An emergence approach to speech acquisition*. Hove: Psychology Press. doi: 10.4324/9780203375303

Diez-Itza, E. (1992). *Adquisición del lenguaje*. Oviedo: Pentalfa.

Diez-Itza, E., Antón, A., Fernández-Toral, J., and García, M. L. (1998). "Language development in Spanish children with Williams syndrome," in *Perspectives on language acquisition*, eds A. Aksu-Koç, E. Erguvanli-Taylan, A. Sumru, and A. Küntay (Beşiktaş: Bogazici University), 309–324.

Diez-Itza, E., and Martínez, V. (2004). Las etapas tardías de la adquisición fonológica: Procesos de reducción de grupos consonánticos. *Anuario Psicol.* 35, 177–202.

Diez-Itza, E., Martínez, V., Cantora, R., Justicia, F., and Bosch, L. (2001). "Late phonological processes in the acquisition of Spanish," in *Research on child language acquisition*, eds M. Almgren, A. Barreña, M. J. Ezeizabarrena, I. Idiazábal, and B. MacWhinney (Somerville, MA: Cascadilla Press), 790–799.

Diez-Itza, E., Martínez, V., Fernández-Urquiza, M., and Antón, A. (2017). "Morphological profile of Williams syndrome: Typical or atypical?," in *Language development and disorders in spanish-speaking children*, Vol. 14, eds A. Auza Benavides and R. G. Schwartz (Cham: Springer International Publishing), 311–327. doi: 10.1007/978-3-319-53646-0_15

Diez-Itza, E., Martínez, V., Miranda, M., Antón, A., Ojea, A. I., Fernández-Urquiza, M., et al. (2014). "The Syndroling Project: A comparative linguistic analysis of typical development profiles and neurodevelopmental genetic syndromes (Down, Williams and fragile X syndromes)," in *Proceedings of the IASCL-XII international congress for the study of child language*, Amsterdam.

Diez-Itza, E., Martínez, V., Pérez, V., and Fernández-Urquiza, M. (2018). Explicit oral narrative intervention for students with Williams syndrome. *Front. Psychol.* 8:2337. doi: 10.3389/fpsyg.2017.02337

Diez-Itza, E., Miranda, M., Pérez, V., and Martínez, V. (2019). "Profiles of grammatical morphology in Spanish-speaking adolescents with Williams syndrome and Down syndrome," in *Atypical language development in romance languages*, eds E. Aguilar-Mediavilla, L. Buil-Legaz, R. López-Penadés, V. A. Sanchez-Azanza, and D. Adrover-Roig (Amsterdam: John Benjamins Publishing Company), 219–234. doi: 10.1075/z.223.13die

Diez-Itza, E., Snow, C., and MacWhinney, B. (1999). La metodología RETAMHE y el proyecto CHILDES: Breviario para la codificación y análisis de lenguaje infantil. *Psicotema* 11, 517–530.

Diez-Itza, E., Vergara, P., Barros, M., Miranda, M., and Martínez, V. (2021). Assessing phonological profiles in children and adolescents with Down syndrome: The effect of elicitation methods. *Front. Psychol.* 12:662257. doi: 10.3389/fpsyg.2021.662257

Diez-Itza, E., Viejo, A., and Fernández-Urquiza, M. (2022). Pragmatic profiles of adults with Fragile X syndrome and Williams syndrome. *Brain Sci.* 12:385. doi: 10.3390/brainsci12030385

Donegan, P. (2013). "Normal vowel development," in *Handbook of vowels and vowel disorders*, eds M. J. Ball, and F. E. Gibbon (London: Psychology Press), 24–60.

D'Souza, D., Cole, V., Farran, E. K., Brown, J. H., Humphreys, K., Howard, J., et al. (2015). Face processing in Williams syndrome is already atypical in infancy. *Front. Psychol.* 6:760. doi: 10.3389/fpsyg.2015.00760

D'Souza, D., D'Souza, H., Jones, E. J. H., and Karmiloff-Smith, A. (2020). Attentional abilities constrain language development: A cross-syndrome infant/toddler study. *Dev. Sci.* 23:e12961. doi: 10.1111/desc.12961

Dunn, L. M., Dunn, L. M., and Arribas, D. (2010). *PPVT-III PEABODY. Test de vocabulario en imágenes*. Madrid: TEA ediciones.

Garayzábal, E., Osório, A., Lens, M., and Sampaio, A. (2014). Concrete and relational vocabulary: Comparison between Williams and Smith–Magenis syndromes. *Res. Dev. Disabil.* 35, 3365–3371. doi: 10.1016/j.ridd.2014.07.055

Grant, J., Karmiloff-Smith, A., Gathercole, S. A., Paterson, S., Howlin, P., Davies, M., et al. (1997). Phonological short-term memory and its relationship to language in Williams syndrome. *Cogn. Neuropsychiatry* 2, 81–99. doi: 10.1080/135468097396342

Grunwell, P. (1981). The development of phonology: A descriptive profile. *First Language* 3, 161–191.

Hargrove, P. M., Pittelko, S., Fillingane, E., Rustman, E., and Lund, B. (2012). Perceptual speech and paralinguistic skills of adolescents with Williams syndrome. *Commun. Disord. Q.* 34, 152–161. doi: 10.1177/1525740112436372

Heiz, J., and Barisnikov, K. (2016). Visual-motor integration, visual perception and motor coordination in a population with Williams syndrome and in typically developing children: Visual-motor abilities in a population with Williams syndrome. *J. Intellect. Disabil. Res.* 60, 945–955. doi: 10.1111/jir.12328

Hidalgo, I. (2019). *El nivel fónico de la población con síndrome de Smith Magenis: Particularidades fonatorias y fonético-fonológicas. Comparativa con síndrome de Williams, síndrome de Down y desarrollo típico*. Ph.D. thesis. Madrid: Universidad Autónoma de Madrid.

Hidalgo, I., and Garayzábal, E. (2019). Diferencias fonológicas entre síndromes del neurodesarrollo: Evidencias a partir de los procesos de simplificación fonológica más frecuentes. *Rev. Invest. Logopedia* 9, 81–106. doi: 10.5209/rlog.62942

Hochmann, J.-R., Benavides-Varela, S., Nespor, M., and Mehler, J. (2011). Consonantsand vowels: Different roles in early language acquisition. *Dev. Sci.* 14, 1445–1458. doi: 10.1111/j.1467-7687.2011.01089.x

Huelmo, J., Martínez, V., and Diez-Itza, E. (2017). Evaluación de perfiles fonológicos en el síndrome X-Frágil mediante índices de error. *Rev. INFAD Psicol.* 4, 67–76. doi: 10.17060/ijodaep.2017.n1.v4.1028

Huffman, M. J. (2019). *Speech articulation in children with Williams syndrome or 7q11.23 duplication syndrome*. Ph.D. thesis. Louisville, KY: University of Louisville.

Ingram, D. (1976). *Phonological disability in children*. London: Edward Arnold.

Jarrold, C., and Baddeley, A. D. (2001). Short-term memory in Down syndrome: Applying the working memory model. *Downs Syndr. Res. Pract.* 7, 17–23. doi: 10.3104/reviews.110

Jarrold, C., Baddeley, A. D., Hewes, A. K., and Phillips, C. (2001). A Longitudinal assessment of diverging verbal and non-verbal abilities in the Williams syndrome phenotype. *Cortex* 37, 423–431. doi: 10.1016/S0010-9452(08)70583-5

Järvinen, A., Korenberg, J. R., and Bellugi, U. (2013). The social phenotype of Williams syndrome. *Curr. Opin. Neurobiol.* 23, 414–422. doi: 10.1016/j.conb.2012.12.006

Karmiloff-Smith, A. (1998). Development itself is the key to understanding developmental disorders. *Trends Cogn. Sci.* 2, 389–398. doi: 10.1016/S1364-6613(98)01230-3

Karmiloff-Smith, A., Grant, J., Berthoud, I., Davies, M., Howlin, P., and Udwin, O. (1997). Language and Williams syndrome: How intact is "intact"? *Child Dev.* 68, 246–262.

Kirchner, R. M., Martens, M. A., and Andridge, R. R. (2016). Adaptive behavior and development of infants and toddlers with Williams syndrome. *Front. Psychol.* 7:598. doi: 10.3389/fpsyg.2016.00598

Kozel, B. A., Barak, B., Kim, C. A., Mervis, C. B., Osborne, L. R., Porter, M., et al. (2021). Williams syndrome. *Nat. Rev. Dis. Primers* 7:42. doi: 10.1038/s41572-021-00276-z

Krishnan, S., Bergström, L., Alcock, K. J., Dick, F., and Karmiloff-Smith, A. (2015). Williams syndrome: A surprising deficit in oromotor praxis in a population with proficient language production. *Neuropsychologia* 67, 82–90. doi: 10.1016/j.neuropsychologia.2014.11.032

Kumin, L. (2006). Speech intelligibility and childhood verbal apraxia in children with Down syndrome. *Down Syndr. Res. Pract.* 10, 10–22. doi: 10.3104/reports.301

Laing, E., Hulme, C., Grant, J., and Karmiloff-Smith, A. (2001). Learning to read in Williams syndrome: Looking beneath the surface of atypical reading development. *J. Child Psychol. Psychiatry* 42, 729–739. doi: 10.1111/1469-7610.00769

Levy, Y., and Eilam, A. (2013). Pathways to language: A naturalistic study of children with Williams syndrome and children with Down syndrome. *J. Child Lang.* 40, 106–138. doi: 10.1017/S0305000912000475

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk.Transcription format and programs*, Vol. I. Hillsdale, NJ: Lawrence Erlbaum.

Majerus, S. (2004). "Phonological processing in Williams syndrome," in *Language Acquisition and Language Disorders*, Vol. 36, eds S. Bartke and J. Siegmüller (Amsterdam: John Benjamins Publishing Company), 125–142. doi: 10.1075/lald.36.10maj

Majerus, S., Barisnikov, K., Vuillemin, I., Poncelet, M., and Linden, M. (2003). An investigation of verbal short-term memory and phonological processing in four children with Williams syndrome. *Neurocase* 9, 390–401. doi: 10.1076/neur.9.5.390.16558

Martínez, V. (2010). *Etapas tardías del desarrollo fonológico infantil: Procesos y límites del trastorno.* Ph.D. thesis. Oviedo: Universidad de Oviedo.

Martínez, V., Antón, A., Miranda, M., Pérez, V., Fernández-Toral, J., and Diez-Itza, E. (2014). "Accelerated phonological development in Williams syndrome: A two case corpus-based study of late phonological processes," in *Proceedings of the IASCL-XII international congress for the study of child language*, Amsterdam.

Martínez, V., and Diez-Itza, E. (2012). Procesos de asimilación en las etapas tardías del desarrollo. *Psicothema* 24, 193–198.

Martínez-Castilla, P., Stojanovik, V., Setter, J., and Sotillo, M. (2012). Prosodic abilities in Spanish and English children with Williams syndrome: A cross-linguistic study. *Appl. Psycholinguist.* 33, 1–22. doi: 10.1017/S0142716411000385

Masataka, N. (2001). Why early linguistic milestones are delayed in children with Williams syndrome: Late onset of hand banging as a possible rate–limiting constraint on the emergence of canonical babbling. *Dev. Sci.* 4, 158–164. doi: 10.1111/1467-7687.00161

Mayall, L. A., D'Souza, H., Hill, E. L., Karmiloff-Smith, A., Tolmie, A., and Farran, E. K. (2021). Motor abilities and the motor profile in individuals with Williams syndrome. *Adv. Neurodev. Disord.* 5, 46–60. doi: 10.1007/s41252-020-00173-8

McLeod, S., and Crowe, K. (2018). Children's consonant acquisition in 27 languages: A cross-linguistic review. *Am. J. Speech Lang. Pathol.* 27, 1546–1571. doi: 10.1044/2018_AJSLP-17-0100

McLeod, S., Van Doorn, J., and Reed, V. (2001). Consonant cluster development in two-year-olds: General trends and individual difference. *J. Speech Lang. Hear. Res.* 44, 1144–1171. doi: 10.1044/1092-4388(2001/090)

Mervis, C. B., and Becerra, A. M. (2007). Language and communicative development in Williams syndrome. *Ment. Retard. Dev. Disabil. Res. Rev.* 13, 3–15. doi: 10.1002/mrdd.20140

Mervis, C. B., and John, A. E. (2008). Vocabulary abilities of children with Williams syndrome: Strengths, weaknesses, and relation to visuospatial construction ability. *J. Speech Lang. Hear. Res.* 51, 967–982. doi: 10.1044/1092-4388(2008/071)

Mervis, C. B., Klein-Tasman, B. P., Huffman, M. J., Velleman, S. L., Pitts, C. H., Henderson, D. R., et al. (2015). Children with 7q11.23 duplication syndrome: Psychological characteristics. *Am. J. Med. Genet. A* 167, 1436–1450. doi: 10.1002/ajmg.a.37071

Mervis, C. B., Robinson, B. F., Rowe, M. L., Becerra, A. M., and Klein-Tasman, B. P. (2004). "Relations between language and cognition in Williams syndrome," in *Language acquisition and language disorders*, Vol. 36, eds S. Bartke and J. Siegmüller (Amsterdam: John Benjamins Publishing Company), 63–92. doi: 10.1075/lald.36.08mer

Mervis, C. B., and Velleman, S. L. (2011). Children with Williams syndrome: Language, cognitive, and behavioral characteristics and their implications for intervention. *Perspect. Lang. Learn. Educ.* 18, 98–107. doi: 10.1044/lle18.3.98

Miezah, D., Porter, M., Batchelor, J., Boulton, K., and Campos, G. (2020). Cognitive abilities in Williams syndrome. *Res. Dev. Disabil.* 104:103701. doi: 10.1016/j.ridd.2020.103701

Moraleda, E., and López, P. (2020). Analysis of receptive vocabulary development in Williams syndrome. *Open J. Mod. Linguist.* 10, 804–812. doi: 10.4236/ojml.2020.106050

Nip, I. S. B., Green, J. R., and Marx, D. B. (2009). Early speech motor development: Cognitive and linguistic considerations. *J. Commun. Disord.* 42, 286–298. doi: 10.1016/j.jcomdis.2009.03.008

Pérez, D., Vivar, P., Bernhardt, B. M., Mendoza, E., Ávila, C., Carballo, G., et al. (2018). Word-initial Rhotic clusters in Spanish-speaking preschoolers in Chile and Granada, Spain. *Clin. Linguist. Phon.* 32, 481–505. doi: 10.1080/02699206.2017.1359852

Pérez Jurado, L. A. (2003). Williams-Beuren syndrome: A model of recurrent genomic mutation. *Horm. Res. Paediatr.* 59, 106–113. doi: 10.1159/000067836

Pérez-García, D., Brun-Gasca, C., Pérez-Jurado, L. A., and Mervis, C. B. (2017). Behavioral profiles of children with Williams syndrome from Spain and the United States: Cross-Cultural similarities and differences. *Am. J. Intellect. Dev. Disabil.* 122, 156–172. doi: 10.1352/1944-7558-122.2.156

Purser, H. R. M., Thomas, M. S. C., Snoxall, S., Mareschal, D., and Karmiloff-Smith, A. (2010). Definitions versus categorization: Assessing the development of lexico-semantic knowledge in Williams syndrome. *Int. J. Lang. Commun. Disord.* 46, 361–373. doi: 10.3109/13682822.2010.497531

Ranzato, E., Tolmie, A., and Van Herwegen, J. (2021). The home learning environment of primary school children with Down Syndrome and those with Williams Syndrome. *Brain Sci.* 11:733. doi: 10.3390/brainsci11060733

Reilly, J., Losh, M., Bellugi, U., and Wulfeck, B. (2004). Frog, where are you?" Narratives in children with specific language impairment, early focal brain injury, and Williams syndrome. *Brain Lang.* 88, 229–247. doi: 10.1016/S0093-934X(03)00101-9

Rose, Y., and Inkelas, S. (2011). "The interpretation of phonological patterns in first language acquisition: The interpretation of phonological patterns in first language acquisition," in *The Blackwell companion to phonology*, eds M. van Oostendorp, C. J. Ewen, E. Hume, and K. Rice (Malden, MA: Wiley-Blackwell), 1–25. doi: 10.1002/9781444335262.wbctp0101

Semel, E., and Rosner, S. R. (2003). *Understanding Williams syndrome*. Mahwah, NJ: Lawrence Erlbaum Associates. doi: 10.4324/9781410607416

Setter, J., Stojanovik, V., Van Ewijk, L., and Moreland, M. (2007). Affective prosody in children with Williams syndrome. *Clin. Linguist. Phon.* 21, 659–672. doi: 10.1080/02699200701539056

Smith, N. V. (1973). *The acquisition of phonology. A case study*. Cambridge: Cambridge University Press.

Stoel-Gammon, C. (2011). Relationships between lexical and phonological development in young children. *J. Child Lang.* 38, 1–34. doi: 10.1017/S0305000910000425

Stojanovik, V. (2006). Social interaction deficits and conversational inadequacy in Williams syndrome. *J. Neurolinguist.* 19, 157–173. doi: 10.1016/j.jneuroling.2005.11.005

Stojanovik, V. (2010). Understanding and production of prosody in children with Williams syndrome: A developmental trajectory approach. *J. Neurolinguist.* 23, 112–126. doi: 10.1016/J.JNEUROLING.2009.11.001

Stojanovik, V., Perkins, M., and Howard, S. (2001). Language and conversational abilities in Williams syndrome: How good is good? *Int. J. Lang. Commun. Disord.* 36, 234–239. doi: 10.3109/13682820109177890

Thomas, M. S. C., Grant, J., Barham, Z., Gsödl, M., Laing, E., Lakusta, L., et al. (2001). Past tense formation in Williams syndrome. *Lang. Cogn. Process.* 16, 143–176. doi: 10.1080/01690960042000021

Thomas, M. S. C., and Karmiloff-Smith, A. (2003). Modeling language acquisition in atypical phenotypes. *Psychol. Rev.* 110, 647–682. doi: 10.1037/0033-295X.110.4.647

Udwin, O., and Yule, W. (1990). Augmentative communication systems taught to cerebral palsied children-a longitudinal study. I. The acquisition of signs and symbols, and syntactic aspects of their use over time. *Int. J. Lang. Commun. Disord.* 25, 295–309. doi: 10.3109/13682829009011979

Velleman, S. L., Currier, A., Caron, T., Curley, A., and Mervis, C. B. (2006). *Phonological development in Williams syndrome*. (Dubrovnik: International Clinical Phonetics and Linguistics Association), 108–116.

Vergara, P., Bernhardt, B. M., Pérez, D., and Diez-Itza, E. (2021). Consonant cluster acquisition in Chilean children with typical and protracted phonological development. *Clin. Linguist. Phon.* 35, 964–982. doi: 10.1080/02699206.2020.1851306

Vicari, S., Brizzolara, D., Carlesimo, G. A., Pezzini, G., and Volterra, V. (1996a). Memory abilities in children with Williams syndrome. *Cortex* 32, 503–514. doi: 10.1016/S0010-9452(96)80007-4

Vicari, S., Carlesino, G., Brizzolara, D., and Pezzini, G. (1996b). Short-term memory in children with Williams syndrome: A reduced contribution of lexical-semantic knowledge to word span. *Neuropsychologia* 34, 919–925. doi: 10.1016/0028-3932(96)00007-3

Volterra, V., Capirci, O., Pezzini, G., Sabbadini, L., and Vicari, S. (1996). Linguistic abilities in Italian children with Williams syndrome. *Cortex* 32, 663–677. doi: 10.1016/S0010-9452(96)80037-2

TYPE Original Research
PUBLISHED 24 November 2022
DOI 10.3389/fpsyg.2022.986937

OPEN ACCESS

EDITED BY
Theodoros Marinis,
University of Konstanz, Germany

REVIEWED BY
Theodora Papastefanou,
Cyprus University of
Technology, Cyprus
Laura Gonnerman,
McGill University, Canada

*CORRESPONDENCE
Deanna C. Friesen
Deanna.Friesen@uwo.ca

SPECIALTY SECTION
This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Psychology

RECEIVED 05 July 2022
ACCEPTED 01 November 2022
PUBLISHED 24 November 2022

CITATION
Friesen DC, Schmidt K, Atwal T and Celebre A (2022) Reading comprehension and strategy use: Comparing bilingual children to their monolingual peers and to bilingual adults. *Front. Psychol.* 13:986937.
doi: 10.3389/fpsyg.2022.986937

Reading comprehension and strategy use: Comparing bilingual children to their monolingual peers and to bilingual adults

Deanna C. Friesen*, Katherine Schmidt, Taninder Atwal and Angela Celebre

Faculty of Education, University of Western Ontario, London, ON, Canada

The current study investigated the predictive ability of language knowledge and reported strategy use on reading comprehension performance in English-speaking monolingual and bilingual students. One hundred fifty-five children in grade 4 through 6 (93 bilinguals and 62 monolinguals) were assessed on receptive vocabulary, word reading fluency, reading comprehension, and reading strategy use in English. An additional 38 adult bilinguals (i.e., English Language Learners) were assessed on the same measures. Although, the bilingual adult group and bilingual children had significantly lower English vocabulary knowledge relative to the monolingual children, the bilingual adults exhibited reading comprehension performance that was on par with the monolingual children; both groups outperformed the bilingual children. This discrepancy was accounted for by reported strategy use, wherein bilingual adults reported more inferencing, more connecting between sections of text and more reference to the text structure than the children. Reported strategy use also accounted for unique variance in reading comprehension performance above and beyond the contributions of English vocabulary knowledge and word reading fluency. Findings highlight the strategies that successful readers report and emphasize the value of promoting effective strategy selection in addition to language instruction in the development of reading comprehension skill.

KEYWORDS

bilingualism, reading comprehension, reading strategies, children, English language learner

Introduction

By definition, bilinguals divide their language exposure between two languages. Consequently, they have fewer opportunities to develop proficiencies in each language, and often exhibit weaker second language reading comprehension performance than native speakers of that language (e.g., Aarts and Verhoeven, 1999; Geva and Farnia, 2012; Raudszus et al., 2021). However, reading comprehension success also depends on deploying strategies to extract meaning from print (McNamara, 2012). For bilinguals, it may be especially important to use strategies to offset weaker second language (L2) knowledge (Kolić-Vehovec and Bajšanski, 2007). In the current study, we investigated how three groups of readers reported their use of reading comprehension strategies in English and whether this reported strategy use predicted reading comprehension success.

Unfortunately, the reading comprehension achievement gap between monolingual readers and L2 readers can widen throughout elementary school (e.g., Droop and Verhoeven, 2003; Farnia and Geva, 2013; Raudszus et al., 2021). In Canada, Farnia and Geva (2013) reported that unlike their monolingual peers, the growth trajectory for L2 learners' reading comprehension performance leveled off from Grade 4 to 6. Droop and Verhoeven (2003) also reported stronger reading comprehension in monolinguals relative to their bilingual peers from grade 3 to 4 in the Netherlands. These differences may be problematic given the importance of reading comprehension for both school and career success (August and Shanahan, 2006). Key then, is to understand the locus of these reading comprehension differences and provide instruction to address students' literacy needs.

The predominant approach to understanding reading comprehension is to examine the relative contributions of component skills. Arguably, the Simple View of Reading model (SVR; Hoover and Gough, 1990) is the most widely cited framework of reading comprehension development. In the SVR model, reading comprehension is the product of decoding ability and linguistic comprehension (D x LC = RC). Decoding refers to word recognition processes (e.g., using grapheme-phoneme correspondences), whereas linguistic comprehension refers to the skills necessary to understand language (e.g., vocabulary, syntax, grammar, discourse processes). Several studies have confirmed the importance of both language knowledge and decoding ability for successful reading comprehension (See Castles et al., 2018 for a review). Indeed, reading comprehension success is unlikely if one of these components is missing or weak (Joshi and Aaron, 2000).

Differences in linguistic comprehension have been isolated as the main source of reading comprehension language group differences. In Droop and Verhoeven (2003), bilinguals exhibited faster word decoding than their monolingual peers but poorer language proficiency and reading comprehension performance. Likewise, Geva and Farnia (2012) reported no word reading differences between groups despite weaker reading comprehension performance and syntax knowledge in English second language learners. Raudszus et al. (2021) found that L2 readers with high vocabulary knowledge showed comparable reading comprehension growth relative to high vocabulary L1 readers. In contrast, L2 readers with low vocabulary exhibited less reading comprehension growth than monolingual readers who also had weak vocabulary knowledge. Importantly, Bialystok et al. (2010) have reported a consistent 9-point difference on a standardized English vocabulary measure between monolinguals ($N = 966$) and bilinguals ($N = 772$) who were between the ages of 3 and 10 years old in Canada. Such findings support Droop and Verhoeven's proposal that group differences in reading comprehension are affected by language knowledge in a top-down fashion, as opposed to bottom-up word decoding.

Importantly, as age increases, language knowledge becomes a better predictor of reading comprehension success than decoding ability (Gough et al., 1996; Storch and Whitechurst, 2002; Proctor et al., 2006; Gunnerud et al., 2022). Proctor et al. (2005) found that with sufficient L2 decoding ability, L2 vocabulary is the critical variable for L2 reading comprehension outcomes. For adults, decoding ability is often not a predictor of reading comprehension performance (e.g., Landi, 2010; Friesen and Frid, 2021). Babayigit and Shapiro (2020) reported that both English vocabulary and grammar knowledge are strong predictors of English reading comprehension performance and should be strongly targeted for L2 instruction. A meta-analysis demonstrated that several studies report that language comprehension skills are stronger predictors of reading comprehension in the L2 than in the L1 (Melby-Lervåg and Lervåg, 2014). Taken together, these findings indicate that these language skills need to be supported and/or offset in bilingual students' education.

Although language variables have received the most attention, more recent work has expanded the scope of reading comprehension predictors to include cognitive (e.g., working memory; Farnia and Geva, 2013), affective (e.g., motivation, e.g., Cho et al., 2019) and meta-cognitive measures (e.g., van Steensel et al., 2016). In their Reading Systems Framework, Perfetti and Stafura (2014) identify knowledge (e.g., vocabulary), processes (e.g., decoding) and cognitive abilities (e.g., executive functions) as reader variables that underlie reading comprehension success. However, reading is dynamic, and individuals modify their reading behaviors as a function of the nature of the text and their reading goals (Rand Reading Study Group, 2002). Presumably then, effective strategy selection in response to the reading demands is also a critical ability for consistent reading comprehension success.

The current study examined whether the type of strategies used by L2 readers and monolinguals predict reading comprehension success beyond what is accounted for by traditional language measures. We focused on strategy use for several reasons. First, as noted by Afflerbach et al. (2008), reading strategies are "deliberate, goal-directed attempts to control and modify the reader's efforts to decode text, understand words and construct meanings of texts" (p. 368). Thus, they are subject to explicit instruction from teachers and strategies can be targeted for improvement. Second, reading comprehension strategy instruction is integral to most language arts curriculums. However, strategies are often listed without identifying which strategies may work together or be most effective to retain content (e.g., Ontario Ministry of Education, 2006). Finally, a meta-analysis found that L2 strategy instruction produces small to moderate effect sizes and that several variables impact the strength of the outcomes, including strategy selection itself (Plonsky, 2011). Here, our focus is on identifying reading comprehension strategies that are correlated with reading comprehension success to help inform effective strategy selection.

Reading strategies have been categorized in several ways (e.g., Mokhtari and Reichard, 2004; Plonsky, 2011). Here we focus on strategies that can be used during reading. For example, Block (1986) divided strategies into general and local strategies. General strategies included prediction, identifying text structure, questioning, and using background knowledge. Local strategies pinpoint particular parts of texts and include paraphrasing, rereading, questioning the meaning of a clause or a sentence, and questioning the meaning of vocabulary. Janzen and Stoller (1998) also identified a set of strategies which included predicting, asking questions, checking predictions, or looking for answers to questions, connecting the text to the prior knowledge, summarizing, connecting within the text, and recognizing text structure. The current study adopted the approach of looking at several of these individual strategies (see Supplementary material) because these strategies are often emphasized in both curriculum documents (Ontario Ministry of Education, 2006) and consequently, in language classrooms.

In the monolingual literature, higher-order processes have been found to predict reading comprehension success. For example, Oakhill et al. (2003) followed monolingual children from ages 7–10. Inferential ability, comprehension monitoring ability and knowledge of text structure at age 7 and 8 predicted reading comprehension success at age 10. Good comprehenders also analyze arguments found in text, utilize background knowledge (Saricoban, 2002), and pose questions (Yopp, 1988). Engaging in visualization has also been shown to result in reading comprehension gains (Pressley, 2000; Erfani et al., 2011). Importantly, poor language proficiency may be offset by engaging in some of this effective strategy use (Carrell, 1989; Padrón, 1992; Kolić-Vehovec and Bajšanski, 2007; Friesen and Haigh, 2018).

Effective reading strategy use in L2 has been studied using both questionnaires (e.g., Mokhtari and Reichard, 2004; Afsharrad and Benis, 2017; see Friesen and Frid, 2021 for a brief review) and think-aloud protocols (e.g., Jiménez et al., 1996; Chamot and El-Dinary, 1999; Park and Kim, 2015). Work with adults has typically favored using questionnaires. Although questionnaire data is easier to collect, data is based on respondents' reflections of their strategy use and may not be accurate (Brown, 2017). These retrospective meta-cognitive processes may also be beyond the capabilities of young readers to evaluate. We favor the think-aloud procedure where readers report their thought processes during reading because rich descriptive data is captured online (Chamot and El-Dinary, 1999). Although there are concerns that readers are only reporting a subset of their strategy use and that comprehension may be altered, think-alouds do enable insight into the types of strategies a reader is able to access during online processing.

In a literature review, Brantmeier (2002) summarized that successful L2 adult readers prefer top-down strategies such as integrating distinct parts of text, referring to text structure, and making links to background knowledge, whereas less successful readers used more bottom-up strategies such as rereading and identifying lexical problems. In a think-aloud study, Lin and Yu (2015) found that more proficient L2 adults engaged in more effective and varied strategies that were aimed toward comprehension in their L2, whereas less proficient L2 users were focused on language-oriented strategies. More proficient L2 readers asked more questions, paraphrased more, translated more, and used more contextual cues than the less proficient bilingual readers.

A few think-aloud studies have been conducted with children as L2 readers. One main concern with this work is that sample sizes are often small. For example, Park and Kim (2015) examined strategy use with four L2 learners of English in Grade 4 or 5. Students used a dialogic approach where students spoke aloud to others to engage in meaning-making. They posed questions, made inferences, relied on previous knowledge, and drew conclusions. In another example, Jiménez et al. (1996) compared eight good L2 readers with three poor L2 readers on their think-aloud strategies. Good L2 readers translated text, resolved unknown vocabulary, monitored comprehension, connected text to previous knowledge and made inferences/conclusions. Poor L2 readers identified unknown words but did not attempt to determine their meaning. In García and Godina's (2017) work, Grade 5 bilinguals with more L2 proficiency used more varied strategies, generated more plausible inferences, referred to background knowledge, paraphrased, and monitored comprehension more than the students with less L2 proficiency. Although these studies are informative, it is not clear whether the strategies explained unique variance in reading comprehension performance in bilingual readers that was not accounted for by their language abilities.

Work by Frid and Friesen (2020) examined which reading strategies were related to reading comprehension success in French Immersion students in Grades 4 and 5. These students present a unique population since English is L1, but L1 reading instruction begins in Grade 4. Participants predicted and generated inferences more in L1. They summarized and referred to unknown vocabulary more in L2 than L1. Additionally, reliance on inferencing behaviors and text analysis accounted for unique variance in reading comprehension performance in each language. A similar emphasis on these strategies was reported by Friesen and Frid (2021) in a study that addressed the same questions in English-French bilingual adults.

The current study

The current study asked (1) whether strategy use differed as a function of language experience by comparing performance across three groups of English speakers and (2) whether strategy use accounted for unique variance in the reading comprehension performance beyond language knowledge (i.e., receptive vocabulary and word reading fluency). To our knowledge, this is the first study to compare L2 children with both age-matched monolingual children and language-matched L2 adults. The monolingual children and bilingual children differed in their English vocabulary knowledge, enabling us to draw conclusions about the role of language in reading comprehension performance. In contrast, the bilingual adults had the same degree of English vocabulary proficiency as the bilingual children but differed in their reading experience from both groups of children, enabling us to examine the role of greater literacy expertise in reading comprehension performance. By making these comparisons, we can gain insight into the importance of both language proficiency and reading experience on reading comprehension performance and strategy use.

Both language measures and strategy use were assessed and used as predictors of reading comprehension performance. Think-aloud reponses were coded for ten strategies (i.e., reference to vocabulary, reference to text structure, reference to background knowledge, connecting to texts or previous think-alouds, summarizing, necessary inferences, elaborative inferences, questioning, predicting, and visualizing). Necessary inferences were drawn conclusions required to maintain text cohesion. Elaborative inferences were reasonable conclusions based on the text but unnecessary for understanding (see Supplementary material for full descriptions and examples). We asked whether monolingual English readers and English second language readers reported similar reading strategies to construct meaning from text. We also asked whether individual reading strategies and language abilities uniquely account for reading comprehension success in all three groups. Ideally, knowledge about how second language learners process text will serve to support reading comprehension development in struggling readers.

Method

Participants

Participants were 155 students in Grades 4 to 6 from a large school board in Ontario, Canada. Of these, 93 students were bilingual with English as one of their languages (M_{age} = 10.6, SD = 1.0; 52 females) and 62 students were English monolingual speakers (M_{age} = 10.7, SD = 1.0; 37 females). To be classified as bilingual, students had to speak a language other than English in the home; English was learned either both in the home and school, or just in school. On average, bilingual children had 4.5 years (SD = 1.9) of English schooling compared to 5.7 years (SD = 1.0) for monolingual children. Home languages of the bilinguals were Albanian (2), Amharic (1), Arabic (42), Bengali (1), Bosnian (2), Chinese (12), Dari (1), Hindi (1), Khmer (1), Korean (7), Kurdish (4), Pashto (2), Portuguese (1), Punjabi (2), Russian (1), Sindhi (1), Tagalog (1), Tamil (1), Turkish (1), and Urdu (5). Parents reported that bilingual children spent an average of 5.0 (SD = 7.2) hours reading in English outside of school per week, whereas monolingual children spent an average of 7.2 (SD = 13.8) hours per week. Bilingual children read for an average of 2.1 h (SD = 4.2) in their other language. Parents reported that their child preferred to read in English [bilinguals: 4.4 (SD = 0.98), monolinguals: 4.5 (SD = 1.4)] on a five-point scale where 5 was strongly agree (see Table 1 for additional demographic information).

TABLE 1 Means and standard deviations (in parentheses) for background and language measures for each group.

Measures	Monolingual children	Bilingual children	Bilingual adults
Questionnaire measures			
English AoA (in years)	–	3.5 (3.0)[a]	8.8 (2.7)
Other language proficiency rating[b]	–	7.6 (2.6)	8.9 (2.6)
Language use (speaking)[c]	1.0 (0.1)	4.2 (2.0)	3.5 (1.5)
Language use (reading)[c]	1.0 (0.1)	2.0 (1.3)	1.8 (1.2)
English language measures			
Receptive vocabulary (max. 204)	141.0 (22.3)[d]	114.6 (37.5)	119.1 (21.0)
Word reading fluency (max. 167)	103.0 (25.4)	96.8 (30.8)	103.3 (20.2)
Reading comprehension (max. 24)	12.3 (4.4)	9.5 (5.6)	12.1 (2.7)

[a]Participants tested in the first cohort are missing this data (∼ a third of participants).
[b]Rating scale from 1 to 10 (where 1 is poor & 10 is native-like).
[c]Rating scale from 1 to 7 (where 1 was all English and 7 was all other language).
[d]As a point of comparison, 141 is equivalent to a standard score of 106 based on monolingual norms.

An additional 38 sequential bilingual adults also participated (M_{age} = 25.2, SD = 3.7; 36 females). Adult participants were all born outside of Canada and had been in Canada for an average of 9.7 months (SD = 3.7). First languages included Chinese (35), Farsi (1), Malay (1), and Persian (1). Participants were completing a graduate degree (35 were enrolled in a Teaching English to Speakers of Other Languages program). Participants reported reading in English an average of 18.4 h per week (SD = 12.5) and in their first language an average of 14.6 h (SD = 10.5) per week (see Table 1 for additional demographic information).

Measures

Language experience questionnaires (parent version & adult version)

The parental questionnaire included rating scales for their child's understanding and speaking ability in their non-English language, questions about language use in the home, and about the child's reading preferences. The adult questionnaire asked participants about language dominance, language proficiency and language use. Both questionnaires also asked about the Age of Acquisition for English (AoA) defined as when the participant started learning the language. Note the scales for proficiency and language use were different in each questionnaire and have been transformed to be on the same scale for ease of interpretation in Table 1.

Receptive English vocabulary

Receptive vocabulary was assessed using the Peabody Picture Vocabulary Test (PPVT-III: Form A; Dunn and Dunn, 1997). The test is designed for ages 2.5–90+ years old. In each trial, four images were presented, and an auditory word was heard. Participants selected the picture that matched the word. Items are ordered by word difficulty and standard administration follows basal and ceiling rules. Within the sample, the Spearman-Brown Split-test coefficient was 0.99 for the children and 0.96 for the adults. Raw scores were used rather than standard scores, since absolute vocabulary knowledge is appropriate to make group comparisons and for regression analyses.

Word reading fluency

English word reading fluency was assessed using the Test of Word Reading Efficiency (TOWRE; Torgesen et al., 1999). The TOWRE includes sight word reading efficiency (104 words) and phonemic decoding efficiency (63 pseudo-words). Participants read as many items as possible in 45 s for each subtest. Inter-rater reliability was calculated on a subset of participants (25% of child data and 38% of adult data). For the children, agreement was 0.98 on the words and 0.92 on the non-words. For the adults, agreement was 0.98 and 0.91, respectively. The raw total for both measures were added together and was used in the analyses as a measure of word reading fluency.

Reading comprehension and strategy use task

Reading comprehension and strategy use were assessed using four texts taken from the Gray Oral Reading Test (GORT—4th Edition, Form B, Wiederholt and Bryant, 2001). Standard administration of the GORT-4 involves the individual reading texts aloud and responding to multiple-choice questions. Entry points are determined based on an individual's age. Basal and ceiling rules are applied for both fluency and comprehension. Since our focus was not on reading fluency but strategy selection, standard administration was not employed; participants read texts silently. The reason GORT-4 texts were selected is because they provided 14 developmentally sequenced texts for ages 6–18 years. Texts 6, 7, 8, 9, were selected as being age-appropriate for Grades 4 through 8. Text 5 (∼ grade 3 level) was used as an exemplar. The texts averaged 119.8 words ($SD = 25.4$) and 8.25 sentences ($SD = 1.3$) each. Two texts were narrative (i.e., one about a turtle on an adventure & one about shipwrecked siblings) and two were expository (i.e., one about problems faced by farmers & one about Harriet Tubman). The range of texts were selected to address concerns about floor/ceiling effects within any one group. Importantly, to compare strategy use and reading comprehension performance directly among groups and avoid confounds based on the materials, the same texts were employed for all participants. Given this decision, a valid concern is how students' language proficiency impacts their performance. This concern was addressed by measuring word reading fluency and receptive vocabulary knowledge, and then accounting for them in the analyses of reading comprehension performance.

Each text was divided into four sections of approximately two sentences each. Participants read the first section silently and then hit the spacebar. A beep prompted a think-aloud response. To facilitate think-aloud behaviors, participants were provided with a list of 10 sentence starters (e.g., This is what is happening…, I wonder…, I predict…) that corresponded with the critical strategies. Think-aloud exemplars for a sample story were presented to familiarize the participants with the procedure. Once they had completed their first think-aloud, participants hit the spacebar to continue to the next section and earlier sections remained on the screen. When participants completed their final think-aloud, they pressed the spacebar to continue to the comprehension questions. Since the study's goal was to examine the association between text comprehension and strategy use (and not readers' ability to search the text for the correct answers), participants did not have access to the text for the comprehension questions. Participants were invited to complete the think-alouds in their preferred language.

Although each text from the GORT-4 had five multiple choice questions, researcher-generated open-ended comprehension questions were employed. Open-ended questions require readers to rely on their mental representation of the text to generate the answers (Collins et al., 2018). Additionally, they avoid the high cognitive load associated with comparing answers in a multiple-choice format (Collins et al., 2018). Each text had three questions that required either providing information found directly in the text (literal questions), generating a necessary inference (i.e., for text cohesion, a reader must make an inference that is not explicitly stated but is assumed by the author) or generating elaborative inference (an inference is made, but it is not necessary to understand the text). For example, in a text about the difficulties growing crops. Readers were asked the literal question "what did the farmers do to protect their crops?" the necessary inference question: "why were the farmers concerned about their crops? and elaborative inference question "how do you think the farmers feel?" To increase the likelihood participants understood and were able to respond to the questions, the examiner offered to read the questions aloud and participants were invited to respond in their preferred language.

Think-aloud data coding

Audio recordings of the think-aloud data were transcribed and coded for ten strategies (see Supplemental material for examples of each strategy in a think-aloud). Only one participant responded in a language other than English and their responses were translated to English. Each participant had four think-alouds per text and raters identified and tallied tokens of strategies in each think-aloud. Two raters met to calibrate their coding. To ensure coding remained consistent, think-alouds were examined in small batches (~10 participants at a time). In the child dataset, to verify the inter-rater reliability of the coding scheme, the second rater independently rated 30% ($N = 48$) of the participants. First and second raters' profiles of strategy use were compared (see below); However, the second rater reviewed the coding for all think-alouds, and the finalized coding was based on agreement from both coders. The procedure was slightly different for the adult data. Here, the primary rater's coding was verified by a second rater to finalize coding and a third rater independently coded a data subset (42%; 14 participants) for reliability.

To gain a profile of strategy use, frequency was calculated by tallying the number of times each strategy was identified for each participant. Inter-rater reliability was computed on the strategy profiles. Total count inter-observer agreement (Cooper et al., 2007) is the percentage agreement for each strategy (agreement/agreement + disagreement) averaged across each participant. Overall agreement was 80% in both the child and adult coding. Agreement of around 80% has been previously observed for think-aloud data (Chamot and El-Dinary, 1999). Finalized coding was based on the consensus of two raters. Importantly, the relative use of each strategy was captured by each rater. There was an average correlation of 0.89 between raters in the child data and 0.88 between raters in the adult data, indicating that raters were similarly distinguishing high users of a strategy from low users of a strategy.

Reading Comprehension responses were scored on a scale of 0 to 2 (*0 being incorrect, 1 being incomplete correct answer, 2 being complete correct answer)*. A scoring rubric was constructed for each question. For example, participants had to identify the two issues mentioned in the text about why farmers were concerned about crops (e.g., insects and weather) to receive two points. Identifying only one problem resulted in a single point. Similarly, in response to how farmers protect their crops, mentioning chemicals got a single point, stating that these chemicals were used to kill insects received two points. One rater initially scored each question, and their scoring was confirmed by a second rater. Disagreements were discussed and the rubric was refined if necessary and applied to all responses. Answers to the same question were compared directly to ensure similar responses were assigned the same scores. Responses from a randomly selected subset of participants (35%) were re-coded by a third rater using the rubric and 86% agreement was achieved. There were 3 questions per story with a potential maximum score of 24. A single score was generated since reading comprehension assessments regularly report a single score and include questions that require both literal and inferential information (Eason et al., 2012).

Procedure

Ethics approval was obtained from the university's non-medical research ethics review board and subsequently by the school board research committee. Data was collected in the spring (at the end of the academic year) of two consecutive school years from two separate cohorts. Following consent, each participant completed testing sessions individually. For the children, testing occurred in their school in a quiet space and was conducted over two sessions for ~30 min each. In the first session, students did the vocabulary measure (i.e., PPVT) and the word reading fluency measure (i.e., TOWRE). In the second session, they completed the reading comprehension task. Of note, these tasks were part of a larger test battery in the schools. For the adults, the testing session took place at the University in a single session.

Results

The descriptive statistics for language measures are reported in Table 1. All values are raw scores. A multivariate analysis of variance with receptive vocabulary, total word reading fluency and reading comprehension scores as dependent measures was conducted. The analysis met the assumptions for no multicollinearity. The Mahalanobis distance revealed one outlier that was subsequently removed from the multivariate analysis. Box' M and Levene's tests were significant, indicating that the assumption of normality was violated. To address this concern, Pillai's trace was used and a p-value of 0.001 was set for both the multivariate and univariate analyses (Allen and Bennett, 2008).

The multivariate analysis revealed an overall main effect of language group, $F(6, 376) = 7.32$, $p < 0.001$, $n_p^2 = 0.11$. Univariate ANOVAs (analysis of variance) revealed that group differences were present in the receptive vocabulary measure, $F(2, 189) = 14.94$, $p < 0.001$, $n_p^2 = 0.14$, and the reading

TABLE 2 Mean and median sums of strategy use as a function of language group.

Strategies	Monolingual children		Bilingual children		Bilingual adults	
	Mean (SD)	Median	Mean (SD)	Median	Mean (SD)	Median
Vocabulary	0.7 (1.7)	0.0	**1.8 (3.7)**	**0.0**	0.3 (0.5)	0.0
Text Structure	0.5 (1.0)	0.0	0.7 (1.5)	0.0	**1.9 (2.0)**	**1.0**
Necessary Inferencing	7.7 (5.4)	6.0	6.9 (5.5)	6.0	**14.3 (6.0)**	**14.0**
Elaborative Inferencing	7.0 (6.7)	5.0	7.4 (7.2)	5.0	**10.5 (4.6)**	**10**
Connecting	1.0 (1.3)	1.0	1.0 (1.5)	0.0	**2.7 (2.2)**	**2.0**
Summarizing	3.8 (4.1)	2.0	5.4 (5.7)	3.0	**5.7 (4.5)**	**5.0**
Background Knowledge	2.5 (3.6)	1.0	1.5 (2.5)	1.0	**2.3 (1.7)**	**2.0**
Predicting	**4.9 (4.7)**	**4.0**	3.1 (3.8)	2.0	4.0 (3.2)	4.0
Questioning	3.1 (4.8)	1.0	2.2 (3.6)	1.0	2.1 (2.6)	1.0
Visualizing	1.9 (3.6)	0.0	1.3 (2.7)	0.0	0.8 (1.6)	0.0

Bolded values indicate which strategies differed among groups and which group used the strategy significantly more than at least one other group.

comprehension measure, $F(2, 189) = 9.10$, $p < 0.001$, $\eta_p^2 = 0.09$, but not in total word reading fluency, $F(2, 189) = 1.79$, $p = 0.171$, $\eta_p^2 = 0.02$. Sheffe *post-hoc* comparisons found that for receptive vocabulary, monolingual children outscored both the bilingual children, $p < 0.001$, and the bilingual adults, $p < 0.01$. No differences were observed between bilingual groups on vocabulary knowledge, $p = 0.75$. In contrast, on the reading comprehension measure, no differences were observed between the monolingual children and the bilingual adults, $p = 0.91$. However, both groups outperformed the bilingual children, *p*s < 0.02. Of note, average reading comprehension scores of ~50% appear to be low, but do in fact reflect that, on average, participants were providing partially correct answers.

Strategy recruitment

Table 2 reports the means, standard deviations, and medians for each strategy type by group. The distributions violated the assumption of normality and therefore Kruskal-Wallis H non-parametric tests were conducted on each strategy by language group. There were eight significant Kruskal-Wallis H tests: reference to vocabulary, $\chi^2(2) = 7.11$, $p < 0.05$, reference to text structure, $\chi^2(2) = 23.57$, $p < 0.001$, necessary inferencing, $\chi^2(2) = 35.16$, $p < 0.001$, elaborative inferencing, $\chi^2(2) = 16.01$, $p < 0.001$, connecting, $\chi^2(2) = 23.99$, $p < 0.001$, summarizing, $\chi^2(2) = 6.84$, $p < 0.05$, reference to background knowledge, $\chi^2(2) = 6.98$, $p < 0.05$, and predicting, $\chi^2(2) = 6.84$, $p < 0.05$. No main effect of group was observed in visualizing or questioning, *p*s > 0.05.

Post-hoc comparisons using Bonferroni adjustments revealed that for reference to vocabulary, the bilingual children reported marginally greater use of this strategy than the bilingual adults, $p = 0.07$, but not significantly more than monolingual children $p = 0.11$. Adult bilinguals reported significantly more reference to text structure, *p*s < 0.001, as well as greater use of necessary inferencing, *p*s < 0.001, elaborative inferencing, *p*s < 0.01, and connecting, *p*s < 0.001, that both child groups; no differences were observed between child groups on these strategies, *p*s > 0.05. For summarizing, the bilingual adults reported this strategy significantly more often than the monolingual children, $p < 0.05$, but not the bilingual children, $p > 0.05$. For background knowledge, bilingual adults expressed more reference to background knowledge than bilingual children, $p < 0.05$; no other group differences were observed on background knowledge. For predicting, monolingual children reported marginally greater use than bilingual children ($p = 0.06$) but no difference from bilingual adults, $p > 0.05$.

Predictors of reading comprehension success

Table 3 reports the bivariate correlations of reading comprehension with both language measures and strategies for each group and for the full sample. Partial correlations are also reported to examine whether the relationships between reading comprehension and strategy use remain when the influence of age, English receptive vocabulary and word reading are removed. For all groups, vocabulary correlated significantly with reading comprehension. Total word fluency correlated significantly with reading comprehension for the children only. Reference to text structure, connecting and necessary inferencing were all correlated to reading comprehension in each group. However, for the monolingual children, the effects disappeared for text structure and connecting when language measures and age were partialled out. Additionally, some differences across groups were observed. Elaborative inferencing was a significant correlate in the children but not the adults. Visualizing was significantly correlated for

TABLE 3 Correlations of language measures and strategy use with RC scores for all three language groups.

Measures	Monolingual children		Bilingual children		Bilingual adults		Full sample	
	Bivariate	Partial	Bivariate	Partial	Bivariate	Partial	Bivariate	Partial
Age	**0.39****	—	**0.51****	—	−0.01	—	−0.02	—
Language measures								
Vocabulary knowledge	**0.66*****	—	**0.77*****	—	**0.48****	—	**0.73*****	—
Word fluency total	**0.65*****	—	**0.61*****	—	0.12	—	**0.55*****	—
Strategies								
Text structure	**0.26***	0.07	**0.52*****	**0.39*****	**0.43****	**0.42***	**0.41*****	**0.33****
Connecting	**0.33****	0.19	**0.48*****	**0.34****	**0.52****	**0.47****	**0.42*****	**0.32*****
Necessary inferencing	**0.47*****	**0.30***	**0.66*****	**0.39*****	**0.50****	**0.45****	**0.55*****	**0.40*****
Elaborative inferencing	**0.52*****	**0.44****	**0.55*****	**0.42*****	0.18	0.33	**0.49*****	**0.44*****
Visualizing	**0.26***	**0.27***	0.20	−0.09	0.15	0.27	**0.21****	0.11
Vocabulary	0.04	0.04	0.03	**0.26***	0.28	0.23	−0.04	**0.20****
Background knowledge	0.17	0.11	**0.27****	**0.29****	0.12	0.05	**0.23****	**0.18***
Summarizing	0.03	0.09	0.14	0.13	**0.38***	0.33	0.09	**0.17****
Questioning	**0.31***	0.19	**0.23***	0.18	−0.15	0.01	**0.22****	**0.17***
Predicting	0.14	0.23	0.16	−0.15	0.29	0.32	**0.19****	0.04

*p < 0.05, **p < 0.01, ***p < 0.001. Bolded values denote significant correlations.

the monolingual children, whereas vocabulary and use of background knowledge were significant for the bilingual children in the partial correlations. Summarizing, questioning, and predicting were not significant partial correlates with reading comprehension for any groups. Taken together, as a full sample, most strategies correlated with reading comprehension to some degree, with the strongest correlations observed for inferencing behaviors, making connections within the text and commenting on text structure.

For the full sample, hierarchical linear regression analyses were performed to examine how strategy use accounted for reading comprehension beyond age, vocabulary knowledge and word reading fluency. To reduce the number of predictors, decrease any multiple collinearities and to determine the relationship between strategies themselves, a principal component factor analysis was conducted. The KMO measure of sampling adequacy value was 0.63 and deemed adequate. Four factors were generated that accounted for 65% of the variance. See Table 4 for the four-factor structure. Factor 1 consisted of making connections within the text and reference to text structure and necessary inferencing. It was called textbase strategies since these strategies involved making sense of the meaning units within the text itself. Factor 2 included reference to vocabulary, reference to background knowledge and questioning. This factor was named accessing knowledge since these behaviors involved accessing both lexical and semantic knowledge. It was often done in the context of questioning the text. Factor 3 was called elaboration as it pertained to elaboration both in terms of inferences but also in terms of creating visual imagery. Finally, Factor 4 (Prediction) consisted of prediction behaviors and failures to summarize. Here, readers were not engaged in behaviors grounded in the text but were anticipating upcoming events or information.

TABLE 4 Factors analysis components for the predictor variables.

Construct	Text base	Accessing knowledge	Elaboration	Prediction
Connecting	**0.86**	0.08	0.05	0.13
Text structure	**0.81**	0.09	−0.13	−0.03
Necessary inferencing	**0.64**	−0.24	0.44	−0.22
Vocabulary	0.08	**0.66**	−0.20	−0.20
Background knowledge	0.21	**0.65**	0.11	0.16
Questioning	−0.15	**0.62**	0.16	0.07
Elaborative inferencing	0.49	0.18	**0.60**	−0.04
Visualization	−0.12	0.03	**0.85**	0.02
Predicting	0.20	−0.14	0.05	**0.87**
Summarizing	0.37	−0.34	0.12	**−0.68**

Bolded values indicate that this variable loaded onto the corresponding factor.

Age, vocabulary, and word fluency were entered in the initial step of the hierarchical regression model (Core model) followed by the four strategy factors in the second step (Full model) using the enter input method. The model assumptions were met (e.g., appropriate sample size for number of predictors, no multicollinearity). Table 5 reports the core and full models. Both models were significant: core model, $R = 0.76$, $F(3, 188) = 89.68$, $p < 0.001$; full model, $R = 0.85$, $F(7, 184) = 67.92$, $p < 0.001$. The full model accounted for 71% of

TABLE 5 Reading comprehension regression models for the full sample.

Predictors	b	SE	β	*t*	Sig.
Core model					
Constant	−3.20	1.11		−2.88	0.004
Age	−0.015	0.005	−0.13	−2.66	0.008
Vocabulary knowledge	0.09	0.008	0.63	11.03	< 0.001
Word fluency	0.04	0.010	0.23	4.02	< 0.001
Final model					
Constant	−0.92	0.983		−0.94	0.35
Age	−0.001	0.005	−0.009	−0.21	0.83
Vocabulary knowledge	0.076	0.008	0.511	9.76	< 0.001
Word fluency	0.025	0.009	0.139	2.90	0.004
Textbase	1.65	0.219	0.340	7.52	< 0.001
Accessing knowledge	0.660	0.190	0.136	3.47	< 0.001
Elaboration	0.954	0.202	0.197	4.72	< 0.001
Prediction	−0.088	0.197	−0.018	−0.446	0.656

the variance in reading comprehension scores, an increase of 13.2% from the core model. The positive regression weights for vocabulary knowledge and word fluency remained when strategy use factors were input; age was no longer a significant predictor. Textbase, Accessing Knowledge and Elaboration were all significant positive predictors of reading comprehension scores with textbase strategies accounted for the most variance followed by Elaboration and then Accessing Knowledge.

Discussion

The present study investigated the relationships between reported strategy use, language knowledge and reading comprehension in three groups. Of interest was (1) how groups differed in their reported strategy use, (2) which strategies were associated with reading comprehension performance and (3) whether reported strategy use could explain unique variance in reading comprehension performance not explained by language variables. By including the three groups with different levels of language proficiency and reading experience, we gained insight into these variables' contributions to strategy use and reading comprehension success. We found that bilingual adults reported more inferencing behaviors, references to text structure and connecting behaviors than the children. Monolingual children made marginally more predictions than their bilingual peers. Finally, although receptive vocabulary knowledge and word reading fluency predicted reading comprehension scores, several strategy factors also accounted for unique variance in reading comprehension scores, with an emphasis on textbase strategies. Such findings highlight the contributions of strategic behaviors to reading comprehension success.

Our results are consistent with the SVR model (Hoover and Gough, 1990). Both word reading fluency and vocabulary were significant predictors of reading comprehension performance. Notably, word reading fluency was correlated with reading comprehension performance for the children only. This finding is consistent with work demonstrating that word reading abilities play a unique role earlier in reading comprehension development (e.g., Storch and Whitechurst, 2002; Proctor et al., 2006; Gunnerud et al., 2022) but less so for adults (Landi, 2010). However, increased word reading automaticity in the bilingual adults is an unlikely explanation since no differences in word reading fluency were observed among the three groups. A more likely possibility is that since our reading task was not speeded, adults modulated their reading speed better than the children.

Higher English receptive vocabulary was associated with better reading comprehension outcomes for all groups. This finding was consistent with previous research (e.g., Kendeou et al., 2009; Babayigit and Shapiro, 2020; Raudszus et al., 2021). Receptive vocabulary was among the strongest correlates of reading comprehension success for each group and as expected was the best predictor in the regression model. Of note, consistent with Melby-Lervåg and Lervåg's (2014) meta-analysis, vocabulary knowledge was a stronger predictor in for the bilingual children in their L2 than the monolingual children in their L1. Given that overall, our bilingual children had less English vocabulary knowledge, it is not surprising that the monolingual children produced higher reading comprehension scores; less L2 vocabulary knowledge makes reading comprehension tasks more challenging for second language learners (Melby-Lervåg and Lervåg, 2014).

In general, the monolingual and bilingual children reported similar strategies with some subtle differences. Monolingual children tended to state more predictions, whereas the bilingual children made more references to vocabulary words. Notably, these results are consistent with Frid and Friesen (2020) who observed that French Immersion students reported more predicting in their dominant language and referred more to vocabulary words in their non-dominant language. Jiménez et al. (1996) noted that identifying vocabulary words and summarizing are often recruited more in a less proficient language. Nonetheless, the overall similar pattern of reported strategy use among groups suggests that the main reason for the reading comprehension differences was due to English language proficiency, where receptive English vocabulary was used as a proxy here.

A comparison of the bilingual children and adults on language measures provides insights into differences in reading comprehension success. Despite being older, the bilingual adults did not differ from bilingual children on receptive English vocabulary knowledge but did outperform the bilingual children on reading comprehension. The lack of difference in English receptive vocabulary is likely because the bilingual children had spent significantly more time in an Anglophone community than the bilingual adults. The adults, in contrast, had significantly more overall literacy experience. They learned to read in their L1 and could use reading strategies to offset their lack of L2 knowledge. The benefit of strong L1 abilities has been originally detailed by Cummins (1981) in his linguistic interdependence hypothesis. Higher-order reading strategies learnt in the L1 can be used in L2 (assuming a minimum level of L2 proficiency). This common underlying proficiency presumes that academic competencies are shared across languages and that individuals can draw on higher-order skills such as analysis, integration, and reasoning in both languages. In contrast, the bilingual children were primarily developing reading skills and higher-order strategies in their L2 (as opposed to their L1) alongside their monolingual peers.

An examination of reported strategy use provides insights into why bilingual adults outscored the bilingual children on reading comprehension. The bilingual adults reported significantly more necessary inferences, elaborative inferences, reference to text structure and connecting behaviors than both groups of children. Importantly, these behaviors were correlated with reading comprehension in each group, providing evidence that overall greater use of these strategies likely facilitated reading comprehension in the bilingual adults. Likewise, these strategies were critical in accounting for unique variance in reading comprehension performance in the full sample. Of note, in Friesen and Frid (2021) when presented with challenging texts at and above their reading levels, English-French bilingual adults also greatly relied on making inferences. Unlike the current study, they also tended to favor more summarizing statements; likely to confirm their understanding.

Taken together, these behaviors are necessary to construct a comprehensive mental representation of the text. As described in the Construction-Integration Model (Kintsch, 2005), readers need to generate and select relevant inferences, and then integrate these meaning units by making connections. Additionally, the ability to identify the text structure enables a reader to create a scaffold on which to insert newly generated information (Gernsbacher et al., 1990; Cain, 2010). Readers who are aware of the text structure can anticipate upcoming information and then organize the information for later retrieval. A clear mental representation of the text involves understanding the relationship between ideas and organizing these ideas; doing so, enables better retrieval, and consequently, better reading comprehension performance.

Despite some overall group differences, there was also variability in reading comprehension performance within each language group. From an educator's perspective, an understanding of which strategies are associated with reading success may be sufficient for the classroom. The bivariate correlations in Table 3 provide insight into the likelihood that students will be successful on a subsequent reading test. Indeed, it was making connections, generating inferences and reference to text structure that are all markers of subsequent reading comprehension success for all readers. Looking for these behaviors during independent or guided reading may serve as a diagnostic or formative assessment of effective strategy use as readers build toward reading comprehension success. Importantly, individual strategies are not used in isolation and may be associated with each other in readers' repertoires (Frid and Friesen, 2021). Thus, isolating significant strategy use demonstrates that strategy use accounts for reading success beyond what is accounted for by language measures. This is important given previous work has failed to isolate unique contributions of strategy use and language ability (e.g., Lin and Yu, 2015; García and Godina, 2017). Here we observed that strategy use accounted for significant variance on top of the language measures.

Inferencing behavior was a significant correlate of reading comprehension performance. The robust nature of these findings highlights the importance of generating inferences as part of developing a situation model (i.e., a meaning-based representation that links text content to the reader's previous knowledge). Indeed, previous work has isolated offline inferential abilities as predictors of reading success (e.g., Oakhill et al., 2003; Ahmed et al., 2016). Raudszus et al. (2019) reported that the ability to build a situation model accounted for significant variance in reading comprehension beyond linguistic and cognitive predictors in both bilingual and monolingual Grade 3 students. Here we demonstrate that greater articulation of inferences during reading is also associated with better performance on a subsequent test that necessitates this inferential knowledge. Importantly, directly teaching inferential skills has been shown to improve reading comprehension performance (Silverman et al., 2014). Silverman et al. (2014) found that teachers' use of instruction that targeted inferential comprehension was positively associated with reading comprehension gains in both monolinguals and bilinguals. Importantly, students should be taught how to engage in effective think-alouds to promote effective strategy use (Kim and Cha, 2015; Friesen and Haigh, 2018) and consolidate their inferences into long-term memory.

In the regression analysis, three of the four strategy factors were positively related to reading comprehension performance. Like previous research (Frid and Friesen, 2020), the factor associated with building a mental representation of the text through constructing and integrating meaning units (i.e., textbase) was the second strongest predictor of reading comprehension performance after vocabulary knowledge. Elaborative behaviors and accessing knowledge were also associated with reading comprehension success but to a less extent. Finally, predicting behaviors were not particularly beneficial. These finding confirms previous work (i.e., Duke and Pearson, 2009; Frid and Friesen, 2020), wherein predicting by itself was not a significant predictor of reading comprehension success in children. Friesen and Frid (2021) reported that with adults, there is a greater tendency to make both predictions and then explicitly connect back to their predictions; this behavior is associated with greater reading comprehension success.

Implications

Our results demonstrate that reported strategy use is associated with reading success beyond language knowledge. Of interest to educators is how to support the development of these skills alongside language instruction. Given the rising numbers of English language learners in schools (Census Canada, 2017), an important avenue of future research will be to understand how strategy use and language proficiency develop interactively throughout schooling. Since English vocabulary knowledge was strongly related to reading comprehension success for all readers, continued language and vocabulary development should facilitate reading comprehension success (see Babayigit and Shapiro, 2020). However, explicit instruction on how to utilize strategies together to build a mental representation of the text is clearly warranted, particularly given the variability in strategy use and reading comprehension performance within all groups. One suggestion would be to jointly work on gaining language proficiency and strategy use by scaffolding language instruction (e.g., guided reading) to focus directly on strategy development.

A few considerations become key when determining what strategies to teach to support reading comprehension performance. Our research and previous research have found that questioning (e.g., Yopp, 1988), visualization (e.g., Pressley, 2000), and reliance on text structure (Oakhill et al., 2003) are all associated with greater comprehension success. However, our work demonstrated that these strategies were less frequently reported and as such may have served as markers of comprehension rather than fully realized strategies in the readers' repertoires (see also Frid and Friesen, 2020 and Friesen and Frid, 2021). Consequently, there may have been unrealized strategies that would have increased comprehension performance that readers in general fail to report or to use. For educators, assessing which strategies each individual student is using becomes essential to understanding which strategies require additional support and which strategies require direct instruction (Friesen and Haigh, 2018).

Our implications should be considered in light of the study's limitations. The correlational nature of the current study makes caution necessary in recommending specific strategies to teach based on correlations or regression models. Likely, there is a bidirectional relationship between reading comprehension performance and reported strategy use, such that a reader's comprehension dictates the strategies that they can report. But in return, the selection of effective strategies consolidates content in memory. Importantly, the act of doing a think-aloud may increase processing beyond what is expected during silent reading. For some readers, this opportunity may be beneficial for reading comprehension. However, for others doing think-alouds may negatively impact comprehension, particularly if respondents were not employing strategies that supported consolidating content in memory. Future research could examine how strategy use reported in think-alouds are related to performance on a different reading comprehension assessment as this approach would reduce concerns about the think-alouds impacting the assessment of reading comprehension ability.

Another concern is that a reader may choose to only report a subset of their thoughts due to perceived time constraints. For example, one possibility is that more successful comprehenders are providing elaborative inferences because they have the foundational skills (e.g., summarizing, drawing necessary inferences) required to think beyond the text and consequently elaborative inferencing stands as the representative of a group of strategies. Thus, it is likely a constellation of strategic behaviors that support comprehension. Indeed, findings from both the current study as well as Frid and Friesen (2021) imply that utilizing strategies in concert is particularly effective in generating a comprehensive mental text representation upon which to base reading comprehension performance.

An additional consideration for teachers is the selection of reading comprehension test format and what knowledge or skills in addition to reading comprehension that it may be assessing. Unfortunately, research has demonstrated that even with standardized measures, only moderate correlations exist between test formats (Keenan and Meenan, 2014) and thus tests tap into different underlying skills in addition to reading comprehension (Spencer et al., 2019). For example, Carlisle and Rice (2004) found that cloze test performance is associated with semantic understanding at the sentence level rather than at the text-level. Given that preferred reading comprehension questions require students to draw on a deeper understanding of the text (Spencer et al., 2019), we selected to use open-ended questions that required drawing conclusions to be successful, and subsequently found that reliance on inferencing, connecting, reference to text structure was particularly fruitful for our reading comprehension measure. It remains to be seen if reliance on these strategies would be as successful with other response formats. Ideally, educators should be mindful of the alignment between the reading comprehension strategies they are teaching and how they are assessing students' understanding.

In conclusion, the current study was able to identify reported strategies that underlie successful reading comprehension in two ways. This was accomplished, first, by examining strategy differences between groups that were matched on either vocabulary knowledge or reading comprehension performance. Secondly, this question was addressed by examining which strategies were associated with reading comprehension performance and which strategies accounted for unique variance beyond vocabulary knowledge, age and word reading fluency. Strategies that enabled readers to identify implicit meaning (i.e., inferences), integrate meaning across the text (i.e., connections) and organize their knowledge (i.e., reference to text structure) served readers best in encoding and retrieving knowledge.

Data availability statement

The datasets presented in this article are not readily available because at the request of the participating school board, guardians' consent was not obtained to share their child's datasets widely. Requests to access the datasets should be directed to DF; Deanna.Friesen@uwo.ca.

Ethics statement

The studies involving human participants were reviewed and approved by Western University: Non-Medical Ethics Review Board. For the child participants, written informed consent to participate in this study was provided by the participants' legal guardian/next of kin. Adult participants provided written informed consent.

Author contributions

DF, KS, and TA contributed to conception and design of the study. A subset of children's data was collected and analyzed by KS for their MA Thesis. The adult data was collected and analyzed by TA for their MA thesis. DF and AC developed the coding scheme. AC coded the majority of the data. DF conducted the combined statistical analyses and wrote the first draft of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

Funding

This research was supported by a Social Sciences and Humanities Research Board Seed Research Grant from Western University. It was also supported by an Social Sciences and Humanities Research Council of Canada Insight grant (435-2018-442).

Acknowledgments

We wish to thank Kayla Edwards, Christina Amico, Tsz-Wing Zita Lau, and Bailey Frid for their help with data collection and processing.

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.

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.2022.986937/full#supplementary-material

References

Aarts, R., and Verhoeven, L. (1999). Literacy attainment in second language submersion context. *Appl. Psycholinguist.* 20, 377–393. doi: 10.1017/S0142716499003033

Afflerbach, P., Pearson, P. D., and Paris, S. G. (2008). Clarifying differences between reading skills and reading strategies. *Read. Teach.* 61, 364–373. doi: 10.1598/RT.61.5.1

Afsharrad, M., and Benis, A. R. S. (2017). Differences between monolinguals and bilinguals/males and females in English reading comprehension and reading strategy use. *Int. J. Biling. Educ. Biling.* 20, 34–51. doi: 10.1080/13670050.2015.1037238

Ahmed, Y., Francis, D. J., York, M., Fletcher, J. M., Barnes, M., and Kulesz, P. (2016). Validation of the direct and inferential mediation (DIME) model of RC in grades 7 through 12. *Contemp. Educ. Psychol.* 44–45, 68–82. doi: 10.1016/j.cedpsych.2016.02.002

Allen, P., and Bennett, K. (2008). *SPSS for the Health & Behavioural Sciences*. Edn 1. Victoria, BC: Thomson Learning.

August, D., and Shanahan, T. (2006). *Developing Literacy in Second-Language Learners: Report of the National Literacy Panel on Language-Minority Children and Youth*. Mahwah, NJ: Lawrence Erlbaum Associates.

Babayigit, S., and Shapiro, L. (2020). Component skills that underpin listening comprehension and reading comprehension in learners with English as a first and additional language. *J. Res. Read.* 43, 78–97. doi: 10.1111/1467-9817.12291

Bialystok, E., Luk, G., Peets, K. F., and Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingual. Lang. Cognit.* 13, 525–531. doi: 10.1017/S1366728909990423

Block, E. (1986). The comprehension strategies of second language readers. *TESOL Q.* 20, 463–494. doi: 10.2307/3586295

Brantmeier, C. A. (2002). Second language reading strategy research at the secondary and university levels: Variations, disparities, and generalizability. *Read. Matrix* 2, 1–14.

Brown, R. (2017). "Comprehension strategies instruction for learners of English: Where we have been, where we are now, where we might go," in *Handbook of Research in Reading Comprehension*, ed S. E. Israel (New York, NY: The Guilford Press).

Cain, K. (2010). *Reading Development and Difficulties*. Cornwall: Blackwell Publishing.

Carlisle, J. F., and Rice, M. S. (2004). "Assessment of reading comprehension," in *Handbook of Language and Literacy: Development and Disorders*, eds C. A. Stone, E. R. Silliman, B. J. Ehren, and K. Apel (New York, NY: Guilford Press).

Carrell, P. L. (1989). Metacognitive awareness and second language reading. *Modern Lang. J.* 73, 121–134. doi: 10.1111/j.1540-4781.1989.tb02534.x

Castles, A., Rastle, K., and Nation, K. (2018). Ending the reading wars: Reading acquisition from novice to expert. *Psychol. Sci. Public Interest* 19, 5–51. doi: 10.1177/1529100618772271

Census Canada (2017). *Immigration and Diversity: Population Projections for Canada and its Regions, 2011 to 2036*. Available online at: http://www.statcan.gc.ca/pub/91-551-x/91-551-x2017001-eng.htm

Chamot, A. U., and El-Dinary, P. B. (1999). Children's learning strategies in language immersion classrooms. *Modern Lang. J.* 83, 319–338. doi: 10.1111/0026-7902.00025

Cho, E., Toste, J. R., Lee, M., and Ju, U. (2019). Motivational predictors of struggling readers' reading comprehension: The effects of mindset, achievement goals, and engagement. *Read. Writi.* 32, 1219–1242. doi: 10.1007/s11145-018-9908-8

Collins, A. A., Lindström, E. R., and Compton, D. L. (2018). Comparing students with and without reading difficulties on reading comprehension assessments: A meta-analysis. *J. Learn. Disabil.* 51, 108–123. doi: 10.1177/0022219417704636

Cooper, J. O., Heron, T. E., and Heward, W. L. (2007). *Applied Behavior Analysis (2nd ed.)*. Columbus, OH: Pearson.

Cummins, J. (1981). Empirical and theoretical underpinnings of bilingual education. *J. Educ.* 163, 16–29. doi: 10.1177/002205748116300104

Droop, M., and Verhoeven, L. (2003). Language proficiency and reading ability in first- and second-language learners. *Read. Res. Q.* 38, 78–103. doi: 10.1598/RRQ.38.1.4

Duke, N. K., and Pearson, P. D. (2009). Effective practices for developing reading comprehension. *J. Educ.* 189, 107–122. doi: 10.1177/0022057409189001-208

Dunn, L. M., and Dunn, L. M. (1997). *Examiner's Manual for the PPVT-III Peabody Picture Vocabulary Test: Form IIIA and Form IIIB*. AGS. doi: 10.1037/t15145-000

Eason, S. H., Lindsay, F., Goldberg, K. M., Young, M. C., Geist, and Cutting, L. E. (2012). Reader–text interactions: How differential text and question types influence cognitive skills needed for reading comprehension. *J. Educ. Psychol.* 104, 515–528. doi: 10.1037/a0027182

Erfani, S. M., Iranmehr, A., and Davari, H. (2011). Deepening ESP reading comprehension through visualization. *J. Lang. Teach. Res.* 2, 270–273. doi: 10.4304/jltr.2.1.270-273

Farnia, F., and Geva, E. (2013). Growth and predictors of change in English language learners' reading comprehension. *J. Res. Read.* 36, 389–421. doi: 10.1111/jrir.12003

Frid, B., and Friesen, D. C. (2020). Reading comprehension and strategy use in fourth and fifth grade French Immersion students. *Read. Writ.* 33, 1213–1233. doi: 10.1007/s11145-019-10004-5

Frid, B., and Friesen, D. C. (2021). An investigation of reading strategy patterns recruited by bilingual adults and children. *Canad. Modern Lang. Rev.* 77, 212–233. doi: 10.3138/cmlr-2020-0039

Friesen, D. C., and Frid, B. A. (2021). Predictors of successful reading comprehension in bilingual adults: The role of strategies and language proficiency. *Languages* 6:18. doi: 10.3390/languages6010018

Friesen, D. C., and Haigh, C. A. (2018). How and why strategy instruction can improve second language RC: A review. *Read. Matrix* 18, 1–18.

García, G. E., and Godina, H. (2017). A window into bilingual reading: the bilingual reading practices of fourth-grade, Mexican American children who are emergent bilinguals. *J. Literacy Res.* 49, 273–301. doi: 10.1177/1086296X17703727

Gernsbacher, M. A., Varner, K. R., and Faust, M. E. (1990). Investigating differences in general comprehension skill. *J. Experi. Psychol.* 16, 430–445. doi: 10.1037/0278-7393.16.3.430

Geva, E., and Farnia, F. (2012). Developmental changes in the nature of language proficiency and reading fluency paint a more complex view of RC in ELL and EL1. *Read. Writ.* 25, 1819–1845. doi: 10.1007/s11145-011-9333-8

Gough, P. B., Hoover, W. A., and Peterson, C. L. (1996). "Some observations on a simple view of reading," in: *Reading Comprehension Difficulties: Processes and Interventions*, eds C. Cornoldi and J. Oakhill (Mahwah, NJ: Lawrence Erlbaum Associates).

Gunnerud, H. L., Foldnes, N., and Melby-Lervåg, M. (2022). Levels of skills and predictive patterns of reading comprehension in bilingual children with an early age of acquisition. *Read. Writ.* 54:e10286. doi: 10.1007/s11145-022-10286-2

Hoover, W., and Gough, P. (1990). The simple view of reading. *Read. Writ.* 2, 127–160. doi: 10.1007/BF00401799

Janzen, J., and Stoller, F. L. (1998). Integrating strategic reading in L2 instruction. *Read. Foreign Lang.* 12, 251–269.

Jiménez, R. T., García, G. E., and Pearson, P. D. (1996). The reading strategies of bilingual Latina/o students who are successful English readers: Opportunities and obstacles. *Read. Res. Q.* 31, 90–112. doi: 10.1598/RRQ.31.1.5

Joshi, R. M., and Aaron, P. G. (2000). The component model of reading: Simple view of reading made a little more complex. *Read. Psychol.* 21, 85–97. doi: 10.1080/02702710050084428

Keenan, J. M., and Meenan, C. E. (2014). Test differences in diagnosing reading comprehension deficits. *J. Learn. Disabil.* 47, 125–135. doi: 10.1177/0022219412439326

Kendeou, P., van den Brock, P., White, M. J., and Lynch, J. S. (2009). Predicting reading comprehension in early elementary school: The independent contributions oral language and decoding skills. *J. Educ. Psychol.* 101, 765–778. doi: 10.1037/a0015956

Kim, H.-I., and Cha, K.-A. (2015). Korean learners' metacognition in reading using think-aloud procedures with a focus on regulation of cognition. *Engl. Lang. Teach.* 8, 178–193. doi: 10.5539/elt.v8n6p178

Kintsch, W. (2005). An overview of top-down and bottom-up effects in comprehension: The CI perspective. *Discourse Process.* 39,125–128. doi: 10.1207/s15326950dp3902&3_2

Kolić-Vehovec, S., and Bajšanski, I. (2007). Comprehension monitoring and reading comprehension in bilingual students. *J. Res. Read.* 30, 198–211. doi: 10.1111/j.1467-9817.2006.00319.x

Landi, N. (2010). An examination of the relationship between reading comprehension, higher-level and lower-level reading sub-skills in adults. *Read. Writ.* 23, 701–717. doi: 10.1007/s11145-009-9180-z

Lin, L.-C., and Yu, W.-Y. (2015). A think-aloud study of strategy use by EFL college readers reading Chinese and English texts. *J. Res. Read.* 38, 286–306. doi: 10.1111/1467-9817.12012

McNamara, D. S. (2012). *Reading Comprehension Strategies: Theories, Interventions, and Technologies*. New York, NY: Lawrence Erlbaum Associates.

Melby-Lervåg, M., and Lervåg, A. (2014). Reading comprehension and its underlying components in second language learners: A meta-analysis of studies comparing first- and second-language learners. *Psychol. Bull.* 140, 409–433. doi: 10.1037/a0033890

Mokhtari, K., and Reichard, C. (2004). Investigating the strategic reading processes of first and second language readers in two different cultural contexts. *System* 32, 379–394. doi: 10.1016/j.system.2004.04.005

Oakhill, J., Cain, K., and Bryant, P. (2003). The dissociation of word reading and text comprehension: Evidence from component skills. *Lang. Cogn. Process.* 18, 443–468. doi: 10.1080/01690960344000008

Ontario Ministry of Education (2006). *The Ontario Curriculum Grades 1-8: Language*. Available online at: www.edu.gov.on.ca/eng/curriculum/elementary/language18currb.pdf

Padrón, Y. N. (1992). The effects of strategy instruction on bilingual students' cognitive strategy use in reading. *Biling. Res. J.* 16, 35–51. doi: 10.1080/15235882.1992.10162636

Park, H. R., and Kim, D. (2015). English language learners' strategies for reading computer-based texts at home and in school. *CALICO J.* 2015, 380–410. doi: 10.1558/cj.v33i3.26552

Perfetti, C., and Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Sci. Stud. Read.*18, 22–37. doi: 10.1080/10888438.2013.827687

Plonsky, L. (2011). The effectiveness of second language strategy instruction: A meta-analysis. *Lang. Learn.* 61, 993–1038. doi: 10.1111/j.1467-9922.2011.00663.x

Pressley, M. (2000). "What should comprehension instruction be the instruction of?" in: *Handbook of Reading Research: Volume III*, eds M. L. Kamil, P. B.

Mosenthal, P. D. Pearson, and R. Barr (Mahwah, NJ: Lawrence Earlbaum Associates, Inc.).

Proctor, C. P., August, D., Carlo, M. S., and Snow, C. E. (2006). The intriguing role of Spanish language vocabulary knowledge in predicting English RC. *J. Educ. Psychol.* 98, 159–169. doi: 10.1037/0022-0663.98.1.159

Proctor, C. P., Carlo, M., August, D., and Snow, C. (2005). Native Spanish-speaking children reading in English: Toward a model of comprehension. *J. Educ. Psychol.* 97, 246–256. doi: 10.1037/0022-0663.97.2.246

Rand Reading Study Group (2002). *Reading for Understanding: Toward an R&D Program in RC*. Santa Monica, CA: RAND.

Raudszus, H., Segers, E., and Verhoeven, L. (2019). Situation model building ability uniquely predicts first and second language reading comprehension. *J. Neurolinguistics* 50, 106–119. doi: 10.1016/j.jneuroling.2018.11.003

Raudszus, H., Segers, E., and Verhoeven, L. (2021). Patterns and predictors of reading comprehension growth in first and second language readers. *J. Res. Read.* 2021, 1–18. doi: 10.1111/1467-9817.12347

Saricoban, A. (2002). Reading strategies of successful readers through the three-phase approach. *Read. Matrix* 2, 1-14.

Silverman, R. D., Proctor, C. P., Harring, J. R., Doyle, B., Mitchell, M. A., and Meyer, A. G. (2014). Teachers' instruction and students' vocabulary and comprehension: An exploratory study with English monolingual and Spanish–English bilingual students in grades 3–5. *Read. Res. Q.* 49, 31–60. doi: 10.1002/rrq.63

Spencer, M., Gilmour, A. F., Miller, A. C., Emerson, A. M., Saha, N. M., and Cutting, L. E. (2019). Understanding the influence of text complexity and question type on reading outcomes. *Read. Writ.* 32, 603–637. doi: 10.1007/s11145-018-9883-0

Storch, S. A., and Whitechurst, G. J. (2002). Oral language and code-related precursors to reading. Evidence from a longitudinal model. *Dev. Psychol.* 38, 934–947. doi: 10.1037/0012-1649.38.6.934

Torgesen, J. K., Wagner, R. K., and Rashotte, C. A. (1999). *Test of Word Reading Efficiency (TOWRE)*. Austin, TX: Pro-ed.

van Steensel, R., Oostdam, R., van Gelderen, A., and van Schooten, E. (2016). The role of word decoding, vocabulary knowledge and meta-cognitive knowledge in monolingual and bilingual low-achieving adolescents' reading comprehension. *J. Res. Read.* 39, 312–329. doi: 10.1111/1467-9817.12042

Wiederholt, J. L., and Bryant, B. R. (2001). *Gray oral reading tests (4th ed.)*. Austin, TX: Pro-Ed.

Yopp, R. E. (1988). Questioning and active comprehension. *Question. Exch.* 2, 231–238.

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 24 November 2022
DOI 10.3389/fpsyg.2022.1003199

Check for updates

OPEN ACCESS

EDITED BY
Maria Garraffa,
University of East Anglia, United Kingdom

REVIEWED BY
Elena Nicoladis,
University of Alberta,
Canada
Agnieszka Otwinowska-Kasztelanic,
University of Warsaw,
Poland

*CORRESPONDENCE
Yanyan Chen
flcyy@scut.edu.cn

SPECIALTY SECTION
This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Psychology

RECEIVED 26 July 2022
ACCEPTED 11 October 2022
PUBLISHED 24 November 2022

CITATION
Xue H, Deng R, Chen Y and
Zheng W (2022) How does bilingual
experience influence novel word learning?
Evidence from comparing L1-L3 and L2-L3
cognate status.
Front. Psychol. 13:1003199.
doi: 10.3389/fpsyg.2022.1003199

How does bilingual experience influence novel word learning? Evidence from comparing L1-L3 and L2-L3 cognate status

Heng Xue[1], Renhua Deng[2], Yanyan Chen[2]* and Wenxin Zheng[3]

[1]College of Education, Hebei Normal University, Shijiazhuang, China, [2]School of Foreign Languages, South China University of Technology, Guangzhou, China, [3]Sino-Danish College, University of Chinese Academy of Sciences, Beijing, China

Bilingual experience exerts a complex influence on novel word learning, including the direct effects of transferable prior knowledge and learning skill. However, the facilitation and interference mechanism of such influence has largely been tangled by the similarity of the previously learned word knowledge. The present study compared Chinese-English bilinguals' paired-associate learning of nonwords in logographic and alphabetic writing systems. The logographic nonwords resemble the form and meaning of L1 Chinese words in varying degrees, being cognates, false cognates, or non-cognates of Chinese. The alphabetic nonwords resemble the form and meaning of L2 English words, being cognates, false cognates, or non-cognates of English. The learning sequence of logographic and alphabetic words was cross-balanced. The learning results were measured in production and recognition tasks. As for learning the logographic nonwords, both the recognition and production results showed that cognates were learned significantly faster than the non-cognates, and the false cognates were also learned significantly faster than the non-cognates. This suggests stronger facilitation rather than interference from L1 on novel word learning. As for learning the alphabetic nonwords, both the recognition and production results revealed that cognates were learned significantly faster than the non-cognates, but false cognates showed no advantage over the non-cognates. This indicates that interference from L2 is stronger than that from L1. Taken together, the results provide new evidence for the dissociable facilitation and interference effects of bilingual experience. These results carry potential educational implications in that learning novel words depends on substantial bilingual experience.

KEYWORDS

bilingual experience, cross-linguistic similarity, facilitation, interference, L1-L3 cognate status, L2-L3 cognate status

Introduction

Bilingual experience is one of the main factors that makes word learning different between bilinguals and monolinguals (Del Pilar Agustín-Llach, 2019; Liu et al., 2020; Hirosh and Degani, 2021). When learning a novel word, bilingual experience mainly refers to bilinguals' extensive experience of mapping novel word form to known concept based on the prior knowledge of the first language (L1) and the second language (L2) as well as their accumulative learning skill of form-meaning mapping (Bartolotti and Marian, 2017; Hirosh and Degani, 2018). Though bilinguals have been found to outperform their monolingual counterparts in vocabulary learning, the bilingual experience of L1 and L2 can exert facilitation or interference effects quite differently (Kaushanskaya et al., 2012; Bartolotti and Marian, 2017). However, the existing research on foreign vocabulary learning is mainly based on the alphabetic writing system (e.g., Bartolotti and Marian, 2019; Zhang et al., 2019; Otwinowska et al., 2020; Vanlangendonck et al., 2020). There are few empirical studies on logographic writing system and even less on both logographic and alphabetic writing systems (though discussed by Ruan et al., 2017; Mok et al., 2018; Liu et al., 2018; Richlan, 2020). Thus, the facilitation or interference effect of bilingual experience may be tangled by the similarity of writing systems (Eng et al., 2019; Jiang, 2021). In this study, 41 Chinese-English bilinguals participated in our experiment to learn logographic and alphabetic novel words through paired-associate learning and completed recognition and production tasks to measure the learning outcomes (Marecka et al., 2021). The logographic nonwords share different degrees of form and meaning overlapping with L1 Chinese, such as " 婚纱." It refers to "wedding dress" in English and slightly differs in radicals from the original L1 Chinese word "婚纱." The alphabetic nonwords share different degrees of form and meaning overlapping with L2 English, such as "pandda." It refers to "panda" in English but slightly differs in the spelling of the original L2 word "panda." This study is unique as the influence from L1 and L2 can be disentangled from learning alphabetic or logographic novel words based on a within-subject experiment design, contributing to identifying the facilitative and interferent mechanism of bilingual experience on novel word learning.

Learning novel words of different cognate status

Learning a new word in a foreign language means acquiring knowledge of the word form and mapping the form to the concept (Schmitt, 2019; Nation, 2020). To measure the learning outcome of such knowledge, both recognition and productive aspects are assessed. The recognition task is used to access whether the learner can recognize the form-meaning mapping of a word, whereas the production task evaluates the learner's ability to produce the word form. Existing studies have basically confirmed that specificity, frequency, and word presentation have an impact on novel word learning (Wang et al., 2020). Thus, it is necessary to use nonwords or artificial words, controlling their semantic specificity, logographic, and other essential information to study the learning effect.

Words to be learned in a foreign language may share varying degrees of overlap with the previously-learned words and thus can be classified into three types of cognate status, i.e., cognate, false cognate, and non-cognate (Simpson Baird et al., 2016; García et al., 2020). A cognate is a word whose form and meaning are almost the same in two different languages, such as the Chinese word "小说" and the Japanese word "小説" (both refer to the meaning of "fiction") as well as the English word "actor" and the French word "acteur" (both refer to the meaning of "actor"). False cognates refer to two words in different languages that have quite similar forms but have different meanings, such as the word "大手" in Chinese and Japanese (refers to "big hands" in Chinese and "large enterprises" in Japanese) as well as the word "magazine" in English and "magasin" in French (it refers to "magazine" in English and "shop" in French). Non-cognates are words that do not share a significant formal similarity with L1 or L2 words. Thus, the cognates are well-matched with the prior language experience and the false cognates are the mismatched ones, when non-cognates are used as baselines for comparison (Iniesta et al., 2021; Marecka et al., 2021).

Cognate status is a well-explored topic in foreign language learning (Van Hell and De Groot, 1998; Hirosh and Degani, 2018). Studies have used non-identical spellings to replace identical spellings in experiments to identify the word form learning more precisely (Dijkstra et al., 2010; Arana et al., 2022). Cognates show advantages over non-cognates in recognition and production tasks in most studies (e.g., Lemhöfer et al., 2004; Van Hell and De Groot, 2008). Recently, the cognate facilitation effect has been reported to be moderated by the bilingual experience (Iniesta et al., 2021). Cross-language orthographic errors have been observed as evidence of cognate interference (Muscalu and Smiley, 2018). Besides, false cognates have also been used to clarify the influence of the previously-learned form and form-meaning mapping in bilingual experience (Marecka et al., 2021; Elias and Degani, 2022). Bilingual experience brings in the transferable knowledge and representations, and also the abilities to acquire the knowledge and create the representations (Hirosh and Degani, 2021; Marecka et al., 2021). Nevertheless, the form-meaning mismatch of false cognates inevitably costs extra efforts to differentiate and re-match the form-meaning mapping (Janke and Kolokante, 2015). Notably, the facilitation and interference from a previously learned language can be interwoven, competing to assist or hinder the novel word learning outcome. Therefore, learning L1-L3 and L2-L3 false cognates, respectively, can help explicate the subtle facilitation-and-interference mechanism of the bilingual experience.

10.3389/fpsyg.2022.1003199

Learning novel words with bilingual experience

Bilingual experience is formed by the accumulation of knowledge, acquisition, and regular use of two languages (Kroll et al., 2014; Subramaniapillai et al., 2019). Hirosh and Degani (2018) proposed a direct–indirect framework to clarify the effects of bilingual experience on learning novel languages. Direct effects include firstly those transferable knowledge and representations from known languages, and secondly the abilities to acquire the knowledge and create the representations. Indirect effects refer to the additional mediating role of bilingual experience as an advantage in a broader sense, such as cognitive and social abilities. Learning novel words in additional languages involves both the direct and indirect effects of bilingual experience. The direct effects critically depend on the degree of cross-linguistic similarity, i.e., the more similar the more direct effects (Antoniou et al., 2015). Besides, the direct effects also depend on the status of the two previously learned languages, i.e., the more frequently a particular pattern of mappings is experienced, the stronger learning advantage can be expected (Koda and Miller, 2018). Nevertheless, the indirect effects mainly refer to the bilinguals' learning advantage over their monolingual counterparts as well as the developmental changes of multilingual language learners. Therefore, in order to reveal the facilitation and interference of the known languages on the to-be-learned language, the direct effects should be the focus of research.

Studies so far provided little conclusive evidence on how the bilingual experience facilitates or interferes with subsequent word learning. Research has shown both L1 and L2 benefit further word learning. Bartolotti and Marian (2017) found that novel word learning benefited from both L1-English and L2-German, in which participants used an English keyword for Englishlike words and a German keyword for Germanlike words. Besides, a novel word's similarity to both L1 and L2 did not provide an additional learning benefit. The direct effects of bilingual experience may even be more complex. Mulík and Carrasco-Ortiz (2021) compared the phonological activation of L2 cognates and L1 cognates through event-related potentials (ERPs). Their research found that both L1 Spanish and L2 English facilitated learning novel L3 Slovak words in similar behavioral results but with different electrophysiological results. Evidence for a stronger facilitative role of L1 originated mainly from translation-related research. Hirosh and Degani (2021) found that bilinguals learned novel words better through L1 translations rather than L2 translations. Bogulski et al. (2019) also found that bilingual advantage in vocabulary learning depended on learning *via* the L1 or dominant language because learning *via* the L1 allows bilinguals to engage regulatory skills that benefit further vocabulary learning. Another line of research concerned the interferent effect of the bilingual experience. They found the L2 (status) effect rooted in the model of inhibitory control (Green, 1986, 1998; de Bot and Jaensch, 2015), which predicted the inhibition of the highly proficient language (usually L1) to retrieve the less developed languages (L2 and L3), thus leaving L2 and L3 in a competing condition with L1.

Furthermore, the role of bilingual experience can be tangled by the similarity among the languages learned and to be learned. Extensive studies have investigated monolinguals or bilinguals using alphabetic writing systems as L1 and even L2 experience, but more attention has been recently paid to those using logographic writing systems (see Table 1 for the relevant articles retrieved from Web of Science since 2017). Hsieh et al. (2017) found that although Japanese-Chinese bilinguals have a bilingual automatic activation mechanism similar to that of alphabetic bilinguals, the process of logographic recognition requires more neural mechanisms for semantic selection and suppression between cognates and false cognates. Zhang et al. (2021) found that Chinese-Japanese speakers' cognate awareness systematically predicts the vocabulary learning outcomes of Japanese words. Nevertheless, there is still a lack of research on how orthographically different L1 and L2, respectively, affect the learning outcomes of novel words similar to either L1 or L2. It would be more transparent to probe into the role of bilingual experience with participants of different writing systems.

The current study

The current study explored the role of prior bilingual experience in learning L3 alphabetic and logographic novel words. To this end, we used a word-learning experiment in which participants were continuously visually exposed to, tested on, and provided feedback about the forms and meanings of the target words (Marecka et al., 2021). Furthermore, we used artificially created logographic and alphabetic nonwords to disentangle L1-L3 cognate status and L2-L3 cognate status. The learning process of alphabetic and logographic words was investigated through cross-balancing within-subject design. Chinese-English bilingual participants learned the nonwords through paired pictures of objects which represented the novel words' meanings. The novel words included cognates, false cognates, and non-cognates of the learners' L1 Chinese and L2 English, respectively. It has been reported that at least 6 to 16 encounters with a word are needed to learn it (Nation, 2013). In our pilot study, the participants showed a decline in attention after 10 times of learning. Thus, our experiment paradigm exposed learners to each new word 9 times. The alphabetic and logographic learning blocks were the same. All the participants' learning sequences of alphabetic and logographic blocks were balanced. At the onset of the study, participants were presented with an exercise block to familiarize them with the production and recognition tasks. Then participants were presented with each target picture-word pair, one at a time. After this initial

TABLE 1 Major relevant articles concerning form-meaning mapping since 2017.

Author(s)	Publication year	Language experience			Writing system	Results	
		L1	L2	L3		Recognition	Production
Otwinowska and Szewczyk	2019	Polish	English		Same	Translation: C > NC > FC	
Otwinowska et al.	2020	Polish	English		Same	Cognate awareness did not boost learning cognates.	
Marecka et al.	2021	Polish	Nonword		Same	C > FC ≈ NC	C > FC > NC
Iniesta et al.	2021	Spanish	English		Same	Word dictation task English: C < NC; Spanish: C > NC	
Li and Golla	2021	Spanish	English		Same	Naming: C > NC	
Robinson Anthony et al.	2022	Spanish	English		Same	Language dominance was found to predict crosslinguistic (cognate) facilitation from Spanish to English.	
Muylle et al.	2022	Dutch	English		Same		C > NC
Allen	2019	Japanese	English		Different	Cognate frequency effect was found.	
Zhang et al.	2019	Chinese	English		Different	C > NC	
Bartolotti & Marian	2017	English	German	Nonword	Same	C > NC	C > NC
Cenoz et al.	2021	Basque	Spanish	English	Same	Cognate awareness did not boost learning cognates.	
Hirosh and Degani	2021	Hebrew	English	German	Different	With L2 translation (Error rates): C > FC > NC; with L1 translation (Error rates): C > NC ≈ FC; with L1, L2 translation (RTs): C > FC ≈ NC	With L1, L2 translation (Error rates, RTs): C > NC ≈ FC;

C = cognate; FC = false cognate; NC = non-cognate; ">" means the performance is better.

offline presentation, the tasks started, and the participants performed a series of production blocks interleaved with recognition blocks (see Figure 1 for a schematic overview of the paradigm design). In the recognition task, together with the target picture, three distractors were also presented, i.e., a semantic distractor, a graphic distractor, and a phonetic distractor. After each trial, the participants were given feedback on the accuracy of their response, so the production and recognition blocks both tested and trained the participants. The accuracy of the last round of the production task and the reaction times (RTs) of the correct answers in the last round of the recognition task was used for statistical analysis. One novel aspect of our study is that we combined testing of the alphabetic and logographic writing systems with the same participants, which has rarely been done in experiments so far.

Materials and methods

Experiment design

This experiment is based on the experiment designed by Marecka et al. (2021), which examined the Polish natives learning alphabetic nonwords. In the current experiment, Chinese-English bilinguals learned both the alphabetic and logographic new words. PsychoPy 3 (Peirce et al., 2019) was used to present the materials on a computer screen, a computer keyboard was used to collect the RTs in recognition tasks, and Han Wang electronic writing tablet served to record the results in production tasks. In this study, the dependent variables were the test scores of the results, i.e., accuracy in production tasks as well as RTs in recognition tasks. The independent variables were word types (cognate, false

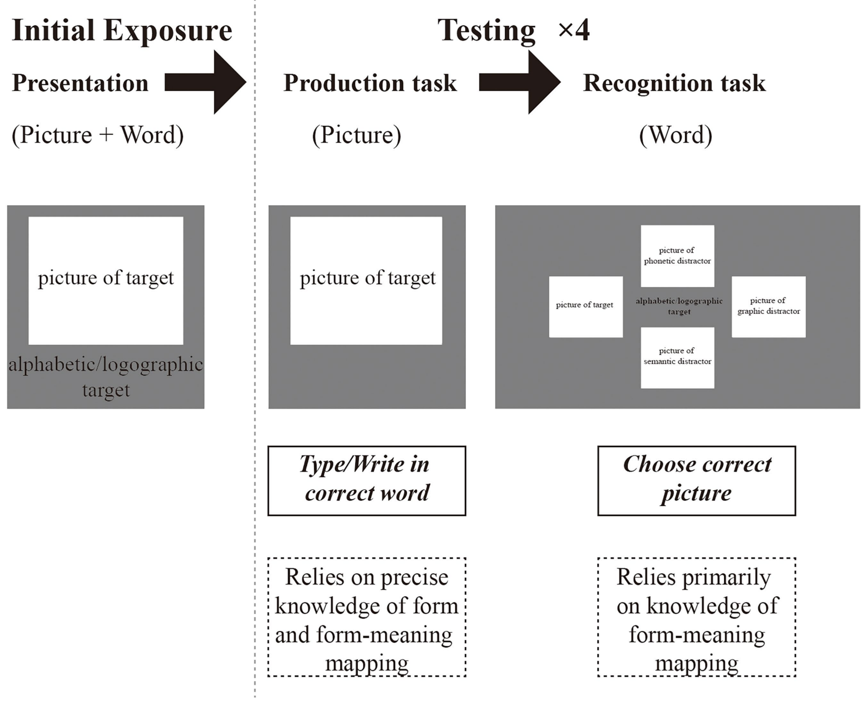


FIGURE 1
A schematic overview of the paradigm design.

cognate, and non-cognate) and writing systems (alphabetic and logographic).

The basic assumption of the research is that learners' prior language experience may influence L3 vocabulary learning differently. The learning outcomes of different word types should be compared within the logographic or the alphabetic division to reveal the language experience of L1 and L2. If L1-L3 cognates and L2-L3 cognates are learned faster than L1-L3 and L2-L3 non-cognates respectively, the bilingual experience can be proved to exert facilitative cross-linguistic influence independently. If L1-L3 false cognates and L2-L3 false cognates are learned faster than L1-L3 and L2-L3 non-cognates respectively, the bilingual experience can be proved to exert more facilitative rather than interferent cross-linguistic influence independently. If the patterns of learning outcomes differ between the writing systems, the bilingual experience can be proved to act in different modes.

Participants

Forty-one Chinese-English bilinguals aged from 18 to 25 were recruited for this study. All participants have been learning English at school since the age of 9 to 11. All participants did LexTALE (Lemhöfer and Broersma, 2012), a test of vocabulary knowledge for speakers of English as a second or foreign language. The results of LexTALE ranged from 43.75 to 87.5%, indicating that participants' proficiency ranged across three levels, i.e., upper advanced, upper intermediate, and lower intermediate (Lemhöfer and Broersma, 2012). They also finished a bilingual language use profile (Gertken et al., 2014) to research their everyday language use of alphabetic and logographic words. According to the language use profile results, all participants used Chinese as the dominant language.

Materials

Creating alphabetic stimuli

The target alphabetic stimuli were 24 nouns paired with pictures (6 cognates, 6 false cognates, and 12 non-cognates of L2 English as shown in Supplementary material). Based on the 2000 common English nouns in the Chinese education curriculum, this study selected nouns of 5 to 7 letters in length. By replacing, adding, or subtracting one letter, the alphabetic stimuli were words of 6 letters in length (Bartolotti and Marian, 2019). For example, "banana" can be changed into "benana." Firstly, approximately 300 English nouns were selected. Next, the words' concreteness and imageability were rated *via* a 5-point Likert scale (5 indicates the most concrete or imaginable) by a group of 8 Chinese-English bilinguals who would not participate in the following experiment. The concreteness or imageability ratings lower than 4 were eliminated. Then the words were put into CLEARPOND to search their frequency and neighbor size (Marian et al., 2012). Those frequencies over 25 occurrences per million and neighbor sizes over 4 were eliminated. Finally, their Chinese translations were matched, and their translation frequencies were log-transformed and controlled between 3.75 to 4.25 per million words in BLCU Chinese Corpus[1], and their translations were all two-character Chinese words. Only 30 nouns were reserved as the alternative meaning of the 24 nonwords in the experiment.

From the selected words, the 6 cognates and 6 false cognates were randomly assigned. The 6 cognates were matched to their original meaning in English. For example, "pandda" was assigned as cognates, meaning "panda." The meanings of the 6 false cognates were randomly selected among the 30 mostly concrete and imaginable nouns rated previously. False cognates' meanings differed remarkably from their initial meanings.

A hundred nonwords with 6 letters were first generated and their neighbor sizes were controlled to less than 4 in the ARC nonword database (Rastle et al., 2002). The non-cognates were 6 non-wordlike non-cognates and 6 wordlike non-cognates based on whether the form is similar to an English word. Finally, the meaning of the words selected before (30 nouns) was randomly assigned to the words generated in ARC.

Creating logographic stimuli

Similar to the alphabetic stimuli, the target logographic stimuli were 24 nouns paired with pictures (6 cognates, 6 false cognates, and 12 non-cognates of L1 Chinese as shown in Supplementary material). Based on the common Chinese nouns in the Chinese education curriculum, this study only used the two-characters nouns as the meaning of the 24 nonwords. Firstly, those frequencies were log-transformed and controlled between 3.75 to 4.25 per million words in BLCU Chinese Corpus. Next, the words' concreteness and imageability were rated *via* a 5-point Likert scale (5 indicates the most concrete or imaginable) by a group of 8 Chinese-English bilinguals who would not participate in the following experiment. The rating of imageability and concreteness less than 4 were deleted.

From the selected words, the meaning of 6 cognates, 6 false cognates, and 12 non-cognates was randomly assigned. Non-characters were created by randomly combining the phonetic and semantic radicals of the actual character stimuli following orthographic rules (Yum et al., 2014). Six cognates were created at first. For example, the form of the cognate "焒纱," which is the transformation of the "婚纱," was created by replacing the radical " 女" into "火." The form of the six false cognates was the same as the cognates. Differently, false cognates' meanings differed remarkably from their initial meanings. For example, the false cognate "烧烤" means "梨子" in the experiment, which is transformed from "烧烤" by replacing "火" into "女." Twelve non-cognates were divided into 6 wordlike non-cognates and 6 non-wordlike non-cognates. The wordlike non-cognates were transformed from a real Chinese word by changing two or three radicals, while the non-wordlike non-cognates did not follow the structure of Chinese words. The orthographic neighborhood size of 18 nonwords (6 cognates, 6 false cognates, and 6 wordlike non-cognates) was controlled between 20 and 30 (Dong et al., 2015). In addition, all the nonwords' strokes were controlled between 11 and 26. Finally, Truetype, a special character editing program in Windows 10, was used to present the nonwords in picture format.

Selecting associative pictures

Pictures were selected to indicate the meanings of the 48 logographic and alphabetic stimuli. Another 130 pictures were selected for the distractors in the recognition task. Pictures for the logographic and alphabetic stimuli were used both in the learning session and the recognition task in the test session. Pictures for the distractors were used only in the recognition task. All the pictures were from Cambridge online dictionary[2] and Bing picture database[3]. The pictures were piloted through an online translation task by 12 Chinese-English bilinguals who would not participate in the formal experiment. Thus, the pictures were validated that they were not ambiguous.

Procedure

Participants took part in a computerized word-learning task. They were asked to learn 24 alphabetic nonwords and 24 logographic nonwords, respectively. The interval between the alphabetic experiment and the logographic experiment was 1 day. The sequence of the alphabetic experiment and logographic experiment was balanced. The procedure could be divided into 4 parts, i.e., exercise trials, initial presentation block, production block, and recognition block. The sequence of the alphabetic and logographic experiments was balanced, and each participant's interval of the logographic and alphabetic experiment was 1 day.

The first block was the exercise trials. Five pictures were displayed in the center of the computer one after another. Participants were asked to write the correct word for the picture. It is similar to a production task. Then 5 words were displayed randomly in the center of the computer. Participants were asked to choose the correct picture using the keyboard. It is similar to the recognition task. All 5 words were irrelated and were actual words paired with pictures.

The second block was the initial presentation block. All 24 nonwords were displayed on the screen randomly with the pictures they represented. Before the presentation, the instruction had told participants to memorize the nonwords. Participants can press "space" to cut the instruction. Next, the screen presented with fixation "+" for 500 ms to remind participants to pay attention. Then on the center of the screen presented a picture with a nonword below it for 1,500 ms.

After the initial presentation block was the production block. All 24 nonwords were displayed on the screen randomly with the pictures. However, there was no time limit for it. Participants were told to write the answer (nonword) as much

1 https://www.blcu.edu.cn

2 https://dictionary.cambridge.org/us/dictionary/english/

3 https://cn.bing.com/images/

as they could or leave them blank if they could recall nothing. There was also a fixation pattern "+" for 500 ms before every picture to remind them. If the answer was correct, the feedback would present "Correct" for 500 ms, followed by the correct nonword and the corresponding picture for 500 ms. If the answer was wrong, the feedback would present "Sorry, you are wrong." for 500 ms, followed by the correct nonword and corresponding to the picture for 500 ms. It is worth noting that only the alphabetic production task had the feedback but not in the logographic production task because the computer could not recognize the handwriting of logographic nonwords.

Next, was the recognition block. In the recognition block, participants were told to choose the corresponding picture to which the nonword refers. There were 4 kinds of pictures (see Figure 2), i.e., target (the correct item), phonetic distractor (the picture corresponding to the word which pronounces similarly to the nonword), semantic distractor (the picture corresponding to the word whose meaning is similar to the nonword), and graphic distractor (the picture corresponding the word which looks like to the nonword). They were randomly displayed on the nonword's top, left, down, and right. We counterbalanced the position of the four categories of pictures across trials and blocks. The participants were asked to press the arrow key to choose the picture. Each nonword was presented for 4,000 ms, meaning that participants must choose the answer in 4,000 ms or it would be regarded as wrong. After that was feedback similar to the production block, which included "Correct" or "Sorry, you are wrong." for 500 ms and the nonword with the correct picture for 1,500 ms. All 24 nonwords were in a random sequence.

The recognition and production test blocks were interleaved. There were 4 loops, which means 4 production blocks and 4 recognition blocks. Participants learned the nonwords 9 times in total (1 time in the initial presentation, 4 times in the production block, and 4 times in the recognition block).

The authors of this article looked through every incorrect response. Participants were interviewed after the experiments to make sure of their performance. The questions asked are presented below:

1. In the recognition task, why did you choose this incorrect answer?
2. In the recognition task, was there any difference in learning the alphabetic and logographic words? And why?
3. In the production task, what made you write the word in this incorrect way?
4. In the production task, was there any difference in learning the alphabetic and logographic words? And why?

Data analysis

Production data analysis (accuracy)

In the alphabetic production test blocks, this study calculated the normalized Levenshtein Distance (nLD) between the correct nonword and the nonword typed by the participants. The Levenshtein Distance (LD) is a metric that indexes the total number of insertions, deletions, and substitutions necessary to transform one string of letters onto another (Levenshtein, 1966). For example, the LD between the word "apple" and "epple" is 1 because they differ in one letter. The LD between the word "apple" and "abble" is 2 because they differ in two letters. The nLD is the LD between the two words divided by the number of target word

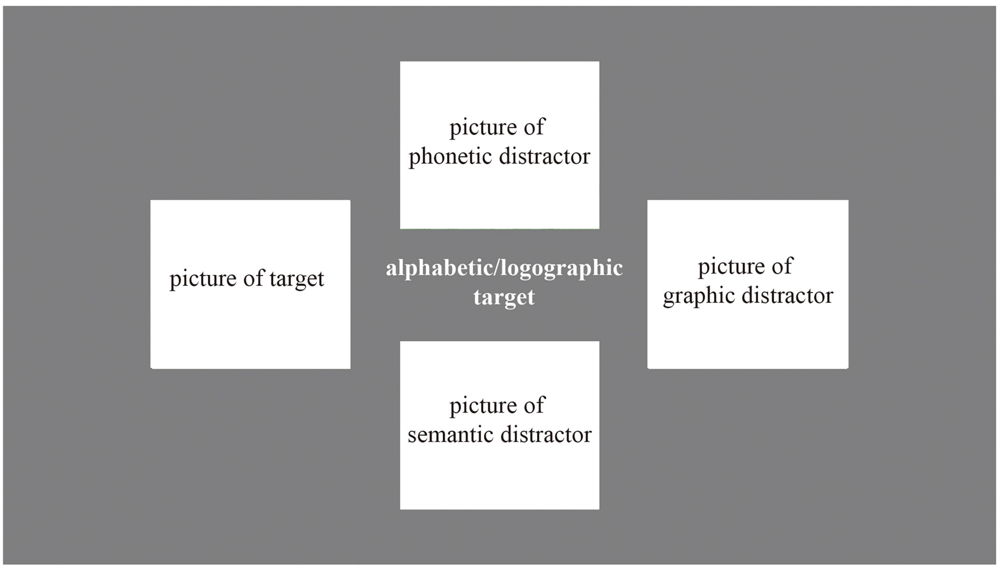


FIGURE 2
The target, semantic distractor, graphic distractor, and phonetic distractor for the alphabetic and logographic nonwords.

letters. For example, the nLD between the word "apple" (as the target) and "aple" is 1/5. This study calculated the production score using nLD.

In the alphabetic production test blocks, the score of every nonword was the average of 4 production tasks' scores. Every production score could be calculated by nLD (production score = 1 − nLD). For example, the nLD between the word "apple" and "epple" is 1/5, and its production score is 4/5. The nLD between the word "apple" and "abble" is 2/5, and its production score is 3/5. If a participant wrote "abbla," "abble," "appla," "apple," his or her score of the word "apple" should be 0.7.

The scoring of the logographic production test was performed similarly but adjusted to the features of logographic words. Logographic words were firstly separated into characters, and radicals of interweaved strokes formed each character. Thus, According to Barcroft (2002), the scoring method was adopted to consider partial correctness, i.e., the radicals. The partly reproduced words were given 0.25, 0.5, or 0.75 based on the radicals. The produced words were scored as 1. For instance, the Chinese character "婚" could be divided into two radicals, i.e., "女" and "昏." A two-character nonword "焻纱" could be considered to disassemble into four parts, i.e., "火," "昏," "纟," and "少." If the target word was wrongly written as "婚纱" in the first production task and correctly written as it should be in the remaining production tasks, the score would be 3/4 in the first production task and 1 in each remaining production task. The final score of this word production would be 0.94 (3.75/4).

Recognition data analysis (reaction time)

In the recognition test, both alphabetic and logographic blocks measured the accuracy of response and its Reaction Times (RTs) as the score of the recognition task. The RTs of every nonword were log-transformed. To balance the speed and accuracy, we only analyzed the correct-chosen nonwords. After calculating all the words' scores, the average score of cognates, false cognates, and non-cognates were calculated.

Results

All relevant data, as well as analysis scripts, are available on the OSF platform.[4]

Production blocks

Mixed model analyses were conducted on R software (version 4.1.2; R Core Team, 2021), using lmer functions from the lme4 package (Bates et al., 2015). The models included dummy-coded fixed effects of writing system as a within-subject variable (alphabetic vs. logographic, with alphabetic set as the reference), word type also as a within-subject variable (cognate, false cognate, non-cognate, with the non-cognate set as the reference), and the interaction between writing system and word type. The formula of the maximal model was lmer[score.pro ~ word type * writing system + (1 + word type + writing system| Subject) + (1| Item)]. The word type and writing system were not set as random slopes for item because an item was presented in one writing system and one word type. The word type and writing system were set as random slopes for subject because a subject responded to two writing systems and three types of words. The maximal model was fitted using the buildmer function in the buildmer package (Version 1.3; Voeten, 2019) in R, which uses the lmer function from the lme4 package (Bates et al., 2015). Using backwards stepwise elimination, the buildmer function starts from the most complex model and systematically simplifies the random structure until the model converges. This resulted in a random intercept for word and subject and a random slope for word type and writing system over subject. The fixed part consisted of the writing system (alphabetic vs. logographic), word type (non-cognate, cognate, false cognate), and their interaction. Model formula is: lmer[score.pro ~ word type * writing system + (1| Item) + (1 + word type + writing system| Subject)]. Model intercept reflects the score of the alphabetic non-cognates. The model for the production block is presented in Table 2 (fixed effects) and Table 3 (random effects). The outcome variable in the model is the score for each item. In general, the ANOVA shows the production task significantly differed between logographic and alphabetic nonwords ($F = 261.58$, $p < 0.001$), indicating that L1-L3 cognate status plays a different role from L2-L3 cognate status. A simple effect (see Table 4) is tested showing that cognates were learned significantly faster than non-cognate in both the alphabetic block (estimate: 0.40, $SE = 0.05$, $t = 7.80$, $p < 0.001$) and logographic block (estimate: 0.63, $SE = 0.05$, $t = 12.03$, $p < 0.001$). The major difference between learning alphabetic and logographic nonwords lies in L1-L3 and L2-L3 false cognates. As Table 4 shows, logographic false cognates showed difference from non-cognates (estimate: 0.34, $SE = 0.05$, $t = 6.69$, $p < 0.001$) while alphabetic did not differ (estimate: 0.08, $SE = 0.05$, $t = 1.53$, $p = 0.14$).

Figure 3 compares the score for the three word types in alphabetic and logographic writing systems. Taken together, the results in production tasks display the significant difference between the two writing systems. Therefore, the results in production task indicate that participants' bilingual experience facilitated the production of well-matched novel words, i.e., the L1 and L2 cognates. Furthermore, in learning the form-meaning mismatched novel words, interference can be overcome by the facilitation of the participants' dominant language, i.e., L1 Chinese; thus, the L1-L3 false cognates were learned faster than the non-cognates. However, the facilitation and interference from

4 https://osf.io/kzsd5/

TABLE 2 Fixed effects from linear mixed model of score with writing systems and word type as fixed effects in the production tasks.

Effect	Estimate	*SE*	*df*	*t*	*p*
(Intercept)	0.49	0.04	74.10	13.24	**<0.001*****
Logographic vs. alphabetic	-0.30	0.04	54.99	−6.72	**<0.001*****
False cognate vs. non-cognate	0.08	0.05	44.26	1.52	0.13
Cognate vs. non-cognate	0.40	0.05	47.75	7.78	**<0.001*****
Logographic vs. alphabetic * false cognate vs. non-cognate	0.26	0.07	42.05	3.74	**<0.001*****
Logographic vs. alphabetic * cognate vs. non-cognate	0.22	0.07	42.06	3.17	**0.002****

**$p < 0.01$ and ***$p < 0.001$.

TABLE 3 Random effects from linear mixed model of score with item and subject as random effects in the production tasks.

Groups	Name	SD	Variance	ICC
Item	*(Intercept)*	0.10	0.01	0.00
Subject	*(Intercept)*	0.15	0.02	0.31
	Writing system (logographic)	0.11	0.01	
	Word type (false cognate)	0.06	0.00	
	Word type (cognate)	0.09	0.01	
Residual		0.18	0.03	

L2 English were quite balanced in learning L2-L3 false cognates, leaving no significant advantage in the production task.

Recognition blocks

The formula of the maximum model was lmer[RTs ~ word type * writing system + (1 + word type + writing system| Subject) + (1| Item)]. The model selection process of recognition task was similar to the production task. It turned out the best model was lmer[RTs ~ word type * writing system + (1| Item) + (1 + writing system + word type| Subject)]. Model intercept reflects the RTs (log-transformed) of the alphabetic non-cognates. The RTs for the recognition blocks are tested and log-transformed before analyses to reduce skew in the distribution. The model is presented in Table 5 (fixed effects) and Table 6 (random effects). Overall, the pattern of the results is similar to that of the production task. A further *post–hoc* comparison between cognates and false cognates (estimate: 0.24, $SE = 0.05$, $t = 5.12$, $p < 0.001$) indicates significant differences. A simple effect (Table 7) is tested suggesting that there was no difference between non-cognate and false cognate in alphabetic writing system (estimate: 0.00, $SE = 0.06$, $t = 0.05$, $p = 0.96$) but the significant difference was detected in logographic writing system (estimate: −0.16, $SE = 0.06$, $t = -2.79$, $p = 0.01$). Similar results were also found in the production tasks. As Table 7 shows, cognates were learned faster than non-cognates in both alphabetic (estimate: -0.30, $SE = 0.06$, $t = -5.17$, $p < 0.001$) and logographic writing systems (estimate: −0.34, $SE = 0.06$, $t = -5.84$, $p < 0.001$).

Figure 4 compares the RTs for the three word types in alphabetic and logographic writing systems. Taken together, the results in recognition tasks display less difference between the two writing systems in learning cognates. Both L1 and L2 facilitated the well-matched novel learning. However, the results of the false cognates are similar to those of the production task, revealing that there was more facilitation than interference from L1, while the facilitation and interference from L2 were quite balanced.

Discussion

In this study, we examined the facilitative and interferent effects of bilingual experience on novel word learning. To this end, nonwords were created either in the logographic form similar to L1 Chinese or in the alphabetic form like L2 English, which can be further divided into cognates, false cognates, and non-cognates in either L1-L3 or L2-L3 cognate status. An L1-L3 cognate has a slightly different logographic form and completely identical meaning to the original L1 Chinese word. In the same vein, an L2-L3 cognate almost coincides with the alphabetic L2 English words in form and meaning. An L1-L3 false cognate has the logographic form that coincides with the already learned L1 but has a different meaning. Similarly, an L2-L3 false cognate shares the alphabetic form of the original L2 word but not the meaning. The learning outcomes of cognates and false cognates were compared to their non-cognates individually in the logographic block and alphabetic block. Therefore, the learning outcome of the L1-L3 and L2-L3 cognates can be individually traced from either L1 Chinese or L2 English. The learning outcomes were made up of the participants' production scores and recognition RTs. The production task is designed to test the precise knowledge of form and form-meaning mapping. The recognition task is designed to mainly test the knowledge of form-meaning mapping and a small amount of form knowledge. Such an experimental design functioned to examine the direct effects of bilingual experience on novel word learning, i.e., the transferable knowledge and representations from the previously learned word forms and concepts as well as the skills of learning them (Hirosh and Degani, 2018).

The results show that the direct facilitation effects from L1 and L2 can be separately traced from the logographic and alphabetic

TABLE 4 Simple effects in the production tasks.

Writing system	Contrast	Estimate	*SE*	*df*	*t*	*p*
Alphabetic	False cognate vs. non-cognate	0.08	0.05	22.30	1.53	0.14
	Cognate vs. non-cognate	0.40	0.05	24.20	7.80	**<0.001*****
Logographic	False cognate vs. non-cognate	0.34	0.05	22.30	6.69	**<0.001*****
	Cognate vs. non-cognate	0.63	0.05	24.20	12.03	**<0.001*****

***$p < 0.001$.

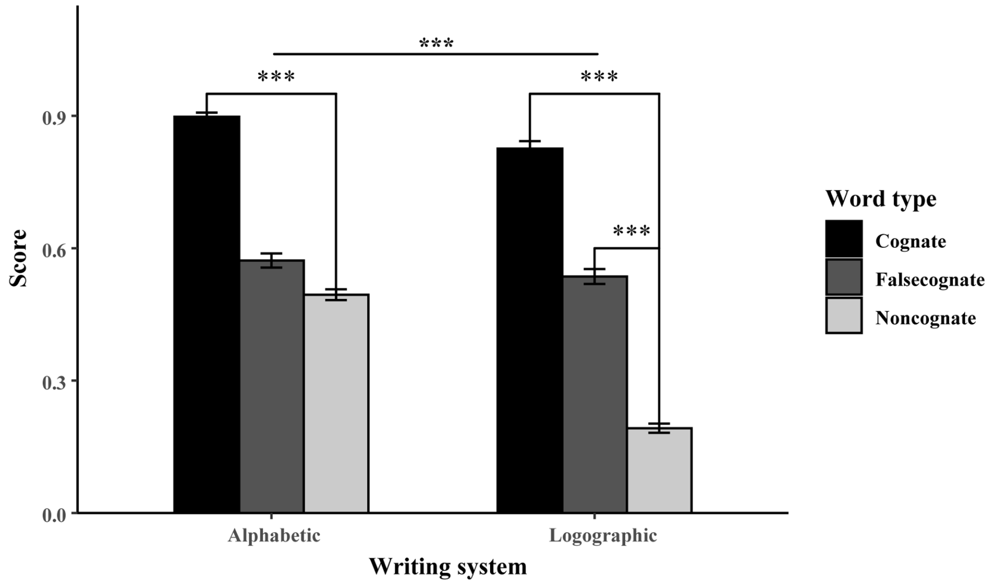


FIGURE 3
The alphabetic and logographic score for the cognate, false cognate and non-cognate in the production block.

novel word learning. Both the L1-L3 and the L2-L3 cognates were learned faster than their non-cognate counterparts in the production and recognition tasks. The results indicate that learners will automatically search for, detect, and use similar information between the languages known and those to be learned. In another word, the bilingual experience can exert its facilitation effect in a dissociable way with the logographic and alphabetic novel words. However, the sophistication of the facilitation and interference from bilingual experience can be more subtly revealed by the learning outcomes of false cognates. In recognition, the learning of false cognates is mostly influenced by the mismatched form-meaning mapping; in production, it's affected by both the form overlap and the mismatched form-meaning mapping (Janke and Kolokante, 2015; Marecka et al., 2021). The current study found no significant difference between L2-L3 false cognates and L2-L3 non-cognates in both the recognition and production tasks. The interference from L2 mismatched form-meaning mapping was possibly offset by the facilitation of the form overlap. Quite differently, a stronger facilitation effect was observed from L1 Chinese, leading to the result that the L1-L3 false cognates were learned significantly better than the non-cognates in both the recognition and the production tasks. Therefore, the current study contributes to providing new evidence to the facilitation and interference mechanism of how bilingual experience affects novel word learning when L1 and L2 word knowledge is not explicitly activated as translations (Cenoz et al., 2021; Hirosh and Degani, 2021).

The facilitation effect of bilingual experience in learning cognates

The current study provides new evidence of bilingual experience with orthographically different languages. Consistent with previous studies, a learning advantage for cognates is found both in laboratory settings and classroom settings (e.g., Bartolotti and Marian, 2019; Zhang et al., 2019; Otwinowska et al., 2020; Vanlangendonck et al., 2020). Regardless of the different writing systems of the bilingual experience, cognates were the quickest to be recognized and produced. Such facilitative effects were not moderated by the difference in writing systems as some studies reported previously (Muscalu and Smiley, 2018; Iniesta et al., 2021). Particularly, this research did not use identical cognates. Instead, the target cognates slightly differ from the participants'

TABLE 5 Fixed effects from linear mixed model of RTs with writing systems and word type as fixed effects in the recognition tasks.

Effect	Estimate	*SE*	*df*	*t*	*p*
(Intercept)	0.50	0.04	69.46	12.38	**<0.001*****
Logographic vs. alphabetic	0.11	0.05	51.93	2.25	**0.03***
False cognate vs. non-cognate	0.00	0.06	43.14	0.05	0.96
Cognate vs. non-cognate	−0.30	0.06	45.44	−5.17	**<0.001*****
Logographic vs. alphabetic * false cognate vs. non-cognate	−0.16	0.08	42.00	−2.02	**0.05***
Logographic vs. alphabetic * cognate vs. non-cognate	−0.04	0.08	42.00	−0.49	0.63

*$p < 0.05$ and ***$p < 0.001$.

TABLE 6 Random effects from linear mixed model of RTs with item and subject as random effects in the recognition tasks.

Groups	Name	*SD*	Variance	*ICC*
Item	*(Intercept)*	0.11	0.01	0.25
Subject	*(Intercept)*	0.14	0.02	0.35
	Writing system (logographic)	0.11	0.01	
	Word type (false cognate)	0.05	0.00	
	Word type (cognate)	0.08	0.01	
Residual		0.19	0.04	

L1 Chinese or L2 English to guarantee the learning process. In this way, our data add new evidence to both the dissociable facilitative role of the L1 and the L2 in recognition and production.

Nevertheless, different from some studies (Muscalu and Smiley, 2018; Iniesta et al., 2021), the cognate interference is not found either with the L1-L3 or L2-L3 cognates in this study. Both L1 and L2 experience has been found to facilitate novel word learning. A possible reason is that the new words were taught and tested on the same day instead of a prolonged period. There was no sleep time for the participants, during which the lexical consolidation and competition would happen (Lindsay and Gaskell, 2013). The current study can be regarded as further evidence of the short-term facilitation advantage of bilingual experience (Marecka et al., 2021). The more similarities shared by the word to-be-learned and the words learned, the easier it can be learned. In the interview immediately after the experiment, all the participants reported that they had tried involuntarily to refer to the L1 or L2 original word of the cognates, especially during the learning phase. Additionally, participants reported more analytic strategies in learning L1-L3 cognates rather than the L2-L3 cognates. Therefore, learning cognates benefits from form overlap and form meaning overlap of the previously learned words. Learners were able to utilize the overlap in form and form-meaning mapping. The bilingual experience facilitates learning both L1-L3 cognates and L2-L3 cognates. In this way, L3 word learning may not be parasitic in a certain language. But rather, it is a process of building new lexical knots with language experience, even though the bilinguals acquire their L2 mainly in classroom contexts (Hirosh and Degani, 2018).

The facilitation-and-interference effect of bilingual experience In learning false cognates

Learning false cognates were expected to entail competing processes in the direct effects of bilingual experience (Fang and Perfetti, 2017; Elias and Degani, 2022). Learning false cognates may benefit from the form overlap, but also need to overcome the meaning interference of words in the acquired language. Since L1 and L2 However, the role of bilingual experience could have been mixed with cross-linguistic similarity and language complexity. To disentangle the confusion of cross-linguistic similarity, this study has researched into the bilingual experience of orthographic difference, i.e., the logographic and alphabetic words. Through such an approach, the direct effects from L1 and L2 can be individually traced. Moreover, the confusing influence of the complexity of the logographic and alphabetic writing systems has been excluded by using the logographic and alphabetic non-cognates as baselines, respectively. Therefore, the facilitation and interference effects have been examined with the same writing system in a within-subject way. The current results of learning false cognates reveal quite different direct effects of L1 and L2 experience on novel word learning. As for the L1-L3 false cognates, the facilitation from L1 form overlapping overcomes the interference from L1 form-meaning mismatch. In both the recognition and production tasks, L1-L3 false cognates were learned significantly better than the non-cognates. However, concerning the L2-L3 false cognates, their learning outcome is almost the same as the non-cognates. Thus, the facilitation and interference from L2 are close to an equal balance. Taken together, there seems to be stronger facilitation from L1 experience rather than L2 experience when bilinguals are learning the mismatched novel words, i.e., the false cognates. The L1 facilitation outperforms its interference with a possibly better and more accurate inhibition instead of higher inhibition of the logographic form-meaning mappings from the prior knowledge (Mulík and Carrasco-Ortiz, 2021).

These results are partially similar to the research of Marecka et al. (2021), in which the learning of false cognates benefits from the overlap in L1-L2 form and is not harmed by L1 interference. In the current research, we found that novel word learning may have weaker facilitation from L2 form overlapping or stronger interference from L2 form-meaning mismatch. When comparing the recognition and production tasks of learning L2-L3 cognates, the production task shows a slightly better learning outcome. Such an advantage in production task over recognition task suggests the potential of a stronger interference from L2 form-meaning mismatch rather than a weaker facilitation from L2 form overlapping. In another word, the semantic discrepancy leads to more difficulty for L2 similar words. However, there is a significant facilitation from L1 experience in learning L1-L3 false cognates. In the production tasks, the learning of false cognates shows a very significant advantage over the learning of non-cognates, while in the production task, such an advantage just reaches the level of being significant. Therefore, the L1-L3 form-meaning mismatch also exerts an interferent effect on novel word learning, but it seems to be much weaker than the L1 facilitation. Similar L1 facilitation in learning false cognates has also been reported in learning novel words both as an L2 and an L3 (Hirosh and Degani, 2021; Marecka et al., 2021). In the study of Hirosh and Degani (2021), they found that learning false cognates through L1 translations was superior to learning them through L2 translations. Taken together, the direct effects from L1 experience seem to exert more facilitation than interference when learning the L1 form-meaning mismatch, while learning the L2 form-meaning mismatch seems to suffer more from its interference effect. Notably, our study employed paired–associate learning with pictures. In both the learning and testing phases, the facilitation and interference effects are not triggered by explicit translations. A possible reason for such learning outcomes may be that the L1 and L2 bilingual experience is quite different concerning their learning conditions, automaticity levels, etc. Therefore, the L2 form-meaning mapping is weaker than the L1 form-meaning mapping. In the interview, participants also reported that among the three distractors in the recognition task, they were rarely confused by the phonetic distractors, but they were mostly misled by the semantic distractors with the graphic distractors as the second most misleading ones. In sum, the current study suggests that L1 and L2 play quite different roles as the direct effects of bilingual experience. The better facilitation effects from L1 may derive from a better inhibition rather than higher inhibition. These findings add new evidence to the facilitation-and-interference mechanism of bilingual experience.

TABLE 7 Simple effects in the recognition tasks.

Writing system	Contrast	Estimate	*SE*	*df*	*t*	*p*
Alphabetic	False cognate vs. non-cognate	0.00	0.06	43.10	0.05	0.96
	Cognate vs. non-cognate	−0.30	0.06	45.40	−5.17	**<0.001*****
Logographic	False cognate vs. non-cognate	−0.16	0.06	43.30	−2.79	**0.01****
	Cognate vs. non-cognate	−0.34	0.06	45.60	−5.84	**<0.001*****

**$p < 0.01$ and ***$p < 0.001$.

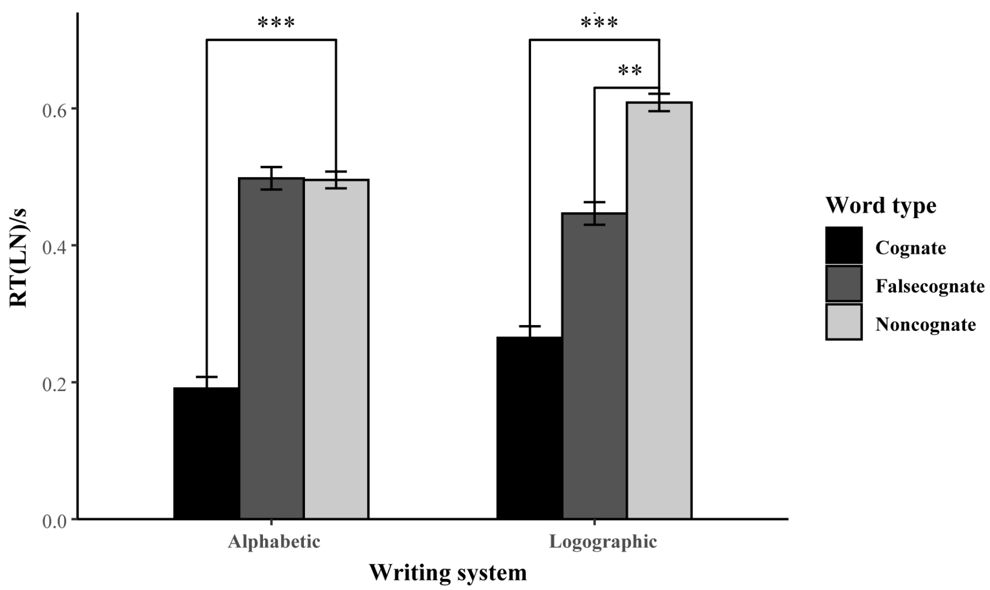


FIGURE 4
The alphabetic and logographic RTs for the cognate, false cognate and non-cognate in the recognition block.

Conclusion

Findings of the study

To the best of our knowledge, the present study is the first to systematically disentangle the influence of bilingual experience *via* examining L1-L3 and L2-L3 cognate status within the same bilingual participants. The L1 and L2 experience has been analyzed, respectively, through comparing the cognates and the false cognates with the baseline of non-cognates within the same writing system to avoid mingling with the different complexity of logographic and alphabetic words. Our results show that the dominant L1 and non-dominant L2 can exert dissociable direct effects as facilitation for learning the form-meaning closely matched novel words, i.e., the cognates. However, in our research, learning the form-meaning mismatch, i.e., the false cognates, reveals the sophistication of the facilitation-and-interference effects sourced from bilingual experience. The form-meaning mismatch potentially triggers interference from both L1 and L2. But the interference is compensated by the facilitation from L1 and L2 prior knowledge and the form-meaning mapping skills. It's worth noticing that the current study provides new evidence to the different subtlety of inhibition with a more accurate inhibition of L1 form-meaning mismatch and a less accurate inhibition of L2 form-meaning mismatch, thus resulting in the different degrees of facilitation-and-interference effect from bilingual experience. These findings carry potential educational implications in that learning novel words depends on substantial bilingual experience and requires a fuller understanding of the subtle difference in the facilitation and interference from L1 and L2. Such findings may provide some insights into foreign language teaching in different contexts (Cenoz et al., 2021; Chen et al., 2022).

Limitations and future study

Firstly, the study is limited to the learning process of paired–associates learning without teacher instruction. Further teaching experiment is needed to identify the cost and benefit of teaching logographic and alphabetic novel words through the dominant and non-dominant languages in different teaching contexts. In addition, according to Marecka et al. (2021), the nonwords are all concrete nouns due to the limitation of the meaning represented by pictures. Therefore, adding abstract nouns, verbs, adjectives, and other types of words would increase the ecological validity of the present findings. Thirdly, we created alphabetic and logographic nonwords based on English and Chinese. However, there are still more writing systems that deserve our further attention. A power analysis can be added to decide the number of participants to address more complicated language experience, such as trilingual experience.

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.

Ethics statement

The studies involving human participants were reviewed and approved by School of Foreign Languages, South China University of Technology. The patients/participants provided their written informed consent to participate in this study.

Author contributions

YC conceived, designed the study, managed and coordinated responsibility for the research activity planning and execution, provided financial support, and drafted the manuscript. HX prepared the stimuli, recruited the participants, implemented the experiment, and performed the statistical analysis. RD and WZ recruited the participants, implemented the experiment, and collected data. YC and HX interpreted the data, wrote the article, and approved the submitted version. All authors contributed to the article and approved the submitted version.

Funding

This work was partly supported by the Department of Education of Guangdong Province (CN; project number 2020GXJK375) and the National Planning Office of Philosophy and Social Science (project number 21BYY052).

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.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2022.1003199/full#supplementary-material

References

Allen, D. (2019). Cognate frequency and assessment of second language lexical knowledge. *Int. J. Biling.* 23, 1121–1136. doi: 10.1177/1367006918781063

Antoniou, M., Liang, E., Ettlinger, M., and Wong, P. C. M. (2015). The bilingual advantage in phonetic learning. *Biling. Lang. Congn.* 18, 683–695. doi: 10.1017/S1366728914000777

Arana, S. L., Oliveira, H. M., Fernandes, A. I., Soares, A. P., and Comesaña, M. (2022). The cognate facilitation effect depends on the presence of identical cognates. *Biling. Lang. Congn.* 25, 660–678. doi: 10.1017/S1366728922000062

Barcroft, J. (2002). Semantic and structural elaboration in L2 lexical acquisition. *Lang. Learn.* 52, 323–363. doi: 10.1111/0023-8333.00186

Bartolotti, J., and Marian, V. (2017). Bilinguals' existing languages benefit vocabulary learning in a third language: bilingual third language learning. *Lang. Learn.* 67, 110–140. doi: 10.1111/lang.12200

Bartolotti, J., and Marian, V. (2019). Learning and processing of orthography-to-phonology mappings in a third language. *Int. J. Multiling.* 16, 377–397. doi: 10.1080/14790718.2017.1423073

Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

Bogulski, C. A., Bice, K., and Kroll, J. F. (2019). Bilingualism as a desirable difficulty: advantages in word learning depend on regulation of the dominant language. *Biling. Lang. Congn.* 22, 1052–1067. doi: 10.1017/S1366728918000858

Cenoz, J., Leonet, O., and Gorter, D. (2021). Developing cognate awareness through pedagogical translanguaging. *Int. J. Biling. Educ. Biling.* 25, 2759–2773. doi: 10.1080/13670050.2021.1961675

Chen, Y., Zhang, P., and Huang, L. (2022). Translanguaging/trans-semiotizing in teacher-learner interactions on social media: making learner agency visible and achievable. *System* 104:102686. doi: 10.1016/j.system.2021.102686

De Bot, K., and Jaensch, C. (2015). What is special about L3 processing? *Biling. Lang. Congn.* 18, 130–144. doi: 10.1017/S1366728913000448

Del Pilar Agustín-Llach, M. (2019). The impact of bilingualism on the acquisition of an additional language: evidence from lexical knowledge, lexical fluency, and (lexical) cross-linguistic influence. *Int. J. Biling.* 23, 888–900. doi: 10.1177/1367006917728818

Dijkstra, T., Miwa, K., Brummelhuis, B., Sappelli, M., and Baayen, H. (2010). How cross-language similarity and task demands affect cognate recognition. *J. Mem. Lang.* 62, 284–301. doi: 10.1016/j.jml.2009.12.003

Dong, J., Yang, S., and Wang, Q. (2015). N400-like effect of stroke-based neighborhood size in Chinese characters. *J. Neurolinguistics* 35, 120–134. doi: 10.1016/j.jneuroling.2015.04.003

Elias, M., and Degani, T. (2022). Cross-language interactions during novel word learning: the contribution of form similarity and participant characteristics. *Biling. Lang. Congn.* 25, 548–565. doi: 10.1017/S1366728921000857

Eng, N., Vonk, J. M., Salzberger, M., and Yoo, N. (2019). A cross-linguistic comparison of category and letter fluency: mandarin and English. *Q. J. Exp. Psychol.* 72, 651–660. doi: 10.1177/1747021818765997

Fang, X., and Perfetti, C. A. (2017). Perturbation of old knowledge precedes integration of new knowledge. *Neuropsychologia* 99, 270–278. doi: 10.1016/j.neuropsychologia.2017.03.015

García, G. E., Sacco, L. J., and Guerrero-Arias, B. E. (2020). Cognate instruction and bilingual students' improved literacy performance. *Read. Teach.* 73, 617–625. doi: 10.1002/trtr.1884

Gertken, L. M., Amengual, M., and Birdsong, D. (2014). "Assessing language dominance with the bilingual language profile," in *Measuring L2 proficiency: perspectives from SLA*. eds. P. Leclercq, A. Edmonds and H. Hilton (Bristol: Multilingual Matters), 208–225.

Green, D. (1986). Control, activation, and resource: a framework and a model for the control of speech in bilinguals. *Brain Lang.* 27, 210–223. doi: 10.1016/0093-934X(86)90016-7

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Biling. Lang. Congn.* 1, 67–81. doi: 10.1017/S1366728998000133

Hirosh, Z., and Degani, T. (2018). Direct and indirect effects of multilingualism on novel language learning: an integrative review. *Psychon. Bull. Rev.* 25, 892–916. doi: 10.3758/s13423-017-1315-7

Hirosh, Z., and Degani, T. (2021). Novel word learning among bilinguals can be better through the (dominant) first language than through the second language. *Lang. Learn.* 71, 1044–1084. doi: 10.1111/lang.12457

Hsieh, M.-C., Jeong, H., Dos Santos Kawata, K. H., Sasaki, Y., Lee, H.-C., Yokoyama, S., et al. (2017). Neural correlates of bilingual language control during interlingual homograph processing in a logogram writing system. *Brain Lang.* 174, 72–85. doi: 10.1016/j.bandl.2017.06.006

Iniesta, A., Rossi, E., Bajo, M. T., and Paolieri, D. (2021). The influence of cross-linguistic similarity and language background on writing to dictation. *Front. Psychol.* 12:679956. doi: 10.3389/fpsyg.2021.679956

Janke, V., and Kolokante, M. (2015). False cognates: the effect of mismatch in morphological complexity on a backward lexical translation task. *Second. Lang. Res.* 31, 137–156. doi: 10.1177/0267658314545836

Jiang, N. (2021). Orthographic friends and lexical strangers in the L2 lexicon. *J. Sec. Lang. Stud.* 4, 224–244. doi: 10.1075/jsls.21012.jia

Kaushanskaya, M., Yoo, J., and Van Hecke, S. (2013). Word Learning in Adults With Second-Language Experience: Effects of Phonological and Referent Familiarity. *J. Speech, Lang. Hear. Res.* 56, 667–678. doi: 10.1044/1092-4388(2012/11-0084

Koda, K., and Miller, R. (2018). "Cross-linguistic interaction in L2 word meaning inference in English as a foreign language" in *Writing systems, reading processes, and cross-linguistic influences: Reflections from the Chinese, Japanese and Korean languages*. ed. H. Pae (Amsterdam: John Benjamins), 293–312.

Kroll, J. F., Bobb, S. C., and Hoshino, N. (2014). Two languages in mind: bilingualism as a tool to investigate language, cognition, and the brain. *Curr. Dir. Psychol. Sci.* 23, 159–163. doi: 10.1177/0963721414528511

Lemhöfer, K., and Broersma, M. (2012). Introducing LexTALE: a quick and valid lexical test for advanced learners of English. *Behav. Res. Methods* 44, 325–343. doi: 10.3758/s13428-011-0146-0

Lemhöfer, K., Dijkstra, T., and Michel, M. (2004). Three languages, one ECHO: cognate effects in trilingual word recognition. *Lang. Cogn. Process.* 19, 585–611. doi: 10.1080/01690960444000007

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Phy. Doklady* 10, 707–710.

Li, C., and Gollan, T. H. (2021). What cognates reveal about default language selection in bilingual sentence production. *J. Mem. Lang.* 118:104214. doi: 10.1016/j.jml.2020.104214

Lindsay, S., and Gaskell, M. G. (2013). Lexical integration of novel words without sleep. *J. Exp. Psychol. Learn. Mem. Cogn.* 39, 608–622. doi: 10.1037/a0029243

Liu, S., Liu, D., Pan, Z., and Xu, Z. (2018). The association between reading abilities and visual-spatial attention in Hong Kong Chinese children. *Dyslexia* 24, 263–275. doi: 10.1002/dys.1584

Liu, H., Zhang, Y., Blanco-Elorrieta, E., He, Y., and Chen, B. (2020). The role of proactive control on subcomponents of language control: evidence from trilinguals. *Cognition* 194:104055. doi: 10.1016/j.cognition.2019.104055

Marecka, M., Szewczyk, J., Otwinowska, A., Durlik, J., Foryś-Nogala, M., Kutyłowska, K., et al. (2021). False friends or real friends? False cognates show advantage in word form learning. *Cognition* 206:104477. doi: 10.1016/j.cognition.2020.104477

Marian, V., Bartolotti, J., Chabal, S., and Shook, A. (2012). CLEARPOND: cross-linguistic easy-access resource for phonological and orthographic neighborhood densities. *PLoS One* 7:e43230. doi: 10.1371/journal.pone.0043230

Mok, P. P. K., Lee, A., Li, J. J., and Xu, R. B. (2018). Orthographic effects on the perception and production of L2 mandarin tones. *Speech Comm.* 101, 1–10. doi: 10.1016/j.specom.2018.05.002

Mulík, S., and Carrasco-Ortiz, H. (2021). Different effects of L1 and L2 phonology on L3 lexical learning: an ERP study. *Second. Lang. Res.* (in press). doi: 10.1177/02676583211054448

Muscalu, L. M., and Smiley, P. A. (2018). The illusory benefit of cognates: lexical facilitation followed by sublexical interference in a word typing task. *Biling. Lang. Congn.* 22, 848–865. doi: 10.1017/S1366728918000792

Muylle, M., Assche, E. V., and Hartsuiker, R. J. (2022). Comparing the cognate effect in spoken and written second language word production. *Biling. Lang. Congn.* 25, 93–107. doi: 10.1017/S1366728921000444

Nation, I. S. P. (2013). *Teaching & learning vocabulary*. Boston: Heinle Cengage Learning.

Nation, P. (2020). "The different aspects of vocabulary knowledge" in *The Routledge handbook of vocabulary studies*. ed. S. Webb (London: Routledge), 15–29.

Otwinowska, A., Foryś-Nogala, M., Kobosko, W., and Szewczyk, J. (2020). Learning orthographic cognates and non-cognates in the classroom: does awareness of cross-linguistic similarity matter? *Lang. Learn.* 70, 685–731. doi: 10.1111/lang.12390

Otwinowska, A., and Szewczyk, J. M. (2019). The more similar the better? Factors in learning cognates, false cognates and non-cognate words. *Int. J. Biling. Educ. Biling.* 22, 974–991. doi: 10.1080/13670050.2017.1325834

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., et al. (2019). PsychoPy2: experiments in behavior made easy. *Behav. Res. Methods* 51, 195–203. doi: 10.3758/s13428-018-01193-y

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: https://www.R-project.org/ (Accessed October 9, 2022).

Rastle, K., Harrington, J., and Coltheart, M. (2002). 358,534 nonwords: the ARC nonword database. *Q. J. Exp. Psychol. Sec. A* 55, 1339–1362. doi: 10.1080/02724980244000099

Richlan, F. (2020). The functional neuroanatomy of developmental dyslexia across languages and writing systems. *Front. Psychol.* 11:155. doi: 10.3389/fpsyg.2020.00155

Robinson Anthony, J. J. D., Blumenfeld, H. K., Potapova, I., and Pruitt-Lord, S. L. (2022). Language dominance predicts cognate effects and metalinguistic awareness in preschool bilinguals. *Int. J. Biling. Educ. Biling.* 25, 922–941. doi: 10.1080/13670050.2020.1735990

Ruan, Y., Georgiou, G. K., Song, S., Li, Y., and Shu, H. (2017). Does writing system influence the associations between phonological awareness, morphological awareness, and reading? A meta-analysis. *J. Edu. Psychol.* 110, 180–202. doi: 10.1037/edu0000216

Schmitt, N. (2019). Understanding vocabulary acquisition, instruction, and assessment: a research agenda. *Lang. Teach.* 52, 261–274. doi: 10.1017/S0261444819000053

Simpson Baird, A., Palacios, N., and Kibler, A. (2016). The cognate and false cognate knowledge of young emergent bilinguals: cognate knowledge of young bilinguals. *Lang. Learn.* 66, 448–470. doi: 10.1111/lang.12160

Subramaniapillai, S., Rajah, M. N., Pasvanis, S., and Titone, D. (2019). Bilingual experience and executive control over the adult lifespan: the role of biological sex. *Biling. Lang. Congn.* 22, 733–751. doi: 10.1017/S1366728918000317

Van Hell, J. G., and De Groot, A. M. B. (1998). Conceptual representation in bilingual memory: effects of concreteness and cognate status in word association. *Biling. Lang. Congn.* 1, 193–211. doi: 10.1017/S1366728998000352

Van Hell, J. G., and de Groot, A. M. B. (2008). Sentence context modulates visual word recognition and translation in bilinguals. *Acta Psychol.* 128, 431–451. doi: 10.1016/j.actpsy.2008.03.010

Vanlangendonck, F., Peeters, D., Rueschemeyer, S.-A., and Dijkstra, T. (2020). Mixing the stimulus list in bilingual lexical decision turns cognate facilitation effects into mirrored inhibition effects. *Biling. Lang. Congn.* 23, 836–844. doi: 10.1017/S1366728919000531

Voeten, C. C. (2019). Buildmer: stepwise elimination and term reordering for mixed-effects regression [computer software]. R package version, 1. Available at: https://CRAN.R-project.org/package=buildmer (Accessed October 9, 2022).

Wang, R., Huang, S., Zhou, Y., and Cai, Z. G. (2020). Chinese character handwriting: a large-scale behavioral study and a database. *Behav. Res. Methods* 52, 82–96. doi: 10.3758/s13428-019-01206-4

Yum, Y. N., Law, S.-P., Su, I.-F., Lau, K.-Y. D., and Mo, K. N. (2014). An ERP study of effects of regularity and consistency in delayed naming and lexicality judgment in a logographic writing system. *Front. Psychol.* 5:315. doi: 10.3389/fpsyg.2014.00315

Zhang, H., Han, Y., Cheng, X., Sun, J., and Ohara, S. (2021). Unpacking cross-linguistic similarities and differences in third language Japanese vocabulary acquisition among Chinese college students. *J. Multiling. Multicult. Dev.* 1–13, 1–13. doi: 10.1080/01434632.2020.1865987

Zhang, J., Wu, C., Zhou, T., and Meng, Y. (2019). Cognate facilitation priming effect is modulated by writing system: evidence from Chinese-English bilinguals. *Int. J. Biling.* 23, 553–566. doi: 10.1177/1367006917749062

frontiers | Frontiers in Psychology

TYPE Systematic Review
PUBLISHED 23 December 2022
DOI 10.3389/fpsyg.2022.1057501

OPEN ACCESS

EDITED BY
Kenneth R. Paap,
San Francisco State University,
United States

REVIEWED BY
Greg Poarch,
University of Groningen, Netherlands
Peter Bright,
Anglia Ruskin University,
United Kingdom

*CORRESPONDENCE
Franziska Köder
✉ f.m.koder@iln.uio.no

SPECIALTY SECTION
This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Psychology

RECEIVED 29 September 2022
ACCEPTED 02 December 2022
PUBLISHED 23 December 2022

CITATION
Köder F, Sharma C, Cameron S and Garraffa M (2022) The effects of bilingualism on cognition and behaviour in individuals with attention deficits: A scoping review.
Front. Psychol. 13:1057501.
doi: 10.3389/fpsyg.2022.1057501

The effects of bilingualism on cognition and behaviour in individuals with attention deficits: A scoping review

Franziska Köder[1]*, Curtis Sharma[1,2], Sarah Cameron[1] and Maria Garraffa[1,3]

[1]Center for Multilingualism in Society Across the Lifespan, Department of Linguistics and Scandinavian Studies, University of Oslo, Oslo, Norway, [2]University of Cambridge, Cambridge, United Kingdom, [3]Faculty of Medicine and Health Sciences, East Anglia University, Norwich, United Kingdom

Background: Weaknesses in executive function have persistently been found to be associated with Attention Deficit Hyperactivity Disorder (ADHD), while bilinguals have been argued to show advantages in executive functions. While there has been some research into how bilingualism affects cognitive skills and behaviour in individuals with attention deficits, the question is still very much open. The aim of this systematic review is to gather, synthesise and evaluate existing evidence on how bilingual language experience and attention deficits affect executive function performance and ADHD-related symptoms in children and adults.

Methods: Following PRISMA guidelines, a comprehensive literature search in relevant databases (PsycInfo, PubMed, Scopus, ERIC, Web of Science, EMBASE, MEDLINE, LLBA) was performed using search strings related to attention difficulties/ADHD and bilingualism. All quantitative studies were included that presented original empirical data on the combined effects of bilingualism and attention levels, regardless of age group and methodology. The screening procedure revealed nine relevant studies.

Results: Across the nine identified studies, a total of 2071 participants were tested. Of these, seven studies involved children and two adults. The studies varied considerably with respect to their design and methodology, the targeted executive function skills or behavioural symptoms, as well as their measure of bilingualism and attention levels. Most studies assessed aspects of executive function performance such as interference control, response inhibition, working memory or cognitive flexibility. Three studies looked at the effects of bilingualism on ADHD-related symptoms or ADHD diagnosis. Across the studies, no systematic advantage or disadvantage of bilingualism on cognitive performance or behaviour in people with attention deficits was observed.

Conclusion: The limited number of identified studies provide no consistent evidence that bilingualism alleviates or intensifies attention difficulties in adults or children with ADHD. Based on the current state of research, individuals with ADHD and their families should not be concerned that learning additional languages has a negative impact on functioning or cognitive performance.

Systematic review registration: https://doi.org/10.17605/OSF.IO/PK768.

KEYWORDS

bilingualism, multilingualism, ADHD, attention, executive function

1. Introduction

It is estimated that more people in the world are bilingual than monolingual (Grosjean, 2010). At the same time studies have shown that people from minority backgrounds, who are often speakers of several languages, lag behind those from non-minority backgrounds when it comes to prevalence and treatment of Attention Deficit Hyperactivity Disorder (ADHD) (e.g., Slobodin and Masalha, 2020). Given that ADHD is one of the most common neurodevelopmental disorders in childhood and also present in the adult population (Polanczyk et al., 2007; Faraone et al., 2021), understanding this disparity should be an urgent pursuit. There may be several factors contributing to this imbalance. This includes proficiency in the majority language by reporters or caregivers, cultural expectations of development, or knowledge about ADHD (Stevens et al., 2004; Rothe, 2005; Eiraldi et al., 2006). Another reason could be that people from minority backgrounds often have a migration background and are therefore more likely to grow up with multiple languages. As suggested in the field of bilingualism research (see Antoniou, 2019), being bilingual could potentially improve cognitive abilities related to attention. Looking at the small but scientifically interesting group of bilingual speakers with attention deficits could provide new insights into the existence and extent of a so-called bilingual advantage.

In their systematic review on multilingualism and neurodevelopmental disorders, Uljarević et al. (2016) did not find any studies on the effect of multilingualism in people diagnosed with ADHD. However, since then several studies on the topic have appeared. The aim of this systematic review is to gather and synthesise existing evidence on the effects of bilingualism on the cognitive abilities and ADHD-related symptoms of people with attention deficits. The findings of this review could also be informative for practitioners and multilingual families who might be worried that exposing a child with ADHD to multiple languages might be detrimental to their cognitive development and functioning.

As background for the studies, we first briefly introduce the debate of advantages of bilingualism, discuss the potential association between ADHD and executive function deficits, and outline three possibilities of how bilingual language experience might affect cognition and behaviour in people with attention difficulties.

1.1. Bilingualism and advantages in executive function

In this review, we use the term "bilingual" to refer to anyone whose language experience includes two or more languages, and also to include multilinguals throughout the review, unless a specific need to distinguish between bilinguals and other multilinguals arises. We use the term "bilingualism" in a wide sense to cover both the early acquisition of two or more languages as well as second languages acquired later in life, but actively used outside of the classroom.

There is now a large body of literature claiming to have found advantages in cognitive skills for bilinguals, namely a group of skills under the umbrella term *executive function* (EF). The three principal executive function skills investigated in the literature are *inhibition*—the ability to inhibit prepotent responses (*response inhibition*) or task-irrelevant information (*interference control*); *cognitive flexibility*—the ability to switch attention between cognitive tasks; and *working memory* – the ability to store, monitor, manipulate, and update information relevant to an initiated or ongoing cognitive task (cf. Miyake et al., 2000). Evidence for an advantage in inhibition has been found for both children (e.g., Bialystok and Martin, 2004; Martin-Rhee and Bialystok, 2008; Poarch and van Hell, 2012) and adults (e.g., Bialystok et al., 2005; Salvatierra and Rosselli, 2011). Similarly, advantages in cognitive flexibility have been reported for children (e.g., Bialystok and Martin, 2004; Carlson and Meltzoff, 2008) and adults (e.g., Bialystok et al., 2004; Marzecová et al., 2013). Finally, there is some evidence that bilingual language experience might improve working memory capacity in children (Morales et al., 2013), but most studies report little or no effect for children and adults (e.g., Namazi and Thordardottir, 2010; Ratiu and Azuma, 2015; Yang, 2017). These cognitive advantages for bilinguals are said to emerge due to the bilingual's need to constantly monitor and manage both of their languages, as the languages not currently in use cannot be "switched off" (e.g., Spivey and Marian, 1999; Colomé, 2001; Starreveld et al., 2014; Bobb et al., 2020).

The so-called "bilingual advantage hypothesis" is controversial and hotly debated. Several studies have failed to replicate bilingual advantages in inhibition, cognitive flexibility, and working memory for both children and adults (e.g., Paap and Greenberg, 2013; Gathercole et al., 2014; Dick et al., 2019; Timmermeister et al., 2020). Furthermore, several meta-analyses report either no evidence of a bilingual advantage or small effect sizes that disappear when correcting for publication bias (de Bruin et al., 2015; Lehtonen et al., 2018; Paap, 2019; Lowe et al., 2021).

On the other hand, Grundy (2020) and others argue that while there are several reports of null findings, the studies that do find group differences far more often report bilinguals outperforming monolinguals than the other way around, even when controlling for factors related to publication bias and task differences. This could indicate that bilingualism might have small positive effects on cognitive performance, however only for certain groups of bilinguals, under certain conditions, and for certain tasks (Grundy, 2020; Ware et al., 2020). Recent studies point to the importance of making finer-grained distinctions assessing for instance bilinguals' language proficiency and usage, and their differential effects on performance in different executive function tasks (Poarch and Krott, 2019; Grundy, 2020). A higher level of bilingual proficiency seems to be significantly associated with better executive function performance (Pot et al., 2018; Thomas-Sunesson et al., 2018). Furthermore, the usage of different languages in different social contexts (Pot et al., 2018) and switching between two languages in the same environment might also lead to certain cognitive benefits (Hartanto and Yang, 2016). This systematic review aims to add to a more nuanced investigation of a "bilingual advantage" by focusing on bilingual speakers from neurodiverse backgrounds, namely people with attention deficits.

1.2. Executive function deficits in ADHD

Attention Deficit Hyperactivity Disorder (ADHD) is one of the most common neurodevelopmental disorders in childhood with estimated prevalence of around 5.29–7.1% of the 18 and under population (Willcutt, 2012; Polanczyk et al., 2014). Difficulties can continue into adulthood, where prevalence is estimated at 2.5% of the adult population (Roberts et al., 2015). ADHD is a clinical umbrella term for a set of behaviours, namely inattentiveness, hyperactivity, and impulsivity, which may or may not occur together (American Psychiatric Association, 2013).

Among several underlying cognitive impairments (cf. Sjöwall et al., 2013), ADHD is commonly linked to weaknesses in key executive function domains such as working memory and inhibitory control (Willcutt et al., 2005; Coghill et al., 2018). However, even though children with ADHD tend to perform below their peers in various executive function measures on a group level, more than 50% of individual children with ADHD do not exhibit executive function deficits (Nigg et al., 2005). Furthermore, effect sizes for group differences in performance on executive function were much smaller (0.46–0.69) than group differences on ADHD symptoms (2.5–4.0) (Willcutt et al., 2005). This refutes strong claims that executive function deficits are the primary cause of ADHD. Rather they seem to be an important component of the complex neuropsychology of ADHD, with potentially multiple pathways leading to similar behavioural symptoms (Sonuga-Barke, 2005).

1.3. Interactions between bilingualism and attention deficits

Considering that bilingualism and ADHD have potentially opposing effects on cognition, with bilingualism benefiting while ADHD hindering executive function performance, the question arises how these two factors might interact across the lifespan. In the following, we will sketch three possibilities on how bilingualism might affect executive functions and ADHD-related symptoms in children and adults with attention deficits.

First, it is possible that bilingual speakers with attention problems might experience a bilingual advantage, showing a better ability in executive functions and other cognitive domains, and exhibiting less severe symptoms linked to ADHD than their monolingual peers. This is based on the idea that bilinguals' constant need to selectively attend to one language (potentially suppressing their other language) trains executive function skills (cf. Bialystok, 2015). In other words, being a bilingual might improve overall executive function and offset (some) ADHD-related symptoms.

Opposed to that, being bilingual could be an additional burden for individuals with attention deficits, negatively affecting both executive function performance and inattention symptoms. This could be due to bilinguals needing to allocate parts of their already limited cognitive resources on inhibiting interference from their other language, making them slower and more error-prone in cognitive tasks. If this were the case, we would expect bilinguals with ADHD to show lower executive function abilities and more ADHD-related symptoms compared to their monolingual peers with ADHD.

Given that several studies have failed to replicate findings of a bilingual advantage (as noted in Section Bilingualism and advantages in executive function), as well as the small or null effects reported in several meta-analyses (Lehtonen et al., 2018; Paap, 2019), the hypothesis that bilinguals and monolinguals with attention deficits do not differ in cognitive or behavioural aspects is also a strong competitor. In this case, we would expect to find an association between ADHD and executive function deficits, but no effect or interaction with bilingual language proficiency or use.

2. Methods

To locate relevant studies on the joint effects of bilingualism and attention deficits, we performed a comprehensive search in June 2021 in databases connected to psychological, clinical, and linguistic research. These include PsycInfo, PubMed, Scopus, ERIC, Web of Science, EMBASE, MEDLINE, and Linguistics and Language Behaviour Abstracts (LLBA). As we expected relatively few relevant hits, the search was done with no restrictions regarding publication year or language. The search string consisted of several keywords relevant to attention difficulties and attentional abilities, and to bilingualism and multilingualism. Three different strings were tested and further expanded with new terms, resulting in the following search string:

[(ADHD or "attention deficit" or "Attention-Deficit" or "attention problem*" or "attention difficult*" or "attentional abilit*" or "attention abilit*") and (bilingual* or multilingual* or "dual language" or "second language" or "minority language" or "home language" or "heritage language")].

After removing duplicates, the remaining 779 titles were screened based on title and abstract. Screening was done independently by FK and SC, using the online systematic review tool Rayyan (Ouzzani et al., 2016) for efficient collaboration. In case of disagreement or uncertainty (title classified as "maybe" by one or both reviewers), a joint decision was reached during discussion. Given the scarcity of research on the combination of attention deficits and bilingualism, we included all empirical studies presenting original data on a combined effect of the two, regardless of age group and methodology. Exclusion criteria can be summarised in two main categories: (1) off topic, which includes any paper not directly dealing with the combination of bilingualism and attention deficits (i.e., studies on either bilingualism or attention deficits, but not both); and (2) ineligible study design, including case studies and methodological or theoretical papers with no empirical data. Category 1 was also used to exclude studies on second language acquisition in a classroom context, as these studies do not consider the active use of two or more languages in everyday life, and thus do not align with our definition of bilingualism.

Details of the number of exclusions for each criterion can be found in the PRISMA flowchart (Page et al., 2021) in Figure 1.

After a full-text screening of the remaining studies, nine were deemed eligible. The three exclusions in this step were a conference abstract and two papers on classroom L2 acquisition. Two additional papers were found by looking through the Google citations of the eligible studies, increasing the total of included studies to 12. Of those 12 studies, 10 were published in peer-reviewed journals, while two were found in an unpublished Ph.D. dissertation which we acquired by contacting the author.

Additional searches were performed in October 2021 and May 2022, in order to locate potential new studies that had been published since the first search. 36 (October) and 116 (May) new studies were identified. All but one were excluded (off topic: 147, ineligible study design: 4). The new inclusion was a journal publication of one of the studies from the formerly unpublished PhD retrieved in the first search, thus replacing this study rather than adding to the list of inclusions.

After extracting key information from the 12 included studies and discussing their relevance with respect to their topic, methodology, and quality, three studies were excluded. Özerk et al. (2011) was excluded due to both its low number of participants and the fact that its main focus was on methodology and assessment. Ramos et al. (2019) was excluded, as while it did include bilingual children who were at high risk for ADHD, the focus of this study was to compare monolingual and bilingual subjects' usage of syntax and semantics rather than their executive function abilities or ADHD-related behavioural symptoms. Finally, Askari et al. (2019) intervention study was excluded because of methodological and statistical concerns as well as due to inconsistencies in the results section, where the presented data did not seem to match the conclusions presented. Therefore, the final number of included studies was nine.

3. Results

3.1. Study characteristics

3.1.1. Participant information, variables, and analysis

Table 1 summarises general characteristics of the included studies. Across the nine studies included in the review a total of 2071 participants were tested. Of these, seven studies involved children (N = 1,823) ranging between 5 and 17 years of age with the exception of Goh et al. (2020), which tested younger children longitudinally from 2 to 4.5 years of age. Only two studies involved adults (N = 248); in both cases young adults typically in their twenties. All included studies were observational, examining the effects of independent variables on targeted variables. All studies included control variables, with age, gender, and socio-economic status (SES) being common, except for Mor et al. (2015), which did not include gender, and Hardy et al. (2021), which did not include SES. The way SES was operationalized varied between studies. Most studies used maternal education (Toppelberg et al., 2002; Bialystok et al., 2017; Sorge et al., 2017; Goh et al., 2020) or average number of parental years of schooling (Mor et al., 2015) as proxy for SES. Only two studies operationalized SES as a combination of affluence and parental education (Sharma, 2019; Sharma et al., 2022). The participating children in Chung-Fat-Yim et al. (2020)'s study were enrolled in a private school with high tuition fees, which they used as proxy for high SES. Six studies included verbal intelligence or language skills, and five studies included non-verbal intelligence. Several other control variables were measured, such as academic performance, ethnicity, and various immigration measures.

The included studies can be divided into two groups based on their outcome variables. Six studies measured cognitive

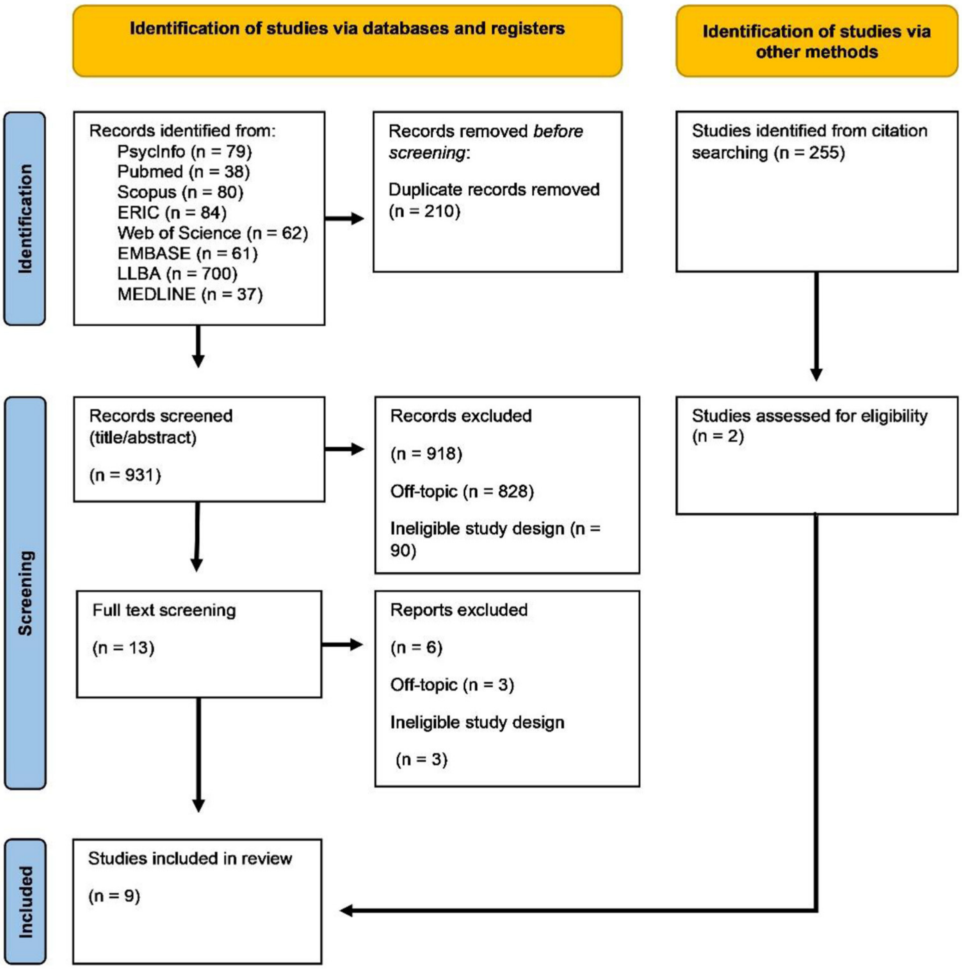


FIGURE 1
Combined results from first, second, and third search.

abilities related to executive function (Mor et al., 2015; Bialystok et al., 2017; Sorge et al., 2017; Sharma, 2019; Chung-Fat-Yim et al., 2020; Hardy et al., 2021), while four studies looked at levels of attention problems (Toppelberg et al., 2002), ADHD-related behavioural symptoms (Sharma, 2019; Sharma et al., 2022) or the odds of receiving an ADHD diagnosis (Goh et al., 2020). The findings for these two groups of studies will be presented separately below.

The studies utilised various methods of statistical analyses. The most common analysis method was (stepwise) linear regression models, inputting control variables sequentially, with language status (categorical) or language ability (continuous) entered last. Separate regressions were run for each dependent variable, whether that be levels of ADHD symptoms/attention difficulties, or executive function measures (Sorge et al., 2017; Sharma, 2019; Chung-Fat-Yim et al., 2020; Hardy et al., 2021; Sharma et al., 2022). None of these studies reported the inclusion of random effects such as random intercepts or random slopes into the models. Other analysis methods included bivariate correlations (Toppelberg et al., 2002), ANOVAs (Mor et al., 2015; Bialystok et al., 2017) and moderated models (Goh et al., 2020).

3.1.2. Bilingualism measures

The term "bilingual" is used across most studies, also to refer to participants who spoke or were exposed to more than two languages. Three studies (Mor et al., 2015; Bialystok et al., 2017; Hardy et al., 2021) analysed bilingualism as a categorical factor, i.e., participants were assigned to either a monolingual or a bilingual group. In Bialystok et al. (2017), participants were considered monolingual if they did not list a second language on the Language and Social Background Questionnaire (LSBQ; Luk and Bialystok, 2013) or reported only limited proficiency in another language. Participants were classified as bilingual if they reported a certain degree of proficiency and usage

TABLE 1 Characteristics of the included studies.

References	Age range (years)	Background/control variables	Independent variables	Outcome variables	Sample size
Bialystok et al. (2017)	Average 21.8	Age; education; SES; verbal and non-verbal intelligence	Bilingualism (cat.) and ADHD diagnosis (cat.)	EF	168
Chung-Fat-Yim et al. (2020)	8–10	Age; SES; verbal intelligence (English and French) and non-verbal intelligence	Bilingualism (cont.) and ADHD symptoms (cont.)	EF	82
Goh et al. (2020)	Longitudinal (2, 3.5, 4.5)	Age; gender; SES; ethnicity; ODD diagnosis; cognitive ability	Bilingualism (cont.)	ADHD diagnosis at 4.5 years	408
Hardy et al. (2021)	6–17	Age; sex; ethnicity	Bilingualism (cat.) and ADHD symptoms (cont.)	EF and visual perception	511
Mor et al. (2015)	19–30	Age; SES; average number weekly hours video gaming	Bilingualism (cat.) and ADHD diagnosis (cat.)	EF	80
Sharma et al. (2022)	5–11	Age; gender; SES; structural language skill in English	Bilingualism (cat., cont.)	ADHD-related behaviour	394
Sharma (2019)	5–11	Age; gender; SES; verbal and non-verbal intelligence; structural language skill in English	Bilingualism (cat., cont.) and ADHD symptoms (cont.)	EF and ADHD-related behaviour	88
Sorge et al. (2017)	8–11	Age; education; SES; verbal and non-verbal intelligence	Bilingualism (cat., cont.) and ADHD symptoms (cont.)	EF	208
Toppelberg et al. (2002)	5–16	Age; gender; ethnicity, SES; immigration data	Bilingualism (cont.)	Attention difficulties	50

SES, socio-economic status; EF, executive function; cat., categorical; cont., continuous.

in another language. In Hardy et al. (2021), children were categorised as bilingual if parents answered that Spanish or Spanish and English were spoken in the home, monolingual if only English. Mor et al. (2015) used a Hebrew version (Prior and Beznos, 2009) of the Language Experience and Proficiency Questionnaire (LEAP-Q; Marian et al., 2007), which includes questions regarding language exposure and ratings of spoken language proficiency. Important to note is that the "monolinguals" in this study were highly proficient in both Hebrew and English, while the "bilinguals" spoke an additional third language.

Three studies (Sorge et al., 2017; Sharma, 2019; Sharma et al., 2022) analysed bilingualism as both categorical (monolingual vs. multilingual) and continuous factor (within their bilingual samples). Sharma (2019) and Sharma et al. (2022) used a language and family background questionnaire based on the Alberta Language Environment Questionnaire (ALEQ; Paradis, 2011). Children of caregivers who indicated their child as bilingual were categorised as bilingual. They were further asked to rate their children on speaking, understanding, reading and writing in all their languages. Information was also gathered on age of onset of bilingualism, and proportion of use of English vs. non-English in the home. A continuous composite language ability score was created from standardising these scores. Sorge et al. (2017) used the LSBQ (Luk and Bialystok, 2013) to quantify the children's language environment and usage.

Finally, three studies (Toppelberg et al., 2002; Chung-Fat-Yim et al., 2020; Goh et al., 2020) analysed bilingualism only as a continuous measure. Chung-Fat-Yim et al. (2020) used a later version of the LSBQ (Anderson et al., 2018). In Goh et al. (2020) a continuous measure of bilingualism was obtained by asking parents to give the aggregate proportion of input their baby received in all their languages, with all scores adding up to 100%. The proportion of time the less-heard languages were used was the measure of bilingualism (0 monolingual, 50—balanced bilingual). The sample in Toppelberg et al. (2002) comprised of English-Spanish speaking-children whose mothers, families and/or caregivers communicated solely or mainly in Spanish. Their language proficiency measure was based on parent ratings of children's use of English or Spanish in different settings and with different people.

3.1.3. Attention/ADHD measures

The included studies differ in whether they analysed attention difficulties as a categorical or continuous factor. Three studies (Mor et al., 2015; Bialystok et al., 2017; Goh et al., 2020) used a categorical distinction between control participants and participants who had a previous ADHD diagnosis or showed clinical levels of attention deficits on diagnostic scales such as the Conners' Adult ADHD Rating Scales (CAARS-S:L; Conners et al., 1999) or the Diagnostic Interview Schedule for Young Children (DISC-YC; Fisher and Lucas, 2006). The remaining six studies used a continuous measure of attention difficulties. Three of these studies (Sorge et al., 2017; Sharma, 2019; Chung-Fat-Yim et al., 2020) used the Strengths and Weaknesses of Attention-Deficit/Hyperactivity Disorder Symptoms and Normal Behaviour Scale (SWAN; Swanson et al., 2012). Hardy et al. (2021) used the Swanson, Nolan and Pelham Parent & Teacher Rating Scale (SNAP-IV, Swanson, Nolan, and Pelham, Fourth Edition). Finally, two studies used diagnostic screeners. Sharma et al. (2022) used the ADHD subscale in the Social Skills Improvement System-Rating Scales (SSIS-RS; Gresham and Elliott, 2008). Toppelberg et al. (2002) used the Child Behaviour Checklist (CBCL) and Teacher's Report Form (TRF, Achenbach and Edelbrock, 1991), which includes an attention difficulties subscale.

3.2. Study results

This section summarises the results of the identified studies, starting with the studies targeting executive function skills or other cognitive abilities (see Table 2), and then turning to the studies examining the effect of bilingualism on ADHD-related behaviour and diagnosis (see Table 3).

An overview of the different executive function components and other cognitive abilities that have been assessed as outcome variables in the included studies can be found in Table 2. Most studies targeted interference control ($n = 6$), some looked at response inhibition ($n = 3$), working memory ($n = 2$), or cognitive flexibility ($n = 2$), while only one study tested decision making ($n = 1$) and delay tolerance ($n = 1$). In the following section we present the findings for each cognitive ability separately, assessing whether there is evidence for a bilingual advantage, disadvantage, or a null effect for that aspect of cognition, and whether bilingualism and attention levels interact. Almost all tasks avoided verbal elements in test trials, apart from the colour-Word Stroop task in Sharma (2019). The study does report, however, that any child unable to read the colour names prior to the task was not allowed to continue.

3.2.1. Interference control

Interference control is the ability to ignore task-irrelevant, competing information. Half of the studies assessing interference control used a version of the Flanker task (Eriksen and Eriksen, 1974). In the critical part of the Flanker task, participants need to indicate the direction of a target stimulus when it is either surrounded by stimuli pointing in the same (congruent) or the opposite (incongruent) direction. Variables of interest are differences in accuracy and reaction times between congruent and incongruent trials.

In their study with four groups of young adults combining language group (monolingual/bilingual) and attention group (ADHD/non-ADHD), Bialystok et al. (2017) found that bilingual participants and participants with no ADHD diagnosis experienced a smaller interference effect once general processing speed was controlled for. Studying multilingual children, Chung-Fat-Yim et al. (2020) found no evidence that higher bilingual proficiency and better attention levels improved interference control. This null finding is in line with Sorge et al. (2017) who studied children of a similar age range and assessed bilingualism both as a categorical factor for the complete sample and continuous factor for the bilingual sample. However, with regards to accuracy in mixed blocks (congruent and incongruent trials mixed), Sorge et al. (2017) found that within the bilingual sample, a higher level of bilingualism improved accuracy and that this boost of bilingualism was more pronounced for children with low attention levels.

Mor et al. (2015) used two different tasks to measure interference control in young adults, a numeric Stroop (Hernández et al., 2010) and a Simon Arrows task (Bialystok et al., 2008). For both tasks, they did not find an overall effect of language group (monolingual/bilingual) or attention group (ADHD/non-ADHD) on reaction times. However, they report a significant three-way interaction between language group, attention group and congruency. Bilinguals with ADHD experienced greater interference from task-irrelevant information compared to bilinguals with no ADHD diagnosis. For the accuracy measure, only attention predicted performance, with ADHD participants being less accurate on incongruent trials than neurotypical controls.

Hardy et al. (2021) used the Inhibition-inhibition subtest from the NEPSY-II test battery (Korkman et al., 2007) to assess interference control in children with clinically significant levels of attention problems. Their results indicate that children with more attention problems experienced more interference from competing stimuli, but that bilingualism (monolingual/bilingual) did not have an effect. Sharma (2019) who used a Colour-Word Stroop (cf. Stroop, 1935) also did not find an effect of bilingualism (both measured categorical and continuous) on reaction times. However, this study reports a marginal effect of bilingualism on accuracy in incongruent trials, with bilingual children being slightly more accurate than monolingual ones.

To sum up, there is mixed evidence for a combined effect of bilingualism and attention levels on interference control, with some evidence for a bilingual advantage in children and adults (Bialystok et al., 2017; Sorge et al., 2017; Sharma, 2019), one study indicating a disadvantage for bilingual adults with ADHD (Mor et al., 2015), and a majority of null findings (Mor et al., 2015; Sorge et al., 2017; Sharma, 2019; Chung-Fat-Yim et al., 2020; Hardy et al., 2021).

TABLE 2 Effect of bilingualism and attention ability/ADHD status on executive function performance.

Targeted EF	Task	References	Results	Bilingual advantage
Interference control	Flanker task	Bialystok et al. (2017)	• RT: positive effect of bilingualism (cat.) and non-ADHD status (cat.) (when controlled for general processing speed), no interaction	+
		Chung-Fat-Yim et al. (2020)	• RT: No effect of bilingualism (cont.) and attention (cont.), no difference between bilingual and trilingual group	0
		Sorge et al. (2017)	Accuracy: • Complete sample: positive effect of attention (cont.), but not bilingualism (cat.), no interaction • Bilingual sample: positive effect of attention (cont.) and bilingualism (cont.), interaction (bigger boost of bilingualism for children with low attention) RT: no effect of attention (cont.) or bilingualism (cat./cont.)	+/0
	Inhibition-inhibition subtest (NEPSY-II)	Hardy et al. (2021)	• Negative effect of attention problems (cont.), no effect of bilingualism (cat.), no interaction	0
	Numeric Stroop	Mor et al. (2015)	• RT: No effect of language group (cat.) or attention group (cat.); three-way interaction between language group, attention group and congruency (greater interference effect for bilinguals with ADHD compared to control) • Accuracy: effect of attention group (cat.), but not language group (cat.)	0/(−)
	Simon Arrows	Mor et al. (2015)	• RT: No effect of language group (cat.) or attention group (cat.); three-way interaction between language group, attention group and congruency (greater interference effect for bilinguals with ADHD compared to control) • Accuracy: effect of attention group (cat.), but not language group (cat.)	0/(−)
	Colour-Word Stroop	Sharma (2019)	Complete sample: • RT: No effect of bilingualism (cat.) • Accuracy: marginal positive effect of bilingualism (cat.) on incongruent but not congruent trials Bilingual sample: • RT/accuracy: no effect of bilingual proficiency (cont.) or length of exposure to strongest non-English language	0/(+)
Response inhibition	Stop-Signal	Bialystok et al. (2017)	• SSRT: effect of attention (cat.) but not language group (cat.), interaction of attention and language group (for bilinguals, bigger difference between ADHD and control group; however, in ADHD group no effect of bilingualism)	0/−
		Sorge et al. (2017)	SSRT: • Complete sample: positive effect of attention (cont.) and bilingualism (cat.); no interaction; positive effect of cognitive ability • Bilingual sample: effect of attention (cont.), interaction between bilingualism (cont.) and attention (cont.) (bigger effect of bilingualism in children with strong attention abilities)	+
	Simon Arrows reverse block	Mor et al. (2015)	• RT: No effect of language group (cat.) or attention group (cat.) • Accuracy: effect of attention group (cat.), but not language group (cat.)	0
Working memory	Animal Sounds Monitoring Task	Sharma (2019)	• Complete sample: No effect of bilingualism (cat.) • Bilingual sample: no effect of bilingual proficiency (cont.) or length of exposure to strongest non-English language	0
	Frog Matrices Task	Sorge et al. (2017)	• Complete sample: positive effect of bilingualism (cat.), no effect of attention (cont.), no interaction, effect of age and cognitive ability • Bilingual sample: positive effect of bilingualism (cont.), no effect of attention, no interaction, effect of age, SES, and cognitive ability	+

(Continued)

TABLE 2 (Continued)

Targeted EF	Task	References	Results	Bilingual advantage
Cognitive flexibility	Trail Making Task	Mor et al. (2015)	• RT/accuracy: no effect of language group (cat.) or attention group (cat.)	0
	Task Switching Paradigm	Mor et al. (2015)	Switching cost (difference in performance between switch and non-switch trials in mixed blocks): • RT: No effect of language group (cat.) or attention group (cat.) • Accuracy: effect of attention group (control more accurate than ADHD group); four-way interaction between attention group, language group, congruency (congruent-incongruent), and trial type (repeat-switch) (only monolingual controls have similar switch costs for congruent and incongruent trials) Mixing cost (difference in performance between single-task blocks and non-switch trials in mixed blocks): • RT: No effect of language group (cat.) or attention group (cat.) • Accuracy: effect of attention group (control more accurate than ADHD group)	0
	Global-Local Task	Sharma (2019)	Complete sample: • RT: no effect of bilingualism (cat.) • Accuracy: positive effect of bilingualism (cat.) Bilingual sample: • RT: no effect of bilingual proficiency (cont.) or length of exposure to non-English language • Accuracy: positive effect of bilingual proficiency, no effect of length of exposure to strongest non-English language	0/+
Decision making (reversal learning)	Child Iowa Gambling Task	Sharma (2019)	• Complete sample: no effect of bilingualism (cat.) • Bilingual sample: no effect of bilingual proficiency (cont.) or length of exposure to strongest non-English language	0
Delay tolerance	Delayed Reward Task	Sharma (2019)	• Complete sample: no effect of bilingualism (cat.) • Bilingual sample: no effect of bilingual proficiency (cont.) or length of exposure to strongest non-English language	0

EF, executive function; cat., categorical; cont., continuous; RT, reaction time. Bilingual advantage: "+", positive effect; "–", negative effect; "0", no effect.

3.2.2. Response inhibition

Response inhibition is the ability to suppress a prepotent response. In the Stop-Signal task, used by Bialystok et al. (2017) and Sorge et al. (2017), participants are trained to quickly and accurately respond to a certain property of a stimulus (e.g., press "F" for a blue circle and "J" for a red circle on a keyboard). In the critical block, some stimuli are followed by an auditory "stop" signal at different intervals after stimulus onset, which requires participants to inhibit their response. The measurement of interest, the so-called Stop-Signal Reaction Time (SSRT), is calculated as the difference between the mean reaction time on "go" trials and the mean Stop-Signal delay, with a lower SSRT indicating better response inhibition.

Using the Stop-Signal task, Bialystok et al. (2017) found an effect of attention group (ADHD/non-ADHD), and a significant interaction between attention and language group. In particular, they found that for bilingual participants, people with an ADHD diagnosis had significantly longer SSRTs than their bilingual peers with no diagnosed attention difficulties. However, bilingualism did not turn out to be an additional burden for people with ADHD since the performance of monolingual and bilingual participants with ADHD did not differ. Using a similar task with children, Sorge et al. (2017) found that both higher attention levels and being bilingual improved response inhibition, with no interaction between these two factors. Within the sample of bilingual children, better attention levels again predicted better performance. In addition, a significant interaction with level of bilingualism indicated that children with strong attention abilities benefitted more from bilingual experience and proficiency than children with weaker attention abilities. Mor et al. (2015) tested habitual response inhibition using a reverse Simon Arrows task, in which participants needed to press the response button in the direction opposite to the one indicated by the arrow. They did not find an effect of attention group or language group on reaction times, and only an effect of attention group on accuracy, indicating that participants with ADHD tended to make more mistakes.

In sum, the available evidence suggests that response inhibition in adults is affected by attention levels, but not bilingualism (Mor et al., 2015; Bialystok et al., 2017). For children, one study indicates that being bilingual might positively affect response inhibition, especially when children have strong attention levels (Sorge et al., 2017).

3.2.3. Working memory

The combined effects of attention and bilingualism on working memory capacity have so far only been assessed in

TABLE 3 Effect of bilingualism on ADHD-related behaviour/ADHD diagnosis.

ADHD-related behaviour/diagnosis	References	Results	Bilingual advantage
ADHD diagnosis at 4.5 y	Goh et al. (2020)	• Association between language delay and ADHD diagnosis only for primarily monolingual children • Children with no language delay: higher odds of ADHD diagnosis with increased bilingualism • Language-delayed children: no significant effect of bilingualism on ADHD diagnosis • No mediating effect of executive function (delay tolerance, cognitive flexibility)	–/0
ADHD-related behaviour	Sharma (2019, Study 1) and Sharma et al. (2022)	• Complete sample: small positive effect of bilingualism (cat.) when controlled for age, sex, and structural language skills • Bilingual sample: no effect of bilingual proficiency (cont.)	+/0
Inattentiveness, hyperactivity/impulsivity	Sharma (2019, Study 2)	• Complete sample: no effect of bilingualism (cat.) on inattentiveness or hyperactivity/impulsivity • Bilingual sample: no effect of bilingual proficiency (cont.) (oral or literacy proficiency in strongest non-English language, length of exposure to strongest non-English language) on inattentiveness or hyperactivity/impulsivity	0
Levels of attention difficulties	Toppelberg et al. (2002)	• Clinical subgroup: inverse correlation between bilingual proficiency (cont.) and attention problems	+

Cat., categorical; cont., continuous. Bilingual advantage: "+", positive effect; "–", negative effect; "0", no effect.

children, with again mixed results. Sharma (2019) created an Animal Sounds Monitoring Task [based on Miyake et al. (2000)'s Tone Monitoring Task], which required children to monitor different animal sounds and to press a designated button when the sound of each particular animal was presented for the third time. No difference in auditory working memory capacity was observed between monolingual and bilingual children, neither effect of bilingual proficiency nor length of exposure within the bilingual sample.

In contrast to that, Sorge et al. (2017) found a bilingual advantage in spatial working memory capacity for both the complete sample and the bilingual sample. They used a Frog Matrices Task (Morales et al., 2013), where children needed to recall how a frog jumped between ponds arranged in a 3 × 3 grid. Bilingual children outperformed their monolingual peers in this task, and within the group of bilingual children a higher degree of bilingualism was related to better working memory capacity. Attention ability did not affect outcomes.

The divergence in findings between Sharma (2019) and Sorge et al. (2017) could be due to differences in their samples or the fact that different aspects of working memory (auditory vs. spatial working memory) were assessed, that could be differentially influenced by bilingualism.

3.2.4. Cognitive flexibility/Shifting

Cognitive flexibility refers to the ability to shift between different concepts or task rules and to adapt the corresponding behavioural response accordingly. Mor et al. (2015) used two tasks to assess cognitive flexibility in adults. In a Hebrew version of the Trail Making Task (Reitan and Davison, 1974), participants were asked to connect numbers and Hebrew letters in alternating order (e.g., 1-"Alef," 2-"Bet,"). In a Task Switching Paradigm, participants needed to switch between sorting figures according to their shape and their colour, depending on a task cue presented visually before each trial. While Mor et al. (2015) did not find an effect of language or attention group on performance on the Trail Making Task, they report several significant results for the Task Switching Paradigm. Looking at switching costs, i.e., the differences in performance when people had to switch between the shape and colour task compared to when no switching was required, they detected that people with ADHD tended to make more mistakes than the control group. In addition, they report a four-way interaction between attention group, language group, congruency (same vs. different response required for colour and shape task), and trial type (repeat vs. switch), in the sense that only participants in the monolingual control group had similar switch costs for congruent and incongruent trials. For mixing costs, defined as the difference in performance between single-task blocks and non-switch trials in mixed blocks, they also found people with ADHD to be less accurate than controls.

Sharma (2019) measured cognitive flexibility with a Global-Local Task (cf. Navon, 1977) in which children needed to shift between paying attention to the overall global shape of a figure and the local shapes it consists of. Sharma (2019) reports an effect of bilingualism on accuracy for both the complete and the bilingual sample. Bilingual children were more accurate than their monolingual peers, and a higher bilingual proficiency was connected to better task performance. However, since attention levels were not included in the model, it is unclear whether both attention and bilingualism are independent predictors of cognitive flexibility.

Taken together, the evidence on how attention abilities and bilingualism affect cognitive flexibility is still sparse and requires further investigation.

3.2.5. Decision-making and delay tolerance

Decision-making and delay tolerance have been assessed in only one of the included studies as outcome variables (Sharma, 2019, but see also Goh et al., 2020 where delay tolerance is used as moderating variable). Decision-making skills were tested *via* a child version of the Iowa Gambling Task (Garon et al., 2006), that required children to decide from which of four decks a card should be turned over, with some decks containing more "good" cards than others. Delay tolerance, the ability to wait for a higher reward, was assessed with a Delayed Reward Task (Cherek et al., 1997). Performance on neither task was related to bilingualism, either for the complete sample or the bilingual subsample.

3.2.6. ADHD-related behaviour

After reviewing the effects of bilingualism on cognitive performance, we now turn to studies looking at ADHD-related symptoms, as reported by parents and teachers (see Table 3). Sharma et al. (2022) assessed ADHD-related behaviour with the ADHD subscale of the SSIS-RS parent form (Gresham and Elliott, 2008), which gives a composite score for several ADHD-related symptoms of inattentiveness, hyperactivity/impulsivity, and oppositional defiant behaviour. They report a small significant effect for bilingualism as a category on levels of ADHD-related behaviour, such that bilingual children showed slightly less ADHD-related behaviour than their monolingual peers, when age, sex, SES, and structural language skills were controlled for. However, within the group of bilingual children, a higher level of bilingual ability (composite of oral proficiency, literacy proficiency, and bilingual use with caregivers) did not predict less ADHD-related behaviour. Sharma (2019, study 2) tested a sample of children both monolingual and bilingual, who scored ≤ -1 SD and $\geq +1$ SD on the SSIS-RS ADHD subscale (as reported in Sharma et al., 2022), using the SWAN (Swanson et al., 2012) to assess ADHD symptomatic behaviour. No relation was found between bilingualism either as a category or continuous measure on inattentiveness or hyperactivity/impulsivity symptoms.

Focusing on bilingual Spanish-English-speaking children referred to psychiatric services, Toppelberg et al. (2002) looked among others at the relationship between bilingual language proficiency and attention problems. For children in the clinical range (CBCL score above the clinical cut-off), limited bilingual skills were associated with heightened attention problems, also when controlling for IQ. For the complete sample, the negative correlation between bilingualism and attention problems was still present, but weaker.

In sum, the limited evidence indicates that exposure to multiple languages could have a positive effect on ADHD-related behaviour such as inattentiveness or hyperactivity/impulsivity symptoms.

3.2.7 ADHD diagnosis

Goh et al. (2020)'s study stands out in its design and methodology from the other identified studies, as they used a longitudinal design examining the prospective association of language delay at 2 years to ADHD diagnosis at 4.5 years. They found that for children primarily exposed to a single language, language delay was significantly associated with increased odds of getting an ADHD diagnosis. For children who did not show signs of language delay at 2 years, higher bilingual exposure increased the odds of getting an ADHD diagnosis at 4.5 years. By contrast, for language-delayed children, increased bilingual exposure did not moderate the association of language delay to ADHD, with a tendency towards increased bilingual exposure reducing the odds of an ADHD diagnosis later in childhood. Executive function skills, as measured by a delay tolerance and a cognitive flexibility task, did not mediate the link between language delay in interaction with bilingualism on ADHD diagnosis.

4. Discussion

This review identified and systematically summarised the available scientific evidence on how bilingualism affects cognitive abilities and ADHD-related behaviours/symptoms in adults and children with high and low attention levels. With only nine identified studies in the literature, the topic is to date not well-studied, and the total number of participants is limited. In addition, there is a big variability in the included studies concerning design and methodology. Not only do the identified studies assess different types of executive functions (e.g., working memory, interference control, cognitive flexibility), the tasks to measure them and the reported outcome variables also differ across studies. Furthermore, there are considerable differences in the conceptualisation and operationalisation of bilingualism and attention levels as continuum, category, or both. This variability makes it difficult to compare and synthesise evidence across studies.

Bilingualism and attention are dimensional constructs, with both having multiple underlying contributing factors, as is evident in the instruments used to assess them. Similarly, clinical ADHD constitutes the end of a dimension or dimensions, that falls along a continuum with the behaviour of neurotypical individuals. It follows that both bilingualism and attention abilities/difficulties, are better understood and explored as dimensional rather than categorical (Coghill and Sonuga-Barke, 2012; Luk and Bialystok, 2013; Roberts et al., 2015). Categorical approaches are no doubt important and useful, for instance for deciding who should be prioritised for intervention or for answering the question whether there is something qualitatively different about being bilingual that influences attention. On the other hand, a dimensional approach may reveal for instance for bilingualism what specific components (e.g., oral or literacy proficiency, age of onset, frequency of use, domains of use), if any, influence levels of behaviour across the different domains of ADHD.

Examining the overall evidence, no clear pattern emerged that individuals with attention deficits show systematic bilingual advantages or challenges in specific executive functions (interference control, response inhibition, working memory, cognitive flexibility), related cognitive abilities (decision-making, delay tolerance), or ADHD-related behaviour. However, among many null findings, several studies reported significant effects of bilingualism, sometimes in interaction with attention levels or ADHD status. For instance, Mor et al. (2015)'s study suggests that for people with ADHD being bilingual might be an extra burden, negatively affecting interference control. However, for this study, it should be borne in mind that the study samples involved might be more accurately described as bilingual vs. trilingual as previously mentioned in the section on bilingual measures.

Two studies found some initial evidence that bilingualism could lead to improvements in ADHD-related symptoms (Toppelberg et al., 2002; Sharma et al., 2022). However, to date there is no theory linking bilingual language experience to behavioural difficulties in conversation such as turn-taking and interrupting. Sharma et al. (2022) suggest that bilingual children's improved perspective-taking skills, as reported in Fan et al. (2015), could play a role. More research is needed to explore potential links between bilingualism and behavioural aspects related to inattention, hyperactivity, and impulsivity.

Some significant effects reported in the identified studies might have been due to a confound between bilingualism and SES. Previous research has shown that lower SES is associated with ADHD (Russell et al., 2015; Michaëlsson et al., 2022). In two of the studies examined, SES may have favoured results for bilinguals over monolinguals. In Bialystok et al. (2017), where bilinguals with ADHD scored better on hyperactivity/impulsivity scales than their monolingual counterparts, their SES was also higher than those of monolinguals with ADHD. In Sharma et al. (2022), bilingual children showed significantly lower levels of ADHD-related behaviour, but they had also a slightly higher SES than monolingual children. In addition, the way SES was operationalized varied considerably between studies, from using a single indicator such as maternal education to using an aggregate of affluence and parental education. We recommend that future studies should carefully consider how to operationalize SES, what components to measure (e.g., income, poverty, wealth, parental education, parental occupation), and whether to include them in the analysis separately or combined (cf. Ensminger and Fothergill, 2003; Duncan et al., 2015). Also other background variables such as gender, age, education, IQ, and structural language abilities need to be measured and compared between groups to prevent any confounds. While most studies assessed and reported at least some of these background factors, they were typically not included as covariates in the statistical analysis.

In light of the overall extremely limited number of available studies, any reported positive or negative effects of bilingualism in people with attention deficits needs to be seen as preliminary and awaits replication. In addition, there are several limitations of the included studies. It remains unclear whether the studies were strictly confirmatory or included explorations of the data, with multiple testing inflating the type I error rate considerably (Ioannidis, 2005). Furthermore, the chance of false-positive outcomes is increased by the statistical approach most studies have selected, analysing their data with correlations, ANOVAs or linear models without a maximal random effects structure (cf. Barr et al., 2013). It therefore needs to be seen whether the significant findings reported in some of the papers can be reproduced with a more suitable type of statistical analysis, taking complex dependencies between observations into account.

Taken together, we did not find support for the hypotheses that bilingualism has systematic positive or negative cognitive or behavioural effects on people with attention deficits, which makes the null hypothesis to date the most plausible candidate. The fact that no clear pattern across the included studies emerged suggests that significant effects might be due to characteristics of individual study samples, or the type of analysis selected rather than being generalizable effects. However, since the current evidence is limited and variability between studies is high, further research on this topic is needed, preferably with pre-registered design and analysis plans. It is particularly important to carefully measure different aspects of bilingualism (e.g., language proficiency, language use in different domains) and attention abilities (e.g., ADHD diagnosis, ADHD-related symptoms) to better understand the conditions under which potential differences in executive function performance emerge (cf. Pot et al., 2018; Grundy, 2020).

Based on the current evidence, exposure to more than one language does not seem to impair the cognitive functioning of people with ADHD or intensify inattention symptoms. This means that people with ADHD do not experience an additional cognitive burden or an added disadvantage by acquiring and using multiple languages. These findings are important beyond the scientific community. There is a widespread fear that children who already face developmental challenges might be overburdened by the demands of learning one or more additional languages. It is not uncommon that parents of children with neurodevelopmental disorders get professional advice that it might be harmful for their child's development if they are exposed to more than one language (e.g., Kay-Raining Bird et al., 2012). In line with Uljarević et al. (2016), we would like to point out that such general recommendations are not backed up by the current state of research. This notwithstanding, there is still a lot we do not know about the bilingual language development of children with developmental disorders [see

Kay-Raining Bird et al. (2016) for a review]. Further research is needed to better understand the challenges that children with attention deficits face when acquiring multiple languages, especially because language and communication disorders are common in this group (Green et al., 2014; Tannock, 2018).

On the other hand, bilingualism also does not "train" attention abilities of people with attention deficits. The results of this review add to the scientific debate on a bilingual advantage by presenting evidence from a subgroup of the population that has attention deficits and might therefore be particularly receptive for attention-related effects of bilingualism. The evidence for the null hypothesis in this population aligns with recent meta-analyses on neurotypical adults and children that did not provide support for domain-general cognitive benefits of bilingual speakers (Lehtonen et al., 2018; Lowe et al., 2021).

Potential bilingual benefits are therefore unlikely to be responsible for lower rates of ADHD diagnosis and treatment in children from minority-language backgrounds (e.g., Slobodin and Masalha, 2020). There may be several other explanations, such as cultural differences in caregiver expectation of development, for instance, not being familiar with ADHD, and also limited knowledge how to access assessment or treatment (Stevens et al., 2004; Rothe, 2005; Eiraldi et al., 2006). Mor et al. (2015) mentioned these as possible explanations for the make-up of their bilingual sample. That is, that only the most severe cases of ADHD among bilinguals appeared to be accounted for, as those with less severe difficulties were unaware of ADHD or decided against seeking help to avoid additional stigma.

In line with previous research on ADHD (Willcutt et al., 2005), the included studies show that children and adults with attention deficits can perform significantly below neurotypical controls in executive functions including interference control, response inhibition, and cognitive flexibility. Interestingly, the effect of attention level was most visible in the accuracy measures. The general tendency across the included studies was that people with limited attentional resources were equally fast—or sometimes slower—than controls, but less accurate. Impairments in the optimisation of the speed-accuracy trade-off have been previously linked to ADHD (Mulder et al., 2010).

5. Conclusion

There is to date little evidence that bilingualism affects cognition or behavioural symptom presentation in children and adults with attention deficits. If bilingual effects are real, their effect sizes will likely be small and practical implications for affected individuals will be limited. Especially, it is unlikely that they will outweigh the clear social and linguistic advantages of bilingualism. Given the impact on individuals, families, and society, however, it is important to continue to investigate why differences exist in assessment, diagnosis, and treatment of ADHD in minority communities, and formulate strategies to address these.

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

FK, SC, and MG contributed to conception and design of the study. SC performed the search in relevant databases. SC and FK screened the papers. FK and CS wrote the first draft of the manuscript. All authors commented on the draft and approved the submitted version.

Funding

The research was supported by a FINNUT grant from the Research Council of Norway (project no. 315368, better attention, better communication? How ADHD and multilingualism influence children's pragmatic development) awarded to FK.

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.

Publisher's note

References

Achenbach, T. M., and Edelbrock, C. (1991). *Manual for the Child Behavior Checklist/4–18 and 1991 Profile*. Burlington, NJ: University of Vermont, Department of Psychiatry.

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders: DSM-V*. doi: 10.1176/appi.books.9780890425596

Anderson, J. A. E., Hawrylewicz, K., and Bialystok, E. (2018). Who is bilingual? Snapshots across the lifespan. *Bilingual. Lang. Cogn.* 23, 929–937. doi: 10.1017/S1366728918000950

Antoniou, M. (2019). The advantages of bilingualism debate. *Annu. Rev. Ling.* 5, 395–415. doi: 10.1146/annurev-linguistics-011718-011820

Askari, M., Fazeli, R., Khademali, M., Aghaee, F., and Piroozan, A. (2019). Comparison of the impact of second language acquisition and Parent Management Training on attention control among kindergartners with Attention Deficiency and Hyperactivity Disorder. *J. Biochem. Technol.* 10, 28–33.

Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: keep it maximal. *J. Mem. Lang.* 68, 255–278. doi: 10.1016/j.jml.2012.11.001

Bialystok, E. (2015). Bilingualism and the development of executive function: the role of attention. *Child Dev. Perspect.* 9, 117–121. doi: 10.1111/cdep.12116

Bialystok, E., Craik, F., and Luk, G. (2008). Cognitive control and lexical access in younger and older bilinguals. *J. Exp. Psychol. Learn. Mem. Cogn.* 34, 859–873. doi: 10.1037/0278-7393.34.4.859

*Bialystok, E., Hawrylewicz, K., Wiseheart, M., and Toplak, M. (2017). Interaction of bilingualism and attention-deficit / hyperactivity disorder in young adults. *Biling. Lang. Cogn.* 20, 588–601. doi: 10.1017/S1366728915000887

Bialystok, E., Klein, R., Craik, F. I. M., 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., and Martin, M. M. (2004). Attention and inhibition in bilingual children: evidence from the dimensional change card sort task. *Dev. Sci.* 7, 325–339. doi: 10.1111/j.1467-7687.2004.00351.x

Bialystok, E., Martin, M. M., and Viswanathan, M. (2005). Bilingualism across the lifespan: the rise and fall of inhibitory control. *Int. J. Biling.* 9, 103–119. doi: 10.1177/13670069050090010701

Bobb, S. C., Von Holzen, K., Mayor, J., Mani, N., and Carreiras, M. (2020). Co-activation of the L2 during L1 auditory processing: an ERP cross-modal priming study. *Brain Lang.* 203, 104739. doi: 10.1016/j.bandl.2019.104739

Carlson, S., and Meltzoff, A. (2008). Bilingual experience and executive functioning in young children. *Dev. Sci.* 11, 282–298. doi: 10.1111/j.1467-7687.2008.00675.x

Cherek, D. R., Moeller, F. G., Dougherty, D. M., and Rhoades, H. (1997). Studies of violent and nonviolent male parolees: II. Laboratory and psychometric measurements of impulsivity. *Biol. Psychiatry* 41, 523–529. doi: 10.1016/S0006-3223(96)00426-X

*Chung-Fat-Yim, A., Sorge, G. B., and Bialystok, E. (2020). Continuous effects of bilingualism and attention on Flanker task performance. *Bilingualism* 23, 1106–1111. doi: 10.1017/S1366728920000036

Coghill, D., and Sonuga-Barke, E. J. S. (2012). Annual research review: Categories versus dimensions in the classification and conceptualisation of child and adolescent mental disorders—Implications of recent empirical study. *J Child Psychol Psychiatry.* 53, 469–489. doi: 10.1111/j.1469-7610.2011.02511.x

Coghill, D., Toplak, M., Rhodes, S., and Adamo, N. (2018). "Cognitive functioning in ADHD: Inhibition, memory, temporal discounting, decision-making, timing and reaction time variability," in *Oxford Textbook of Attention Deficit Hyperactivity Disorder*, eds T. Banaschewski, D. Coghill, and A. (Zuddas: Oxford University Press).

Colomé, À. (2001). Lexical activation in bilinguals' speech production: language-specific or language-independent? *J. Mem. Lang.* 45, 721–736. doi: 10.1006/jmla.2001.2793

Conners, C. K., Erhardt, D., and Sparrow, E. P. (1999). *Conners' Adult ADHD Rating Scales: Technical Manual.* New York, NY: Multi-Health Systems.

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

Dick, A. S., Garcia, N. L., Pruden, S. M., Thompson, W. K., Hawes, S. W., Sutherland, M. T., et al. (2019). No evidence for a bilingual executive function advantage in the nationally representative ABCD study. *Nat. Hum. Behav.* 3, 692–701. doi: 10.1038/s41562-019-0609-3

Duncan, G. J., Magnuson, K., and Votruba-Drzal, E. (2015). "Children and socioeconomic status," in *Handbook of Child Psychology and Developmental Science, Vol. 4*, eds R. M. Lerner, M. H. Bornstein, and T. Leventhal (Hoboken, NJ: Wiley), 534–573. doi: 10.1002/9781118963418.childpsy414

Eiraldi, R. B., Mazzuca, L. B., Clarke, A. T., and Power, T. J. (2006). Service utilization among ethnic minority children with ADHD: a model of help-seeking behavior. *Administr. Policy Mental Health Mental Health Serv. Res.* 33, 607–622. doi: 10.1007/s10488-006-0063-1

Ensminger, M. E., and Fothergill, K. (2003). "A decade of measuring SES: what it tells us and where to go from here," in *Socioeconomic Status, Parenting, and Child Development*, eds M. H. Bornstein and R. H. Bradley (Mahwah, NJ: Erlbaum), 13–27.

Eriksen, B. A., and Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Percept. Psychophys.* 16, 143–149. doi: 10.3758/BF03203267

Fan, S. P., Liberman, Z., Keysar, B., and Kinzler, K. D. (2015). The exposure advantage: early exposure to a multilingual environment promotes effective communication. *Psychol. Sci.* 26, 1090–1097. doi: 10.1177/0956797615574699

Faraone, S. V., Banaschewski, T., Coghill, D., Zheng, Y., Biederman, J., Bellgrove, M. A., et al. (2021). The world federation of ADHD international consensus statement: 208 evidence-based conclusions about the disorder. *Neurosci. Biobehav. Rev.* 128, 789–818. doi: 10.1016/j.neubiorev.2021.01.022

Fisher, P., and Lucas, C. (2006). *Diagnostic Interview Schedule for Children (DISC-IV)-Young Child.* New York, NY: Columbia University.

Garon, N., Moore, C., and Waschbusch, D. A. (2006). Decision making in children with ADHD only, ADHD-anxious/depressed, and control children using a child version of the Iowa Gambling Task. *J. Atten. Disord.* 9, 607–619. doi: 10.1177/1087054705284501

Gathercole, V. C. M., Thomas, E. M., Kennedy, I., Prys, C., Young, N., Guasch, N. V., et al. (2014). Does language dominance affect cognitive performance in bilinguals? Lifespan evidence from preschoolers through older adults on card sorting, Simon, and metalinguistic tasks. *Front. Psychol.* 5, 11. doi: 10.3389/fpsyg.2014.00011

*Goh, S. K. Y., Yang, H., Tsotsi, S., Qiu, A., Chong, Y. S., Tan, K. H., et al. (2020). Mitigation of a prospective association between early language delay at toddlerhood and ADHD among bilingual preschoolers: evidence from the GUSTO cohort. *J. Abnorm. Child Psychol.* 48, 511–523. doi: 10.1007/s10802-019-00607-5

Green, B. C., Johnson, K. A., and Bretherton, L. (2014). Pragmatic language difficulties in children with hyperactivity and attention problems: an integrated review. *Int. J. Lang. Commun. Disord.* 49, 15–29. doi: 10.1111/1460-6984.12056

Gresham, F., and Elliott, S. N. (2008). *Social Skills Improvement System Rating Scales.* Minneapolis, MN: NCS Pearson.

Grosjean, F. (2010). *Bilingual: Life and Reality*. Cambridge, MA: Harvard University Press. doi: 10.4159/9780674056459

Grundy, J. G. (2020). The effects of bilingualism on executive functions: an updated quantitative analysis. *J. Cult. Cogn. Sci.* 4, 177–199. doi: 10.1007/s41809-020-00062-5

*Hardy, L. M., Tomb, M., Cha, Y., Banker, S., Muñoz, F., Paul, A., et al. (2021). Bilingualism may be protective against executive function and visual processing deficits among children with attention problems. *J. Atten. Disord.* 25, 865–873. doi: 10.1177/1087054719861745

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

Hernández, M., Costa, A., Fuentes, L. J., Vivas, A. B., and Sebastián-Gallés, N. (2010). The impact of bilingualism on the executive control and orienting networks of attention. *Biling.: Lang. Cogn.* 13, 315–325. doi: 10.1017/S1366728909990010

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Med.* 2, e124. doi: 10.1371/journal.pmed.0020124

Kay-Raining Bird, E., Genesee, F., and Verhoeven, L. (2016). Bilingualism in children with developmental disorders: a narrative review. *J. Commun. Disord.* 63, 1–14. doi: 10.1016/j.jcomdis.2016.07.003

Kay-Raining Bird, E., Lamond, E., and Holden, J. (2012). Survey of bilingualism in autism spectrum disorders. *Int. J. Lang. Commun. Disord.* 47, 52–64. doi: 10.1111/j.1460-6984.2011.00071.x

Korkman, M., Kirk, U., and Kemp, S. (2007). *NEPSY-II: A Developmental Neuropsychological Assessment, 2nd Edn.* San Antonio, TX: Harcourt Assessment.

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, 394–425. doi: 10.1037/bul0000142

Lowe, C. J., Cho, I., Goldsmith, S. F., and Morton, J. B. (2021). The bilingual advantage in children's executive functioning is not related to language status: a meta-analytic review. *Psychol. Sci.* 32, 1115–1146. doi: 10.1177/0956797621993108

Luk, G., and Bialystok, E. (2013). Bilingualism is not a categorical variable: interaction between language proficiency and usage. *J. Cogn. Psychol.* 25, 605–621. doi: 10.1080/20445911.2013.795574

Marian, V., Blumenfeld, H., and Kaushanskaya, M. (2007). The language experience and proficiency questionnaire (LEAP-Q): assessing language profiles in bilinguals and multilinguals viorica. *J. Speech Lang. Hear. Res.* 50, 940–967. doi: 10.1044/1092-4388(2007/067)

Martin-Rhee, M. M., and Bialystok, E. (2008). The development of two types of inhibitory control in monolingual and bilingual children. *Bilingualism* 11, 81–93. doi: 10.1017/S1366728907003227

Marzecová, A., Bukowski, M., Correa, Á., Boros, M., Lupiáñez, J., and Wodniecka, Z. (2013). Tracing the bilingual advantage in cognitive control: the role of flexibility in temporal preparation and category switching. *J. Cogn. Psychol.* 25, 586–604. doi: 10.1080/20445911.2013.809348

Michaëlsson, M., Yuan, S., Melhus, H., Baron, J. A., Byberg, L., Larsson, S. C., et al. (2022). The impact and causal directions for the association between diagnosis of ADHD, socioeconomic status, and intelligence by use of a bi-directional two-sample Mendelian randomization design. *BMC Med.* 20, 106. doi: 10.1186/s12916-022-02314-3

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., and Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: a latent variable analysis. *Cogn. Psychol.* 41, 49–100. doi: 10.1006/cogp.1999.0734

*Mor, B., Yitzhaki-Amsalem, S., and Prior, A. (2015). The joint effect of bilingualism and ADHD on executive functions. *J. Atten. Disord.* 19, 527–541. doi: 10.1177/1087054714527790

Morales, J., Calvo, A., and Bialystok, E. (2013). Working memory development in monolingual and bilingual children. *J. Exp. Child Psychol.* 114, 187–202. doi: 10.1016/j.jecp.2012.09.002

Mulder, M. J., Bos, D., Weusten, J. M. H., Van Belle, J., Van Dijk, S. C., Simen, P., et al. (2010). Basic impairments in regulating the speed-accuracy tradeoff predict symptoms of attention-deficit/hyperactivity disorder. *Biol. Psychiatry* 68, 1114–1119. doi: 10.1016/j.biopsych.2010.07.031

Namazi, M., and Thordardottir, E. (2010). A working memory, not bilingual advantage, in controlled attention. *Int. J. Biling. Educ. Biling.* 13, 597–616. doi: 10.1080/13670050.2010.488288

Navon, D. (1977). Forest before trees: the precedence of global features in visual perception. *Cogn. Psychol.* 9, 353–383. doi: 10.1016/0010-0285(77)90012-3

Nigg, J. T., Willcutt, E. G., Doyle, A. E., and Sonuga-Barke, E. J. S. (2005). Causal heterogeneity in attention-deficit/hyperactivity disorder: do we need neuropsychologically impaired subtypes? *Biol. Psychiatry* 57, 1224–1230. doi: 10.1016/j.biopsych.2004.08.025

Ouzzani, M., Hammady, H., Fedorowicz, Z., and Elmagarmid, A. (2016). Rayyan-a web and mobile app for systematic reviews. *Syst. Rev.* 5, 1–10. doi: 10.1186/s13643-016-0384-4

Özerk, M. R., Handorff, J. A., and Özerk, K. (2011). Assessment of bilingual children with "inattention", "over activity" and "impulsivity" - challenges and solutions. *Int. Electron. J. Elementary Educ.* 3, 193–212.

Paap, K. R. (2019). "The bilingual advantage debate: quantity and quality of the evidence," in *The Handbook of the Neuroscience of Multilingualism*, ed J. W. Schwieter (Hoboken, NJ: Wiley-Blackwell), 701–735. doi: 10.1002/9781119387725.ch34

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

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 372, n71. doi: 10.1136/bmj.n71

Paradis, J. (2011). Individual differences in child English second language acquisition. *Ling. Approaches Biling.* 1, 213–237. doi: 10.1075/lab.1.3.01par

Poarch, G. J., and Krott, A. (2019). A bilingual advantage? An appeal for a change in perspective and recommendations for future research. *Behav. Sci.* 9, 95. doi: 10.3390/bs9090095

Poarch, G. J., and van Hell, J. G. (2012). Executive functions and inhibitory control in multilingual children: evidence from second-language learners, bilinguals, and trilinguals. *J. Exp. Child Psychol.* 113, 535–551. doi: 10.1016/j.jecp.2012.06.013

Polanczyk, G., Lima, M. S., de, Horta, B. L., Biederman, J., and Rohde, L. A. (2007). The worldwide prevalence of ADHD: a systematic review and metaregression analysis. *Am. J. Psychiatry* 164, 942–948. doi: 10.1176/ajp.2007.164.6.942

Polanczyk, G. V., Willcutt, E. G., Salum, G. A., Kieling, C., and Rohde, L. A. (2014). ADHD prevalence estimates across three decades: an updated systematic review and meta-regression analysis. *Int. J. Epidemiol.* 43, 434–442. doi: 10.1093/ije/dyt261

Pot, A., Keijzer, M., and de Bot, K. (2018). Intensity of multilingual language use predicts cognitive performance in some multilingual older adults. *Brain Sci.* 8, 92. doi: 10.3390/brainsci8050092

Prior, A., and Beznos, M. (2009). *Hebrew Version of the Language Experience and Proficiency Questionnaire*. Unpublished instrument.

Ramos, E., Suarez, M., Hart, K., and Graziano, P. (2019). Syntactic and semantic abilities of bilingual versus monolingual preschoolers with language impairment and ADHD. *Int. J. Lang. Ling.* 6, 1–8. doi: 10.30845/ijll.v6n2p1

Ratiu, I., and Azuma, T. (2015). Working memory capacity: is there a bilingual advantage? *J. Cogn. Psychol.* 27, 1–11. doi: 10.1080/20445911.2014.976226

Reitan, R. M., and Davison, L. A. (1974). *Clinical Neuropsychology: Current Status and Applications*. Washington, DC: Winston.

Roberts, W., Milich, R., and Barkley, R. A. (2015). "Primary symptoms, diagnostic criteria, subtyping, and prevalence of ADHD," in *Attention-Deficit Hyperactivity Disorder: A Handbook for Diagnosis and Treatment*, ed R. A. Barkley (New York, NY: The Guilford Press), 51–80.

Rothe, E. M. (2005). Considering cultural diversity in the management of ADHD in Hispanic patients. *J. Natl. Med. Assoc.* 97, 17–23.

Russell, A. E., Ford, T., and Russell, G. (2015). Socioeconomic associations with ADHD: Findings from a mediation analysis. *PLoS ONE.* 10, 1–16. doi: 10.1371/journal.pone.0128248

Salvatierra, J. L., and Rosselli, M. (2011). The effect of bilingualism and age on inhibitory control. *Int. J. Biling.* 15, 26–37. doi: 10.1177/1367006910371021

*Sharma C. J. M. (2019). *The effect of bilingualism on Attention Deficit Hyperactivity Disorder (ADHD)-related behaviour, ADHD symptoms, and executive functions in a general primary school population* [Unpublished doctoral dissertation]. Cambridge, UK: University of Cambridge.

*Sharma, C. J. M., Katsos, N., and Gibson, J. L. (2022). Associations between bilingualism and attention-deficit hyperactivity disorder (ADHD)-related behavior in a community sample of primary school children. *Appl. Psycholinguist.* 43, 707–725. doi: 10.1017/S0142716422000108

Sjöwall, D., Roth, L., Lindqvist, S., and Thorell, L. B. (2013). Multiple deficits in ADHD: Executive dysfunction, delay aversion, reaction time variability, and emotional deficits. *J. Child Psychol. Psychiatry* 54, 619–627. doi: 10.1111/jcpp.12006

Slobodin, O., and Masalha, R. (2020). Challenges in ADHD care for ethnic minority children: a review of the current literature. *Transcult. Psychiatry* 57, 468–483. doi: 10.1177/1363461520902885

Sonuga-Barke, E. J. S. (2005). Causal models of attention-deficit/hyperactivity disorder: from common simple deficits to multiple developmental pathways. *Biol. Psychiatry* 57, 1231–1238. doi: 10.1016/j.biopsych.2004.09.008

*Sorge, G. B., Toplak, M. E., and Bialystok, E. (2017). Interactions between Levels of attention ability and levels of bilingualism in children's executive functioning. *Dev. Sci.* 20, 1–28. doi: 10.1111/desc.12408

Spivey, M. J., and Marian, V. (1999). Cross talk between native and second languages: partial activation of an irrelevant lexicon. *Psychol. Sci.* 10, 281–284. doi: 10.1111/1467-9280.00151

Starreveld, P., De Groot, A., Rossmark, B., and Van Hell, J. (2014). Parallel language activation during word processing in bilinguals: evidence from word production in sentence context. *Biling. Lang. Cogn.* 17, 258–276. doi: 10.1017/S1366728913000308

Stevens, J., Harman, J. S., and Kelleher, K. J. (2004). Ethnic and regional differences in primary care visits for attention-deficit hyperactivity disorder. *J. Dev. Behav. Pediatr.* 25, 318–325. doi: 10.1097/00004703-200410000-00003

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *J. Exp. Psychol.* 18, 643–662. doi: 10.1037/h0054651

Swanson, J. M., Schuck, S., Porter, M. M., Hartman, C. A., and Sergeant, J. A. (2012). Categorical and dimensional definitions and evaluations of symptoms of ADHD: history of the SNAP and the SWAN rating scales. *Int. J. Educ. Psychol. Assess.* 10, 51–70.

Tannock, R. (2018). "ADHD and communication disorders," in *Oxford Textbook of Attention Deficit Hyperactivity Disorder*, eds T. Banaschewski, D. Coghill, and A. Zuddas (Oxford: Oxford University Press). doi: 10.1093/med/9780198739258.003.0028

Thomas-Sunesson, D., Hakuta, K., and Bialystok, E. (2018). Degree of bilingualism modifies executive control in Hispanic children in the USA. *Int. J. Biling. Educ. Biling.* 21, 197–206. doi: 10.1080/13670050.2016.1148114

Timmermeister, M., Leseman, P., Wijnen, F., and Blom, E. (2020). No bilingual benefits despite relations between language switching and task switching. *Front. Psychol.* 11, 1–13. doi: 10.3389/fpsyg.2020.01832

*Toppelberg, C. O., Medrano, L., Morgens, L. P., and Nieto-Castañon, A. (2002). Bilingual children referred for psychiatric services: associations of language disorders, language skills, and psychopathology. *J. Am. Acad. Child Adolesc. Psychiatry* 41, 1832. doi: 10.1097/00004583-200206000-00011

*indicate studies that were included in the systematic review.

Uljarević, M., Katsos, N., Hudry, K., and Gibson, J. L. (2016). Practitioner Review: multilingualism and neurodevelopmental disorders – an overview of recent research and discussion of clinical implications. *J. Child Psychol. Psychiatry Allied Discipl.* 57, 1205–1217. doi: 10.1111/jcpp.12596

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

Willcutt, E. G. (2012). The prevalence of DSM-IV attention-deficit/hyperactivity disorder: a meta-analytic review. *Neurotherapeutics* 9, 490–499. doi: 10.1007/s13311-012-0135-8

Willcutt, E. G., Doyle, A. E., Nigg, J. T., Faraone, S. V., and Pennington, B. F. (2005). Validity of the executive function theory of attention-deficit/hyperactivity disorder: a meta-analytic review. *Biol. Psychiatry* 57, 1336–1346. doi: 10.1016/j.biopsych.2005.02.006

Yang, E. (2017). Bilinguals' working memory (WM) advantage and their dual language practices. *Brain Sci.* 7, 86. doi: 10.3390/brainsci7070086

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 03 February 2023
DOI 10.3389/fpsyg.2023.1096399

Check for updates

OPEN ACCESS

EDITED BY
Gary Morgan,
Fundació per a la Universitat Oberta de Catalunya, Spain

REVIEWED BY
Eliseo Diez-Itza,
University of Oviedo,
Spain
Musa Nushi,
Shahid Beheshti University,
Iran
Balachandran Vadivel,
Cihan University-Duhok,
Iraq

*CORRESPONDENCE
Weiying Huang
✉ huangariel@163.com

†These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

RECEIVED 12 November 2022
ACCEPTED 17 January 2023
PUBLISHED 03 February 2023

CITATION
Huang W and Lu X (2023) The difficulty of oral speech act production tasks in second language pragmatics testing.
Front. Psychol. 14:1096399.
doi: 10.3389/fpsyg.2023.1096399

The difficulty of oral speech act production tasks in second language pragmatics testing

Weiying Huang[1*†] and Xiaofei Lu[2†]

[1]School of Foreign Languages, East China University of Technology, Nanchang, China, [2]Department of Applied Linguistics, The Pennsylvania State University, University Park, PA, United States

This study examined the relative difficulty of oral speech act production tasks involving eight different types of speech acts for Chinese English as a foreign language (EFL) learners and the effects of three contextual variables, namely, power, social distance, and imposition, on such difficulty. Eight Oral Discourse Completion Task items, each representing a unique combination of the three contextual variables, were designed for each speech act. Eighty Chinese EFL learners responded to these items and their responses were rated for appropriateness by two native-speaking college English instructors. A Many-facet Rasch Measurement analysis suggested that the eight speech acts can be ordered by ascending difficulty as follows: Thank, Request, Suggestion, Disagreement, Invitation, Refusal, Offer, and Apology. Significant effects on performance scores were found for the interaction between each of the three contextual variables and speech act, and the specific effects observed varied by speech act. The implications of our findings for L2 pragmatics testing are discussed.

KEYWORDS

pragmatic ability, speech acts, situational variables, task difficulty estimates, L2 pragmatics testing

Introduction

Pragmatic ability, that is, the ability to understand the intended meanings communicated by the speaker and to use language appropriately in various communicative contexts (Ross and Kasper, 2013; Ren, 2022), is a crucial component in models of communicative language ability (Purpura, 2004; Bachman and Palmer, 2010). Albeit recent developments in second language pragmatics testing have shown a growing interest in interactive, discursively oriented assessment of interactional competence (for instance, Grabowski, 2009, 2013; Youn, 2015, 2019; Ikeda, 2017; Galaczi and Taylor, 2018), an important part of second language (L2) pragmatics testing involves assessing L2 learners' ability to realize different speech acts under different circumstances (Ross and Kasper, 2013). Research in this area has attended to the effects of different task features and contextual variables on the difficulty of pragmatic tasks (e.g., Hudson, 2001; Taguchi, 2007; Youn, 2019). At the same time, while language users' ability to perform various speech acts has been recognized as the universality of pragmatics (Searle, 1969), linguistic means to engage in those speech acts and the socio-pragmatic norms associated with them exhibit considerable variation across languages and cultures (Taguchi, 2012). This variation poses challenges for learning L2 speech acts and points to the need to take first language (L1) cultural background into account in assessing task difficulty. As identified in Roever's (2007) study, one fourth of his test items in a pragmatics test showed differential functioning for test takers of Asian and European background. Indeed, a few studies have designed or evaluated L2 pragmatics tests with learners' L1 background in mind (e.g., Fulcher and Reiter, 2003; Liu, 2006, 2007). However, systematical explorations of the difficulty of L2 oral production tasks involving a diverse range of speech acts and representing

10.3389/fpsyg.2023.1096399

diverse combinations of contextual factors for learners from a specific L1 cultural background remain scant.

Task difficulty in oral proficiency assessment

Commonly used frameworks of task difficulty within second language acquisition (SLA) have focused on analyzing the degree of cognitive load and complexity of tasks (e.g., Skehan, 1998; Robinson, 2001). Skehan's (1998) Limited Attentional Capacity Model and Robinson's (2001) Cognition Hypothesis both hypothesize that manipulating the cognitive complexity and communicative requirements of a task will produce differential cognitive and communicative demands and affect the accuracy and complexity of the language that learners use to perform the task. Skehan (1998) proposed three dimensions of task difficulty: code complexity (i.e., the variety and difficulty of the linguistic forms required for performing the task), cognitive complexity (i.e., the cognitive processing demands of the task content, such as the type of information to be processed), and communicative stress (i.e., stress caused by task-related factors such as time pressure). His model predicts a competition between accuracy and complexity as a result of limited attentional resources. Robinson's (2001) triadic framework distinguishes task complexity features affected by cognitive factors (e.g., number of elements to deal with) from task condition features affected by interactional factors (e.g., power difference of the interlocutors) and task difficulty features affected by learner factors (e.g., learner motivation). His Cognition Hypothesis claims that increased task complexity may simultaneously promote linguistic complexity and accuracy as learners will activate and allocate more attentional resources to handle the higher cognitive load.

A few language assessment studies have applied these cognitive models of task complexity to examine the effect of varying task conditions on task difficulty in speaking tests. Based on Skehan's (1998) cognitive complexity framework, Iwashita et al. (2001) manipulated the performance conditions of a series of picture-based narrative task in terms of perspective (first vs. third person perspective), immediacy (here and now vs. there and then), adequacy (a complete set of pictures vs. an incomplete set), and planning time (no planning time vs. 3 min planning time). They found no significant effect of the varying performance conditions on either the test-takers' discourse in terms of fluency, complexity, or accuracy or the quality ratings of their performance. Elder et al. (2002) further reported that the varying performance conditions did not affect task difficulty as perceived by the test-takers. They concluded that their results did not support Skehan's framework in the case of oral proficiency assessment. The lack of score sensitivity to varying task conditions in speaking tests has also been reported in other studies (Fulcher, 1996; Fulcher and Reiter, 2003). Accordingly, Fulcher and Reiter (2003) suggested that L2 pragmatics test designers "may look to pragmatic categories and cultural factors to develop task types" (p. 339).

Speech acts, contextual variables, and task difficulty in L2 pragmatics testing

A common way to attend to pragmatic categories in L2 pragmatics testing has been to look at different speech acts. Indeed, the speech act paradigm has played an important role in pragmatics testing since the 1980s, with the influence of studies in the Cross-Cultural Speech Act Realization Patterns (CCSARP) project initiated to investigate cross-cultural variations in speech act realization (Cohen and Olshtain, 1981; Blum-Kulka et al., 1989). Given that the linguistic realization patterns of speech acts have been found to differ from culture to culture (Gass and Neu, 1996; Taguchi, 2012), L2 learners' pragmatic ability to realize different speech acts in the target language has been recognized as an essential component of their L2 communicative language ability (Bachman, 1990; Bachman and Palmer, 1996, 2010) and a prominent target construct of L2 pragmatics testing (Roever, 2011).

Pragmatics tests of speech act realization have drawn heavily from Speech Act theory (Searle, 1969) and Politeness theory (Brown and Levinson, 1987). Speech Act theory views as the minimum unit of human communication the performance of different acts through language (e.g., apology and refusal) and distinguishes direct speech acts, where the speaker directly states the intended meaning, usually with certain conventionalized linguistic forms, from indirect ones, where the speaker says more than or something other than the intended meaning (Searle, 1975). In Politeness theory, the directness of speech acts is seen to vary systematically with three contextual properties defined *a priori*, i.e., power, social distance, and rank of imposition (Brown and Levinson, 1987). L2 pragmatics tests commonly examine L2 learners' realization of different speech acts in situations with different contextual properties, although the most commonly investigated types of speech acts have centered around apology, refusal, and request (Hudson et al., 1992, 1995; Yamashita, 1996; Yoshitake, 1997; Ahn, 2005; Roever, 2005, 2006; Liu, 2006, 2007).

Among the task types used to test speech act production in pragmatics testing, Discourse Completion Tasks (DCTs) are used more widely than other types such as role plays and sociopragmatic judgment tasks (Martínez-Flor and Usó-Juan, 2010). Although DCTs are artificial in nature (Brown, 2001; Golato, 2003), they allow for the evaluation of learners' pragmatic knowledge and are the most prevalent data collection method in L2 pragmatics. Hudson et al. (1992, 1995) designed a prototypical pragmatics test battery for apology, refusal, and request, which included six types of DCTs, namely, Written Discourse Completion Tasks (WDCT), Multiple-Choice Discourse Completion Tasks (MDCT), Oral Discourse Completion Tasks (ODCTs), Discourse Role-Play Tasks (DRPT), Discourse Self-Assessment Tasks (DSAT), and Role-Play Self-assessments (RPSA). All tasks other than self-assessments were designed around high/low settings of power, social distance, and imposition (Brown and Levinson, 1987), rendering eight combinations of these contextual variables. Each task required test-takers to produce an oral or written response to a specific scenario representing a particular combination of contextual variables.

A limited number of studies have examined how pragmatic production tasks involving different speech acts compared with each other in terms of difficulty or how different contextual variables affect the difficulty of such tasks, sometimes with attention to the effects of assessment methods and/or L1 cultural background. Hudson (2001) examined the effects of three assessment methods (i.e., WDCTs, language lab DCTs, and role-play scenarios) and three contextual variables (i.e., power, social distance, and imposition) on the scores assigned to pragmatic productions tasks involving three speech acts (i.e., apologies, refusals, and requests) among Japanese English as a second language (ESL) learners. He found that lab DCTs were slightly more difficult than the other two methods and that apologies were rated slightly higher than refusals and requests. He reported minimal effects of the contextual variables on the scores, with only imposition showing a slight effect, and attributed the lack of effects to the homogeneity of the participants' proficiency level. Fulcher and Reiter (2003) examined how social power and imposition as well as their interaction with learners' L1 background affect test-takers' pragmatic performance. Six role-play tasks representing six combinations of the two contextual variables were used to elicit L2 English learners' realization of request. Significant effects were found for both contextual variables, the two-way interaction between social power and L1 background, and the three-way interaction between social power, imposition and L1 background. Roever (2004) reviewed item difficulty in pragmatics tests including learners' interpretation of routines, implicature and production of speech acts and identified degree of imposition as a source of speech act difficulty. The effect of degree of imposition on the difficulty of speech act performance was also evident in Taguchi's (2007) study, in which she examined the effects of task difficulty on Japanese EFL learners' oral production of requests and refusals. She operationalized task difficulty as two situation types, one with an equal power relationship, small social distance, and a small degree of imposition (PDR-low), and the other with greater power for the listener, large social distance, and a large degree of imposition (PDR-high). She reported that L2 learners produced speech acts significantly more easily and quickly in the PDR-low situation than in the PDR-high situation. In a study designed to evaluate the reliability of three test methods (WDCT, MDCT, and DST) for assessing the pragmatic knowledge of Chinese EFL learners, Liu (2006) reported that the three methods were reasonably reliable, and that the apology subtest proved consistently more difficult than the request subtest across three test methods. However, compliment responses and refusals were found relatively easy while requests were more difficult for L2 Chinese learners in Li et al. (2019). Krish and May (2020) identified interference of L1 cultural knowledge and linguistic rules in L2 Chinese learners' pragmatic performance of five speech acts: compliments, requests, refusals, apologies, and complaints.

Taken together, these studies have provided evidence that pragmatic tasks involving different speech acts may have varying degrees of difficulty for L2 learners and that their relative difficulty may be affected by the learners' L1 background and proficiency level, the assessment method used, and the contextual variables of power, social distance, and imposition. Meanwhile, it can also be seen that the range of speech acts and the range of combinations of different contextual variables that have been investigated in previous studies were both small, and the interaction between the contextual variables and speech acts has been underexamined. How learners' native culture may influence their performance in pragmatics tests has barely been touched upon.

Objectives

The current study contributes to the limited body of research in this area by examining the difficulty of oral production tasks involving different types of speech acts for Chinese English as foreign language (EFL) learners. In response to the call for broadening the range of pragmatic tasks and attending to the effects of relevant contextual variables in assessing task difficulty in pragmatics testing (Taguchi, 2007; Youn, 2019), we include eight speech acts and three contextual variables in designing the oral production tasks. It is our hope that our analysis will provide useful insight into the relative difficulty of oral production tasks involving different speech acts for Chinese EFL learners and the effects of the interaction between the contextual variables and speech act on task difficulty in L2 pragmatics tests. Informed by findings of previous studies, we explored these issues with a single assessment method and a group of learners from a single L1 background (i.e., Chinese EFL learners) representing diverse proficiency levels.

Research questions

The present study explores the difficulty of oral speech act production tasks for Chinese EFL learners in L2 pragmatics testing by addressing the following research questions:

(1) What is the order of the difficulty estimates for oral speech act production tasks involving the speech acts of Apology, Disagreement, Thank, Request, Suggestion, Invitation, Offer and Refusal?
(2) How do social distance, relative power, and imposition interact with speech act to affect the difficulty of oral speech act production tasks?

Methodology

Participants

Eighty Chinese EFL learners (24 male, 56 female) with an average age of 20.6 from three universities in south China responded to an open call to participate in the current study. The participants represented a range of disciplinary backgrounds, years in college, and language proficiency levels, with 35 first-and second-year non-English major undergraduate students from various arts and science disciplines, 40 first-and third-year English major undergraduate students, and five applied linguistics postgraduate students who majored in English in college. No participant had been abroad for over 1 month.

Instruments

Given that our participants were all undergraduate and postgraduate students, we decided to test their pragmatic performance on speech acts commonly used in university settings. To this end, we identified 20 speech acts commonly discussed in the Interlanguage Pragmatics (ILP) literature and invited 28 L1 English American college students to rate the frequency of using each of them in their university life on a five-point scale. Based on their ratings, we included the following eight highest ranked speech acts in the current study: Apology, Disagreement, Thank, Request, Suggestion, Invitation, Offer, and Refusal.

We elicited the participants' performance in producing target speech acts orally using Oral Discourse Completion Tasks (ODCTs). DCTs have been criticized for limited generalizability (Roever, 2011), but ODCTs can measure online performance under time pressure (Roever, 2004), which improves their authenticity and generalizability. To test the participants' pragmatic ability to cope with different contexts, we incorporated different combinations of three contextual variables, i.e., relative power, social distance, and imposition in the ODCTs, with the values of these variables specified for each speech act production task. Relative power (P) refers to the power of the speaker with respect to the hearer (Brown and Levinson, 1987), and P+, P−, and P= denote the speaker has more, less, or equal power relative to the hearer, respectively, with more power defined as a higher rank, title, or social position or greater control of the assets in the situation. We excluded scenarios with the P+ feature in the current study as we limited the discourse context to the university setting, in which such scenarios were uncommon for our participants. Common scenarios with the P= feature included talking to classmates and roommates, and common scenarios with the P− feature included talking to faculty and staff members. Social Distance (D) refers to the degree of familiarity and solidarity between the speaker and the hearer (Brown and Levinson, 1987). D+ indicates that the speaker and hearer are unfamiliar with each other, and D-indicates that they are familiar with each other. Imposition (R) refers to the expenditure of goods and/or services by the hearer or the obligation of the speaker to perform an act (Brown and Levinson, 1987). Given that the nature of this variable varies with different speech acts, we determined the value of this variable for each item in two steps. The speech events in the ODCTs were first ranked for imposition by two native speaker consultants through collaborative discussion. The rankings were then used to code the task items pertaining to the same speech act as either R+ (high imposition) or R− (low imposition), depending on whether each item was ranked in the top or bottom half among the items for that speech act.

We initially developed eight ODCT items for each target speech act, each with a scenario reflecting a unique combination of the three contextual variables, as summarized in Table 1. Each item was checked by two native speaker consultants for authenticity. The consultants recommended the removal of four items for Disagreement on the basis that they represented unrealistic scenarios. One consultant indicated that "it's better to remain quiet if you do not agree in these cases." Therefore, only four items were retained for Disagreement (Item 1, 2, 3, 5). All other items were accepted by the consultants as authentic. The final test battery thus consisted of 60 ODCT items (see Appendix).

Procedure

The pragmatics test was first piloted with five Chinese EFL learners enrolled in the same university who did not participate in the actual study. They all found the scenario descriptions clear, but two participants identified several words in the descriptions that caused some comprehension difficulties. We thus added Chinese glosses to those words to minimize potential comprehension problems. Based on the maximum time they took to complete each item, we set the time limit to 20 s for the first 50 items and 50 s for the last 10 items due to the extended length of these items.

The final test was administered to the 80 participants in a large language lab in 12 groups of six to seven, with ample space between any two participants to minimize interference from each other. At the beginning of each session, one researcher provided instructions in English, illustrated the scenario descriptions and the types of oral response expected with an example, and confirmed that all participants understood the instructions and requirements. The researcher then presented the scenario descriptions and their corresponding time limits using PowerPoint slides on a screen in the front of the lab one by one. There was a signal for the participants to stop speaking at the end of the time limit for each item, and the next slide was shown. The entire session lasted about 1 h for each group. Each participant's responses were recorded by the computer and then saved in a separate audio file for rating and further analysis.

TABLE 1 Combinations of the three contextual variables represented by the eight ODCT items for each speech act.

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8
D	−	+	−	+	−	+	−	+
P	=	=	−	−	=	=	−	−
R	−	−	−	−	+	+	+	+

D, social distance; P, relative power; R, imposition.

Data analysis

Each participant's oral response to each item was firstly transcribed and their written responses were independently rated for pragmatic appropriateness by two native speakers of American English, both of whom were experienced English instructors at the university. A holistic five-point scale was adopted from the five-level rating scale constructed to evaluate Chinese EFL's written speech act performance by Chen and Liu (2016). Inter-rater reliability, assessed using Spearman's rank correlation, reached 0.823 ($p < 0.001$). The final score of each response was the mean of the two scores, and the overall test score of each participant was the sum of the scores for all responses by that participant.

We subjected the scores of the 80 participants' responses to the 60 ODCT items to a Many-facet Rasch Measurement (MFRM) analysis within Item Response Theory (McNamara and Knoch, 2012) using the FACETS 3.71.3 (Linacre, 2013) for the analyses, with participants, speech acts, and item types as facets to assess the difficulty of items for each speech act as well as items of each of the eight types representing a specific combination of the three contextual variables. We further performed a series of two-way ANOVAs, each with speech act and one of the three contextual variables as independent variables and participants' response scores as the dependent variable, to examine the effects of the interaction between each contextual variable and speech act on the difficulty of oral speech act production tasks. Cohen's *D*, or standardized mean difference, was adopted as an effect size measure. Following Cohen (1969), we characterized effect sizes as small, medium, and large if the η_p^2 values were larger than 0.0099, 0.0588, and 0.1379, respectively.

Results

Research question 1: Order of the difficulty estimates for tasks involving different speech tasks

The MFRM analysis placed the estimates of the three facets (i.e., participants, speech acts, and item types) on a single measurement scale, as shown in Figure 1. The range of the measurements was within two logits, likely due to the narrow range of the ILP competence of our participants. The average person measure was 0.16, with a standard deviation of 0.22. Only four misfitting persons were identified with Z scores larger than two.

For the speech act measures, the mean measure was set at zero and the standard deviation was calculated to be 0.30. Thank and Request were found to be the easiest, followed by Suggestion, Disagreement, and Invitation. Refusal, Offer, and Apology were found to be the most difficult among the eight speech acts.

Facets also generates an overall estimate of the extent to which items are at reliably different levels of difficulty. The reliability of separation

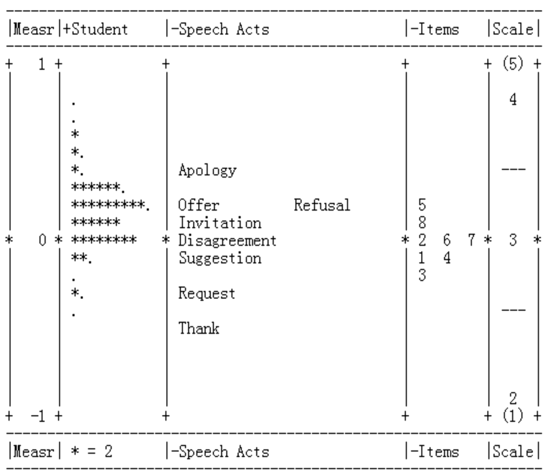


FIGURE 1
Results of the Many-facet Rasch Measurement analysis of participant performance.

index denotes the reliability with which the items included in the analysis are separated (i.e., how different the item difficulty measures are), and the fixed chi-square test for the items tests the hypothesis that all items are of the same level of difficulty, after accounting for measurement error. The reliability of separation was reported as 0.90 [$\chi^2(7) = 74.0$, $p = 0.000$], indicating significant differences among the test items in terms of difficulty.

For the item type measures, the mean measure was set at zero and the standard deviation was calculated to be 0.11, indicating a low range of difficulty. Item 3 (D−, P−, and R−) was the easiest item type, followed by items 1 (D−, P=, R−) and 4 (D+, P−, R−). Item 5 (D−, P=, R+) was the most difficult item type, followed by item 8 (D+, P−, R+). These results suggest that items with lower imposition (R−) tended to be easier than those with higher imposition (R+).

To sum up, the MFRM analysis results suggested that the eight speech acts can be ordered by ascending difficulty as follows: Thank, Request, Suggestion, Disagreement, Invitation, Refusal, Offer, and Apology. The results also suggested a potential effect of imposition on learners' oral speech act production performance.

Research question 2: Effects of the interaction between each of the three contextual variables and speech act on the difficulty of oral speech act production tasks

Three separate two-way ANOVAs were conducted to investigate the effects of the interaction between each contextual factor and speech act on the difficulty of oral speech act production tasks. The four items for Disagreement were excluded from these analyses because not all values for all three variables were represented among these items as a result of the removal of four Disagreement items. The Levene test indicated that the assumption of equal variance across groups was violated ($p < 0.05$). However, the ANOVA *F* test has been shown to be robust if the sample is large, the group sizes are equal, and the largest group standard deviation is not larger than twice the smallest group standard deviation (e.g., Agresti et al., 2017). Given that our dataset met these criteria, we proceeded with the two-way ANOVAs followed by pairwise comparisons using the Tamhane's T2 *post hoc* test, which does not assume equal variances across groups.

Social distance

As shown in Table 2, the main effect of speech act was statistically significant with a large effect size [$F(6,153) = 68.243$, $p = 0.000$, $\eta_p^2 = 0.270$], but the main effect of social distance was insignificant [$F(1,158) = 0.316$, $p = 0.574$, $\eta_p^2 = 000$]. The interaction effect between the two factors was significant with a medium effect size [$F(1,158) = 12.127$, $p = 0.000$, $\eta_p^2 = 0.062$]. Pairwise comparisons revealed that, compared to items with the D+ feature, those with the D-feature were significantly easier for Offer and Request but significantly harder for Suggestion and Thank. These results are also visualized in Figure 2.

Power

As shown in Table 3, the main effect of speech act was statistically significant with a large effect size [$F(6,153) = 65.843$, $p = 0.000$, $\eta_p^2 = 0.263$], but the main effect of power was insignificant [$F(1,158) = 1.986$, $p = 0.159$, $\eta_p^2 = 0.002$]. The interaction effect between the two factors was significant with a medium effect size [$F(1,158) = 23.575$, $p = 0.000$, $\eta_p^2 = 0.113$]. Pairwise comparisons revealed that, compared with items with the P= feature, those with the P-feature were significantly easier for Offer and Suggestion but significantly harder for Refusal. These results are also visualized in Figure 3.

Rank of imposition

As shown in Table 4, the main effects of speech act [$F(6,153) = 63.918$, $p = 0.000$, $\eta_p^2 = 0.257$] and Imposition [$F(6,153) = 39.300$, $p = 0.000$, $\eta_p^2 = 0.034$] were both statistically significant, with large and small effect sizes, respectively. The interaction effect between the factors was also statistically significant with a medium effect size [$F(6,153) = 23.635$, $p = 0.000$, $\eta_p^2 = 0.114$]. Pairwise comparisons revealed that, compared with items with the R+ feature, those with the R-feature were significantly easier for Offer, Request, and Suggestion but significantly harder for Refusal. These results are also visualized in Figure 4.

Discussion

ODCTs are a special type of oral assessment that elicit one-sided responses in hypothesized conversations. Following the suggestion by Fulcher and Reiter (2003), we included both pragmatic categories (i.e., the eight speech acts) and cultural factors (i.e., the combinations of the three social variables in different scenarios) in developing ODCT tasks in the current study. The analysis of the appropriateness ratings of our participants' responses to the ODCT items revealed several substantive findings. First, the MFRM analysis showed that the eight speech acts investigated can be ranked in ascending order of difficulty for Chinese EFL learners as follows: Thank, Request, Suggestion, Disagreement, Invitation, Refusal, Offer, and Apology. Second, the two-way ANOVAs revealed significant main effects of speech act and rank of imposition (R), but not of power (P) and social distance (D). These analyses also revealed significant interaction effects between speech act and each of

TABLE 2 Comparison of mean task performance by speech act and social distance.

Speech act	*N*	Mean/SD		Pairwise comparisons	Analysis of variance				
		D−	D+	*p*		df	*F*	*p*	η_p^2
Apology	80	2.788/0.63	2.728/0.73	0.581	Speech act	6	68.243	0.000	0.270
Invitation	80	3.024/0.51	3.155/0.53	0.114	Social distance	1	0.316	0.574	0.000
Offer	80	3.203/0.59	2.775/0.67	0.000	Interaction	6	12.127	0.000	0.062
Refusal	80	3.123/0.48	2.963/0.65	0.080					
Request	80	3.798/0.57	3.444/0.50	0.000					
Suggestion	80	3.188/0.49	3.580/0.54	0.000					
Thank	80	3.662/0.57	4.003/0.61	0.000					

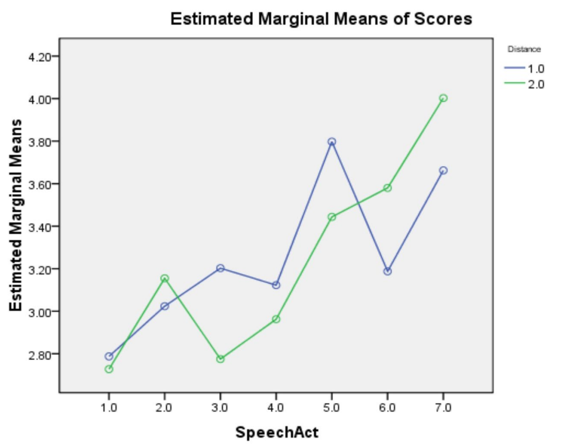


FIGURE 2
Profile plots for the interaction between speech act and social distance. Speech act codes: 1=Apology; 2=Invitation; 3=Offer; 4=Refusal; 5=Request; 6=Suggestion; 7=Thank. Social distance codes: 1=D−; 2=D+.

the three contextual variables, confirming the importance of including both pragmatic categories and cultural factors in ODCT task design (Fulcher and Reiter, 2003). We discuss our findings on the relative difficulty of the tasks for different speech acts and the interaction effects between speech act and the three contextual variables below.

Difficulty of ODCTs for different speech acts

Previous findings on the relative difficulty of pragmatic tasks on different speech acts are limited and inconsistent. In testing learners' pragmatic knowledge of three speech acts: apology, request, and refusal, Hudson (2001) found that apologies were slightly easier than requests and refusals for Japanese ESL learners, which was echoed by Roever's pragmatics test of ESL/EFL learners with diverse language background (Roever, 2004). Hudson accounted for this difference with the explanation that apologies tended to be more formulaic than the other two speech acts and attributed the absence of other difficulty differences to the homogeneity of the participants' proficiency level. Using data from Ahn (2005) on L1 English learners of Korean as a foreign language (KFL) at diverse proficiency levels, Brown (2008) and Brown and Ahn (2011) reported that the average ratings of apologies, requests, and refusals were comparable. Liu (2006), however, found apologies to be consistently more difficult across three test formats (MDCT, DSAT, and WDCT) than requests for Chinese EFL learners at diverse proficiency levels. The different findings pertaining to the difficulty of apologies relative to other speech acts on learners with different L1 backgrounds and the agreement between Liu's finding and our finding that apologies were harder than requests for Chinese EFL learners suggest a potential effect of the learners' L1 cultural background on speech act production task difficulty. This conclusion aligns with the prediction that the culture-specific nature of pragmatic ability may give rise to unique challenges for learning L2 speech acts (Taguchi, 2012). Youn and Brown's (2013) finding that pragmatics test item difficulty remained consistent across two different studies by Ahn (2005) and Youn (2008) on two different groups of L1 English KFL learners also offers support for this conclusion, as it suggests more consistency of task difficulty among learners of the same L1 background.

Apology was found to be the most difficult speech act for Chinese EFL learners in the present study. A closer examination of the production data revealed that our participants had no difficulty in using the formulaic head act strategy (i.e., *I'm sorry*), but many struggled with producing appropriate supporting moves. As illustrated in Example 1, many students followed *I'm sorry* with an explanation that the cause was accidental, often with the structure "didn't … on purpose", likely translated from the Chinese expression *búshì gùyì de* (不是故意的, "didn't do it on purpose"), which is commonly used in apologies in Chinese. This strategy, however, was not considered conventional by the L1 English raters.

(1)	a. I'm sorry. I did it by accident.
	b. I'm so sorry. I did not do it on purpose. I promise it will not happen again.
	c. I'm so sorry. I did not knock over the cup on purpose.

In addition, some participants provided grounders that were considered by the L1 English raters to be too casual to the extent that they jeopardize the sincerity of the apology, as illustrated by Example 2:

(2)	a. Sorry, Miss May, I had something important to do just now. So I'm coming late.
	b. Sorry, Miss May, I had something on the way. I'm very sorry.
	c. Sorry, I have something urgent. Please forgive me.

TABLE 3 Comparison of mean task performance by speech act and power.

Speech act	*N*	Mean/SD		Pairwise comparisons	Analysis of variance				
		P–	P=	*p*		df	*F*	*p*	η_p^2
Apology	80	2.753/0.66	2.752/0.72	0.991	Speech act	6	65.843	0.000	0.263
Invitation	80	3.125/0.59	3.056/0.53	0.434	Power	1	1.986	0.574	0.002
Offer	80	3.100/0.64	2.873/0.71	0.036	Interaction	6	23.575	0.000	0.113
Refusal	80	2.623/0.54	3.458/0.55	0.000					
Request	80	3.630/0.52	3.614/0.58	0.852					
Suggestion	80	3.735/0.54	3.027/0.52	0.000					
Thank	80	3.917/0.61	3.751/0.58	0.081					

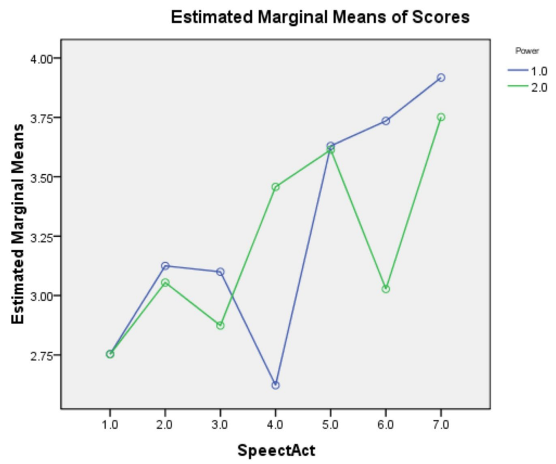


FIGURE 3
Profile plots for the interaction between speech act and power. Speech act codes: 1=Apology; 2=Invitation; 3=Offer; 4=Refusal; 5=Request; 6=Suggestion; 7=Thank. Power codes: 1=P–; 2=*p*=.

These grounders also appeared to display an L1 transfer effect, as the expressions *yǒudiǎn shì* (有点事, "have something") and *yǒudiǎn jíshì* (有点急事, "have something urgent") are commonly used excuses in apologies in Chinese. These examples support Blum-Kulka's (1982) claim that L2 learners' speech act production is often influenced by pragmatic transfer from their L1 and that negative transfer may result in pragmatic failures and cross-cultural communication breakdowns.

Offer was found to be the second most difficult speech act for Chinese EFL learners. Previous research on L2 learners' realization of offers is scant. As offers have a directive nature in that they involve the speaker attempting to persuade the hearer to accept the offer in question, the use of head act strategies for offers resembles that for requests. However, a major difference between offers and requests is that offers presumably benefit the hearer while requests impose on the hearer. As such, the use of direct strategies may be considered more acceptable for offers than for requests, which is also the case in Chinese. Additionally, it has been noted that in some cultures, Chinese included, an offer is not considered sincere until it has been reiterated (Barron, 2003). As noted by the L1 English raters, the participants' offers received low ratings primarily because they sometimes sounded overly direct and eager to help to the extent that the hearer might feel being imposed on. In Example 3, one participant offered to help a sick classmate with the use of *must*, which the raters felt was overly strong.

(3)	You are sick. I must take you to the hospital.

Refusals were found to be the third most difficult among the eight speech acts. As a typical face-threatening speech act (Brown and Levinson, 1987), refusals have been recognized as a major cross-cultural obstacle (Babai Shishavan and Sharifian, 2016). Ekiert et al. (2018) reported that advanced L1 Japanese and Spanish ESL learners achieved comparable pragmatic appropriacy for refusals, complaints, and advice, but lower proficiency ESL learners with those L1 backgrounds achieved lower pragmatic appropriacy for refusals than for complaints and advice. Our results showed that refusals were harder than suggestions for Chinese EFL learners. Refusal was again found more difficult than most speech acts in the present study. Previous research found that grounder and regret strategies are the most frequently used for refusals by Greek foreign language learners (Bella, 2014) as well as by Chinese learners of English in both at-home and study aboard contexts (Ren, 2015). A close analysis of the participants' production data indicated that they relied heavily on expressions of gratitude but rarely used empathetic or positive statements, as illustrated in the participant's response to the item on refusing a chance to take part in a speech contest in Example 4. One L1 English rater commented that a positive statement before the refusal (e.g., *I know the speech contest is a great opportunity for me to practice my English*, *but…*) would improve its pragmatic appropriacy.

(4)	I'm sorry. I do not think I can take part in it. Thank you for your trust.

Request, Suggestion, Disagreement, and Invitation were found to be relatively easier, and Thank was found to be the easiest speech act. The participants demonstrated good familiarity with the pragmatic formulas associated with these speech acts, and they used the most formulaic expressions for Thank among all speech acts. The higher frequency of use of these speech acts in the university setting in general and in the language classroom in particular may have also contributed to the lower difficulty of these speech acts.

The interaction effects between speech act and The three contextual variables

The difficulty of the ODCT items was found to be affected by the interaction between speech act and each of the three contextual variables. This finding is consistent with Taguchi's (2007) finding that social factors may make certain types of situations for pragmatic tasks more demanding than others. The finding also supports Fulcher and Reiter's (2003) claim that different contextual variables may have distinct effects on particular speech acts.

TABLE 4 Comparison of mean task performance by speech act and rank of imposition.

Speech act	*N*	Mean/SD		Pairwise comparisons	Analysis of variance				
		R–	R+	*p*		df	*F*	*p*	η_p^2
Apology	80	2.827/0.84	2.679/0.57	0.193	Speech act	6	63.918	0.000	0.257
Invitation	80	3.079/0.69	3.106/0.53	0.763	Imposition	1	39.300	0.000	0.034
Offer	80	3.121/0.64	2.857/0.59	0.008	Interaction	6	23.635	0.000	0.114
Refusal	80	2.829/0.64	3.256/0.51	0.000					
Request	80	3.822/0.50	3.422/0.57	0.000					
Suggestion	80	3.928/0.53	2.835/0.53	0.000					
Thank	80	3.894/0.69	3.762/0.58	0.195					

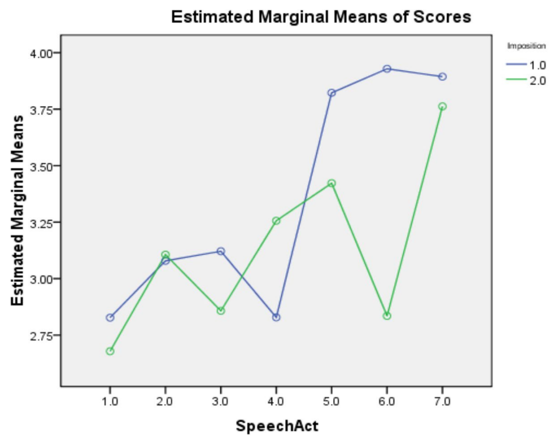


FIGURE 4
Profile plots for the interaction between speech act and rank of imposition. Speech act codes: 1=Apology; 2=Invitation; 3=Offer; 4=Refusal; 5=Request; 6=Suggestion; 7=Thank. Imposition codes: 1=R–; 2=R+.

Social distance exhibited different effects on different speech acts. Compared with items with the D+ feature, items with the D-feature were significantly easier for Offer and Request but significantly harder than Suggestion and Thank. These results indicate that the participants produced more appropriate offers and requests to familiar hearers but more appropriate suggestions and thanks to unfamiliar hearers. A close analysis of the learner production data suggested that the participants tended to use similar types of formulaic strategies for items with D+ and D– features. For example, they frequently used "Would you like to …" for Suggestion and "Thank you very much" for Thank, which were considered more appropriate for unfamiliar hearers (D+) but sometimes overly polite for very familiar peers (D–). Li (2010), for example, indicated that native Australian students tended to use ability statements such as "You can" to realize suggestions in D-scenarios.

With respect to power, items with the P-feature were significantly easier for Offer and Suggestion, while items with the P= feature were significantly easier for Refusal. These results indicate that the participants produced more appropriate offers and suggestions to hears with more power but more appropriate refusals to hears with equal power. These results may not be surprising, as they align with the common understanding that it is easier to make an offer to than to refuse someone with more power in the university setting (e.g., a teacher) in the Chinese culture. Overall, our participants demonstrated some struggle with consistently deploying politeness strategies appropriate for these speech acts to hearers with different power status, sometimes showing negative pragmatic transfer from Chinese. For example, they tended to extend offers to teachers using polite, indirect forms and to their peers using highly direct forms (e.g., *Come to dinner with me*). While such direct strategies for making offers to peers are commonly used to show sincerity and hospitality or to preserve the speaker's positive face in the Chinese culture, they may sound intruding in western cultures where the hearer prefers to be left alone (Gu, 1990; Mao, 1994).

Imposition was the only contextual variable that showed a significant main effect, with items with the R+ feature showing a higher level of difficulty than those with the R– feature overall. Hudson (2001) and Liu (2006, 2007) also reported that R+ items received lower scores than R– items across multiple test methods, although they did not examine the interaction between speech act and imposition. Our analysis showed that, compared to R– items, R+ items were significantly harder for Offer, Request, and Suggestion, significantly easier for Refusal, and comparably difficult for other speech acts. While these findings are not necessarily surprising (e.g., as the degree of imposition increases, requests become harder while refusals become easier), they nonetheless provide evidence for the need and usefulness to look at the interaction effect between speech act and individual situational variables.

Limitations

The current study has several limitations that can be addressed in future research. First, while we included participants with diverse levels of English proficiency in the study to have a heterogenous sample, we did not systematically examine the effect of proficiency on the difficulty of speech act production tasks, a topic that can be useful to investigate in future research. Second, our analysis focused on the appropriateness ratings of the participants' responses only, and it may be useful for future research to consider learners' perceptions of task difficulty and to qualitatively explore the reasons why learners see certain speech acts and contextual variable combinations are more difficulty than others. Third, we employed two raters in the current study only, and greater reliability in the judgments of language learners' pragmatic performance could be achieved by using a larger pool of raters. Fourth, a certain degree of interference existed in the data collection phase as oral samples of a group of participants were elicited simultaneously in a language lab, which can be avoided by applying headphones or collecting data separately. Finally, given that the difficulty of oral speech act production tasks may vary by L1 cultural background, the order of relative difficulty established in the current study for the eight speech acts may not be directly applicable to English learners of other L1 backgrounds. Future research can investigate how the order of relative difficulty may vary by L1 background by including participants from diverse L1 backgrounds.

Conclusion

This study examined the relative difficulty of oral speech act production tasks involving eight types of speech acts for Chinese EFL learners and the effects of three situational variables, namely, power, social distance, and imposition, on such difficulty. A Many-facet Rasch Measurement analysis suggested that the eight speech acts can be ordered by ascending difficulty as follows: Thank, Request, Suggestion, Disagreement, Invitation, Refusal, Offer, and Apology. Significant effects on performance scores were found for the interaction between each of the three contextual variables and speech act, and the specific effects observed varied by speech act. Learner responses also reflected influences of their L1 cultural background. Our findings on the relative difficulty of oral production tasks involving different speech acts and the effects of relevant situational variables on such difficulty have useful implications for L2 pragmatics test design.

Our findings have useful implications for L2 pragmatics testing. Given that different speech act types are not equally difficult to EFL learners, it is important to not generalize results from testing the realization of a particular speech act or a small set of speech acts to the learners' pragmatic ability in performing other speech acts. Furthermore, given the effects of the situational variables on the task difficulty for different speech acts, it is critical to test learners' speech act production with different combinations of contextual variables. Finally, the evaluation of task difficulty in L2 pragmatics assessment need to take learners' L1 background into account.

Our findings also have useful implications for L2 pragmatics pedagogy in the Chinese EFL context. From a task-based language teaching perspective, as advocated by Taguchi and Kim (2018), the relative difficulty of tasks provides highly useful information for task selection and task sequencing in teaching L2 pragmatics. The rank of difficulty estimates of the pragmatic tasks for different speech acts observed in the present study can be used to inform the order in which the speech acts are introduced and the allocation of classroom time to different speech acts in L2 pragmatics pedagogy. Our findings regarding the effects of the three contextual factors on the task difficulty for different speech acts can be used to inform the design of different situation types in teaching speech acts. Our findings further showed the need to help Chinese EFL learners become more sensitive to different situation types and to avoid negative L1 transfer in their choices of speech act strategies. To this end, it will be especially helpful to deploy learning activities designed to help learners become more aware of the pragmatic appropriacy of different speech act strategies in different situation types as well as differences between the pragmatic appropriacy of different speech act realizations in the learners' L1 and the target language.

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

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

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.

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.2023.1096399/full#supplementary-material

References

Agresti, A., Franklin, C. A., and Klingenberg, B. (2017). *Statistics: The art and science of learning from data*. Boston: Pearson.

Ahn, R. C. (2005). Five measures of interlanguage pragmatics in KFL (Korean as a foreign language) learners. PhD thesis. University of Hawai'i at Manoa, USA.

Babai Shishavan, H., and Sharifian, F. (2016). The refusal speech act in a cross-cultural perspective: a study of Iranian English-language learners and Anglo-Australian speakers. *Lang. Commun.* 47, 75–88. doi: 10.1016/j.langcom.2016.01.001

Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.

Bachman, L. F., and Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests* (Vol. *1*). Oxford: Oxford University Press.

Bachman, L. F., and Palmer, A. S. (2010). *Language assessment in practice*. Oxford: Oxford University Press.

Barron, A. (2003). *Acquisition in interlanguage pragmatics: Learning how to do things with words in a study abroad context*. Amsterdam: John Benjamins.

Bella, S. (2014). Developing the ability to refuse: a cross-sectional study of Greek FL refusals. *J. Pragmat.* 61, 35–62. doi: 10.1016/j.pragma.2013.11.015

Blum-Kulka, S. (1982). Learning to say what you mean in a second language: a study of the speech act performance of learners of Hebrew as a second language. *Appl. Linguis.* 3, 29–59. doi: 10.1093/applin/3.1.29

Blum-Kulka, S., House, J., and Kasper, G. (Eds.). (1989). *Cross-cultural pragmatics: Requests and apologies*. Norwood, NJ: Ablex Publishing Corporation.

Brown, J. D. (2001). "Six types of pragmatics tests in two different contexts," in *Pragmatics in language teaching*. eds. G. Kasper and K. Rose (New York: Cambridge University Press), 301–325.

Brown, J. D. (2008). "Raters, functions, item types and dependability of L2 pragmatics tests," in *Investigating pragmatics in foreign language learning, teaching and testing*. eds. E. A. Soler and A. Martinez-Flor (Clevedon: Multilingual Matters), 224–248.

Brown, J. D., and Ahn, R. C. (2011). Variables that affect the dependability of L2 pragmatics tests. *J. Pragmat.* 43, 198–217. doi: 10.1016/j.pragma.2010.07.026

Brown, P., and Levinson, S. C. (1987). *Politeness: Some universals in language usage*. Cambridge: Cambridge University Press.

Chen, Y., and Liu, J. (2016). Constructing a scale to assess L2 written speech act performance: WDCT and E-mail tasks. *Lang. Assess. Q.* 13, 231–250. doi: 10.1080/15434303.2016.1213844

Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.

Cohen, A. D., and Olshtain, E. (1981). Developing a measure of sociocultural competence: the case of apology. *Lang. Learn.* 31, 113–134. doi: 10.1111/j.1467-1770.1981.tb01375.x

Ekiert, M., Lampropoulou, S., Révész, A., and Torgersen, E. (2018). "The effects of task type and L2 proficiency on discourse appropriacy in oral task performance," in *Task-based approaches to teaching and assessing pragmatics*. eds. N. Taguchi and Y. Kim (John Benjamins: Amsterdam/New York), 247–264.

Elder, C., Iwashita, N., and McNamara, T. (2002). Estimating the difficulty of oral proficiency tasks: what does the test-taker have to offer? *Lang. Test.* 19, 347–368. doi: 10.1191/0265532202lt235oa

Fulcher, G. (1996). Testing tasks: issues in task design and the group oral. *Lang. Test.* 13, 23–51. doi: 10.1177/026553229601300103

Fulcher, G., and Reiter, R. M. (2003). Task difficulty in speaking tests. *Lang. Test.* 20, 321–344. doi: 10.1191/0265532203lt259oa

Galaczi, E., and Taylor, L. (2018). Interactional competence: Conceptualisations, Operationalisations, and outstanding questions. *Lang. Assess. Q.* 15, 219–236. doi: 10.1080/15434303.2018.1453816

Gass, S. M., and Neu, J. (Eds.) (1996). *Speech acts across cultures: Challenges to communication in a second language*. Berlin: Mouton de Gruyter.

Golato, A. (2003). Studying compliment responses: a comparison of DCTs and recordings of naturally occurring talk. *Appl. Linguis.* 24, 90–121. doi: 10.1093/applin/24.1.90

Grabowski, K. C. (2009). *Investigating the construct validity of a test designed to measure grammatical and pragmatic knowledge in the context of speaking (unpublished dissertation)*. Columbia University, New York.

Grabowski, K. (2013). "Investigating the construct validity of a role-play test designed to measure grammatical and pragmatic knowledge at multiple proficiency levels," in *Assessing second language pragmatics*. eds. S. Ross and G. Kasper (New York: Palgrave Macmillan), 149–171.

Gu, Y. (1990). Politeness phenomena in modern Chinese. *J. Pragmat.* 14, 237–257. doi: 10.1016/0378-2166(90)90082-O

Hudson, T. (2001). "Indicators for pragmatics instruction: some quantitative tools," in *Pragmatics in language teaching*. eds. G. Kasper and K. Rose (Cambridge: Cambridge uniersity press)

Hudson, T., Detmer, E., and Brown, J. D. (1992). *A framework for testing cross-cultural pragmatics* (Vol. *2*). Honolulu: University of Hawaii, Second Language Teaching and Curriculum Center.

Hudson, T., Detmer, E., and Brown, J. D. (1995). *Developing prototypic measures of cross-cultural pragmatics* (Vol. *7*). Honolulu: University of Hawaii, Second Language Teaching and Curriculum Center.

Ikeda, N. (2017). Measuring L2 oral pragmatic abilities for use in social contexts: Development and validation of an assessment instrument for L2 pragmatics performance in university settings. Unpublished PhD thesis. University of Melbourne, Australia.

Iwashita, N., McNamara, T., and Catherine, E. (2001). Can we predict task difficulty in an oral proficiency test? Exploring the potential of an information-processing approach to task design. *Lang. Learn.* 51, 401–436. doi: 10.1111/0023-8333.00160

Krish, P., and May, O. C. (2020). A case study of L1 interference in speech acts among Chinese L2 students. *3L. Lang. Linguist. Literature* 26, 106–118. doi: 10.17576/3L-2020-2601-08

Li, E. S. (2010). Making suggestions: a contrastive study of young Hong Kong and Australian students. *J. Pragmat.* 42, 598–616. doi: 10.1016/j.pragma.2009.07.014

Li, S., Taguchi, N., and Xiao, F. (2019). Variations in rating scale functioning in assessing speech act production in L2 Chinese. *Lang. Assess. Q.* 16, 271–293. doi: 10.1080/15434303.2019.1648473

Linacre, M. (2013). *A user's guide to FACETS Rasch-model computer programs (Version 3.71.0)*. Chicago, IL. Winsteps.Com.

Liu, J. (2006). *Measuring interlanguage pragmatic knowledge of EFL learners*. Frankfurt am Main: Peter Lang.

Liu, J. (2007). Developing a pragmatics test for Chinese EFL learners. *Lang. Test.* 24, 391–415. doi: 10.1177/0265532207077206

Mao, L. R. (1994). Beyond politeness theory: 'face' revisited and renewed. *J. Pragmat.* 21, 451–486. doi: 10.1016/0378-2166(94)90025-6

Martínez-Flor, A., and Usó-Juan, E. (Eds.). (2010). *Speech act performance: Theoretical, empirical, and methodological issues*. Amsterdam: John Benjamins.

McNamara, T., and Knoch, U. (2012). The Rasch wars: the emergence of Rasch measurement in language testing. *Lang. Test.* 29, 555–4576. doi: 10.1177/0265532211430367

Purpura, J. E. (2004). *Assessing grammar*. Cambridge: Cambridge University Press.

Ren, W. (2015). *L2 pragmatic development in study abroad contexts*. Berlin: Peter Lang.

Ren, W. (2022). *Second language pragmatics*. Cambridge: Cambridge University Press.

Robinson, P. (2001). Task complexity, task difficulty and task production: exploring interactions in a componential framework. *Appl. Linguis.* 22, 27–57. doi: 10.1093/applin/22.1.27

Roever, C. (2004). "Difficulty and practicality in tests of interlanguage pragmatics," in *Studying speaking to inform language learning*. eds. D. Boxer and A. Cohen (Clevedon: Multilingual Matters Ltd), 283–301.

Roever, C. (2005). *Testing ESL pragmatics*. Frankfurt-am-Main: Peter Lang.

Roever, C. (2006). Validation of a web-based test of ESL pragmalinguistics. *Lang. Test.* 23, 229–256. doi: 10.1191/0265532206lt329oa

Roever, C. (2007). DIF in the assessment of second language pragmatics. *Lang. Assess. Q.* 4, 165–189. doi: 10.1080/15434300701375733

Roever, C. (2011). Testing of second language pragmatics: past and future. *Lang. Test.* 28, 463–481. doi: 10.1177/0265532210394633

Ross, S. J., and Kasper, G. (Eds.) (2013). *Assessing second language pragmatics*. Hampshire, UK: Palgrave Macmillan.

Searle, J. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge: Cambridge University Press.

Searle, J. (1975). "Indirect speech acts," in *Syntax and semantics 3: Speech acts*. eds. P. Cole and J. Morgan (New York: Academic Press), 59–82.

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press.

Taguchi, N. (2007). Task difficulty in oral speech act production. *Appl. Linguis.* 28, 113–135. doi: 10.1093/applin/aml051

Taguchi, N. (2012). *Context, individual differences and pragmatic competence*. Bristol: Multilingual Matters.

Taguchi, N., and Kim, Y. (2018). "Task-based approaches to teaching and assessing pragmatics: an overview," in *Task-based approaches to teaching and assessing pragmatics*. eds. N. Taguchi and Y. Kim (Amsterdam/New York: John Benjamins), 1–26.

Yamashita, S. (1996). *Six measures of JSL pragmatics (technical report #14)*. Honolulu: University of Hawaii, Second Language Teaching and Curriculum Center.

Yoshitake, S. S. (1997). *Measuring interlanguage pragmatic competence of Japanese students of English as a foreign language: A multi-test framework evaluation (unpublished doctoral dissertation)*. Columbia Pacific University, Novata, CA.

Youn, S. J. (2008). Rater variation in paper vs. web-based KFL pragmatic assessment using FACETS analysis. Unpublished munuscript, University of Hawai'i, Honolulu, HI.

Youn, S. J. (2015). Validity argument for assessing L2 pragmatics in interaction using mixed methods. *Lang. Test.* 32, 199–225. doi: 10.1177/0265532214557113

Youn, S. J. (2019). "Assessment in L2 pragmatics" in *The Routledge handbook of second language acquisition and pragmatics*. ed. N. Taguchi (New York: Routledge), 308–321.

Youn, S. J., and Brown, J. D. (2013). "Item difficulty and heritage language learner status in pragmatic tests for Korean as a foreign language," in *Assessing second language pragmatics*. eds. S. J. Ross and G. Kasper (Hampshire, UK: Palgrave Macmillan), 98–123.

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 27 February 2023
DOI 10.3389/fpsyg.2023.924056

Check for updates

OPEN ACCESS

EDITED BY
Maria Garraffa,
University of East Anglia,
United Kingdom

REVIEWED BY
Muhammad Aqeel,
Foundation University Rawalpindi,
Pakistan
Zhichao Xia,
University of Connecticut,
United States

*CORRESPONDENCE
Angel Chan
✉ angel.ws.chan@polyu.edu.hk

SPECIALTY SECTION
This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Psychology

RECEIVED 20 April 2022
ACCEPTED 17 January 2023
PUBLISHED 27 February 2023

CITATION
Chan A, Chen S, Hamdani S, Tse B and Cheng K (2023) Story telling in bilingual Urdu–Cantonese ethnic minority children: Macrostructure and its relation to microstructural linguistic skills.
Front. Psychol. 14:924056.
doi: 10.3389/fpsyg.2023.924056

Story telling in bilingual Urdu–Cantonese ethnic minority children: Macrostructure and its relation to microstructural linguistic skills

Angel Chan[1,2,3,4]*, Sarah Chen[1,3,4], Saboor Hamdani[1,2,4], Bernard Tse[1] and Kelly Cheng[1,2,4]

[1]Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Kowloon, Hong Kong SAR, China, [2]Speech Therapy Unit, The Hong Kong Polytechnic University, Hong Kong, Hong Kong SAR, China, [3]The Hong Kong Polytechnic University – Peking University Research Centre on Chinese Linguistics, Hong Kong, Hong Kong SAR, China, [4]Research Centre for Language, Cognition, and Neuroscience, The Hong Kong Polytechnic University, Hong Kong, Hong Kong SAR, China

Introduction: The ability to produce a well-structured, coherent and informative narrative requires the integration of lexical and grammatical skills at different levels of complexity. Investigating how narrative macrostructure competence is predicted by microstructural linguistic skills is conceptually enlightening; yet there have been very few, if any, studies documenting the associations between macrostructure and microstructure in both languages of the same bilinguals. In this paper we attempt to address this research gap and report on the first empirical study of Urdu-Cantonese bilingual children's narrative abilities, bringing in data from a new language pair that is currently understudied.

Methods: Twenty-four bilinguals (mean age=9.17years) acquiring Urdu as first, family and heritage minority language, and Cantonese as second, school and majority language were assessed *via* Multilingual Assessment Instrument for Narratives (MAIN). We examined these children's macrostructural competence and its relations to microstructural skills in both languages (Urdu and Cantonese). Three macrostructure components were scored as response variables: Story Structure (SS), Story Complexity (SC), Internal State Terms (IST). Four microstructural measures were scored as predictor variables: number of different words (NDW), mean length of Communication Units (MLCU), proportion of grammatical Communication Units (Gproportion), proportion of correct connectives linking the major episodic elements (Cproportion).

Results: In regression analyses, NDW emerged consistently as a positive predictor of SS, SC and IST in both languages. MLCU and NDW were positive predictors of SS in the stronger L1, but NDW was the only positive predictor of SS in L2. By contrast, NDW and an index of syntactic competence (MLCU in L1, but Cproportion in L2) were significant or close-to-significant positive predictors of SC in both languages. NDW was the only positive predictor of IST in both languages. These findings suggested that the relationships between narrative macrostructure and specific microstructural abilities could manifest both similarly and differently between L1 and L2.

Discussion: We discuss the findings by considering the unique nature of each macrostructure component and how each component might be related to specific microstructural linguistic skills. We suggest directions for further research and discuss how the current findings bring deeper implications for educators and clinicians in assessment, pedagogy, and intervention.

KEYWORDS

bilingual ethnic minority children, Cantonese, macrostructure, narrative, Urdu

1. Introduction

Children's linguistic competence in narrative production can be analyzed at two levels: macrostructure and microstructure. Macrostructure refers to a higher-order global organization of a story such as episodic structure and story grammar components (Heilmann et al., 2010). Microstructure involves more local level of language use and a more language-specific analysis of the internal linguistic structure such as lexical items, morphosyntax and connectives used in constructing a coherent narrative production (Gagarina et al., 2016). Although it has been shown that macrostructure and microstructure represent two distinct areas underlying narrative competence, they are not mutually exclusive (Liles et al., 1995). Given that the ability to produce a well-structured, coherent and informative narrative requires the integration of lexical and grammatical skills at different levels of complexity, examining the associations between narrative macrostructure and microstructural linguistic skills is conceptually illuminating. This study aims to investigate how macrostructural competence is predicted by microstructural skills in both languages of a group of bilingual ethnic minority children.

1.1. Analysis of macrostructure and microstructure

There can be more than one way of coding story macrostructure depending on the framework, e.g., Applebee (1978)'s six-levels framework, High-point analysis (Labov, 1972), and episodic analysis (Stein and Glenn, 1979). The commonly used episodic analysis, also the framework adopted in this study, analyzes a story based on story grammar, where story grammar elements/components (e.g., setting, initiating event, internal response, internal plan, attempt, consequence and reaction) constitute the episodic structures of a story. The terminology regarding narrative macrostructure is highly variable in the literature. Studies have used terminologies such as story content, event content, story structure, and story complexity. Due to these variations, we discuss the core concept of macrostructure below, to help readers relate the current study to the earlier studies.

One major dimension is the content structure of a story. Under an episodic analysis, this dimension identifies the macrostructure of a story by evaluating the presence of story grammar elements/components. Because the intentions and events represented by these story grammar elements/components involve logical temporal and causal relationships, being able to verbalize more of these story grammar elements/components would contribute to the coherence and richness of relevant content of a story. It therefore has a quantitative dimension on one hand (counting the number of story grammar elements present), while also contributes to the quality of a story (in terms of richness and coherence of story content) on the other hand.

The second dimension is to consider the complexity of a story concept. This notion is related to how a good story is defined. For instance, Stein and Glenn (1979) argued that a goal-directed action is the necessary basis for a minimal definition of a story. A good story has to make reference to the following dimensions of goal-based action: (i) an animate protagonist that can initiate intentional action, (ii) an explicit statement of the goal or desire of the protagonist (the story grammar component "Goal"), (iii) the overt action(s) performed to serve the protagonist's goal (the story grammar component "Attempt"), and (iv) the outcome(s) as a consequence of the goal being attained or not attained (the story grammar component "Outcome.") Goal-Attempt-Outcome are therefore identified as critical components or dimensions of goal-directed action that form a complete episode. Following this reasoning, Stein (1988) and Westby (2005) constructed decision trees that incorporated these concepts and showed how "a systematic increase in the number of dimensions of a goal-directed action sequence increases the complexity of a story concept" (Stein and Albro, 1997, p: 8). This dimension considers how well Goal-Attempt-Outcome is expressed according to these decision trees. It indicates at which level the child's narrative macrostructure is according to the different levels of structural complexity: (a) are there complete episodes, which include all three Goal-Attempt-Outcome statements; (b) are there abbreviated or incomplete episodes, which include Goal, but lack a complete Goal-Attempt-Outcome structure (i.e., Goal, Goal-Attempt, Goal-Outcome); (c) are there only action or reaction sequences, which do not include Goal (i.e., Attempt-Outcome); and (d) are there are only isolated descriptions (i.e., only Attempt or Outcome statements) or statements reflecting none of the episodic components. Under these considerations, stories can systematically increase in their complexity, with (d) corresponding to the lowest level of complexity, and (a) the highest level of complexity.

The third dimension is to consider the use of internal state terms (IST) to explicitly refer to a character's internal states in a story. IST overlap with terminologies such as mental state language (Bartsch and Wellman, 1995), internal states (Miller and Aloise, 1989), evaluations and inferences (Burns et al., 2012). They provide information about a child's understanding of a character being a mental being having intentionality, goals, mental states, and feelings. They also provide information of a child's understanding of the goals and intentions of characters as a child conceives a character's actions as goal-directed. As such, IST not only draw upon linguistic abilities to verbalize knowledge about intentional actions and mental states of characters, but also theory of mind abilities as a child conceives a character's internal states (see also study two of Reilly et al., 2004 for a socio-cognitive perspective). In story-telling, IST are often linked to story grammar elements such as goals, initiating events, and reactions at the macrostructural level, as a child attempts to structure an episode to include reference to an initiating event that may involve the internal state of a character, which triggers an intentional goal of a character that leads to a goal-directed attempt, which in turn leads an outcome as a consequence of the attempt, and then a character's reaction as a result of the outcome. However, IST are also linked to narrative microstructure, because they require semantic skills to use the appropriate and diverse lexical items to verbalize the internal states, and syntactic skills as IST often involve metalinguistic (e.g., say, ask, etc.) and metacognitive (e.g., decide, believe, etc.) verbs that occur in complex syntactic structures. As such, IST are not always included as a narrative macrostructural index in the literature (e.g., Altman et al., 2016). Studies such as Silliman et al. (2002) considered IST as microstructure elements. Unlike the first two dimensions that consider primarily the episodic structure of a story, IST are closely related to linguistic measures due to their unique close connections to microstructure in addition to macrostructure. Their acquisition is therefore relatively more dependent on language-specific experiences. Since bilingual children may differ in the acquisition of mental terms between the two languages (Silliman et al., 2002; Altman et al., 2016), it is possible to find different degrees of association with linguistic measures in the two languages.

There are also variations between earlier studies in terms of how macrostructure was assessed methodologically. Regarding story content, although story grammars are often used, macrostructure can also be coded differently in terms of measures of main ideas (Bishop and Donlan, 2005), events (O'Neill et al., 2004), information units (Renfrew, 1997), or plot structure (Berman and Slobin, 1994), with a common aim of assessing the amount of relevant information in a story for these latter analyses. For instance, Mäkinen et al. (2014) assessed macrostructural competence by evaluating the amount of relevant information used in a narrative and used the term "event content" to refer to the dimension of story content, although their information units are not entirely identical to story grammar elements. In another study by Karlsen et al. (2016), macrostructure was coded based on the presence of eight plot elements, although they overlap with but are not entirely the same as the conventional story grammar elements. Even when story grammar elements are used as the unit of relevant informational content, there are also methodological variations between studies in terms of how they scored story grammar. For instance, Altman et al. (2016) assessed macrostructure using two parameters. One parameter involves the story content counting only Goal, Attempt and Outcome expressed but not the other story grammar elements. The second parameter concerns the complexity of the narrative in terms of the Goal-Attempt-Outcome episodic elements, where Attempt/Attempt-Outcome sequences received 1 point, incomplete episodes like Goal/Goal-Attempt/Goal-Outcome received 2 points, and complete Goal-Attempt-Outcome received 3 points. Bonifacci et al. (2018) also had two macrostructural parameters, but the scoring methods were different. The first parameter was termed number of macro-structural elements, counting the presence of a wider set of macro-structural elements (Goal, Attempt, Outcome, Mental States, Setting). The second parameter was termed level of macro-structural complexity. Four levels of scores ranging from low to high were identified (0, 1, 2, 3) corresponding to absence, low, medium and high complexity levels, respectively. Specifically, absence refers to absence of at least one Attempt and one Outcome, low refers to presence of both Attempt and Outcome, without verbalizing Goal, medium refers to presence of both Goal and Attempt or both Goal and Outcome as incomplete episodes, and high refers to presence of all three core components Goal-Attempt-Outcome in a complete episode. One unwanted consequence of these methodological differences is that they make it harder to assess the extent of which differences in findings between studies could be attributable to the differences in the methodology used. More preferable would be to make use of a common set of assessment materials and methods that are applicable cross-linguistically and cross-culturally, allowing one to draw comparisons across languages, cultures, and acquisition contexts with more stringent methodological controls (see Multilingual Assessment Instrument for Narratives under Method).

Microstructure, on the other hand, targets the narrator's ability in using the target language to construct a coherent narrative. Microstructure measures typically assess competency in the following dimensions when constructing a narrative: productivity (or story length) and lexis, syntactic complexity, grammaticality, and discourse cohesion. Higher microstructural competence is therefore characterized by a person's ability to use diverse vocabulary, syntactically complex and grammatically well-formed utterances, and greater discourse cohesion to construct a longer narrative. Since microstructure features target language-specific proficiency, they are subject to more variations between languages and between bilinguals and monolinguals, compared to macrostructure (Altman et al., 2016; Gagarina et al., 2016; Rodina, 2017). Due to space constraints, below we introduce those measures that have been commonly examined in narrative studies, particularly those that will be targeted in the current study.

Story length and lexis are often measured by the total number of clauses or Communication Units, total number of words with and without mazes, and the Number of Different Words (NDW). NDW represents the different types of word tokens used in a language sample and has been frequently examined in microstructure. Studies have reported that NDW is a sensitive developmental measure in bilingual acquisition (Uccelli and Paéz, 2007) and a sensitive measure to differentiate between children with and without language disorders in both monolinguals (Auza et al., 2018; Torng and Sah, 2020) and bilinguals (Altman et al., 2016; Gagarina et al., 2019c). While NDW can be seen as a measure of productivity (Justice et al., 2006; Mäkinen et al., 2014), it can be seen as a measure of semantic diversity in other studies (Westerveld and Gillon, 2010; Westerveld and Roberts, 2017), and many others including the current study see it also as a measure of lexical diversity (e.g., Altman et al., 2016; Auza et al., 2018).

Syntactic complexity can be indexed by different measures, for instance, Mean Length of Utterance (MLU), Mean Length of Terminable Units (MLTU), and Mean Length of Communication Units (MLCU). They are computed by the total number of word tokens without mazes divided by the number of the structural units selected, where the base structural unit could be an utterance (for MLU), a terminable unit (for MLTU), or a communication unit (for MLCU). The rationale is that a higher level of syntactic complexity is often indexed by a longer mean length (in words, sometimes in morphemes) of a structural unit in a language sample, especially for younger children. Among these three options, the current study, like others (e.g., Mäkinen et al., 2014; Altman et al., 2016), chose MLCU to facilitate more direct comparisons of results with other research groups. In MLCU, communication unit, defined as an independent clause with its modifiers (Loban, 1976), is taken as the base structural unit. There are also other indices of syntactic complexity, e.g., proportion of subordinating/coordinating constructions, but are beyond the scope of the current study (see Gagarina et al., 2015 for details).

Grammaticality can also be indexed by different measures, for instance, proportion of grammatically well-formed error-free utterance (Bedore et al., 2010; Eisenberg and Guo, 2013), proportion of grammatical Terminable Units (Zwitserlood et al., 2015), and proportion of grammatical Communication Units (Fiestas and Peña, 2004). They are computed by the number of error-free structural units divided by the total number of the structural units, where the base structural unit could be an utterance, a terminable unit, or a communication unit. The rationale is that a higher level of grammatical competence is indexed by a higher proportion of grammatical error-free structural units in a language sample. Among these options, the current study, like others (Fiestas and Peña, 2004), chose proportion of grammatical Communication Units (Gproportion) to facilitate more direct comparisons of results with other research groups. There are also other measures that focused instead on errorful (not error-free) units, e.g., percentage of ungrammatical clauses or sentences (Auza et al., 2018; Sheng et al., 2020), addressing grammatical competence from the reverse side.

Discourse cohesion is defined as "a semantic relation between an element in the text and some other element that is crucial to the interpretation of it" (see the seminal work by Halliday and Hasan, 1976, p: 8). The relation is marked by language-specific devices including conjunctions/connectives, reference, substitution, ellipsis and lexis which contribute to the cohesion of a text. Discourse cohesion has been reported to be a vulnerable domain in L2 acquisition and children with language disorders (Liles et al., 1995; Kupersmitt et al., 2014). Among the various candidate measures of cohesion, the current study focused on the proportion of correctly used connectives linking the major episodic macrostructure components Goal, Attempt, Outcome (Cproportion, see 2.4 under Method for computations). Cproportion was chosen because it captures how the more global macrostructures interact with the more local microstructures in discourse structuring to produce a coherent narrative—a measure that is closely related to the theme of this paper.

1.2. Associations between microstructure and macrostructure

The associations between microstructural and macrostructural abilities in narrative production have been examined in the literature. For instance, Stein and Albro (1997) reported that the longest stories, measured by the number of clauses as an index of productivity at the microstructural level, were also structurally the best developed goal-based stories at the macrostructural level in English-speaking children's narrative production. Soodla and Kikas (2011) examined the relationships between macro- and micro- structural measures in Estonian-speaking children. With the quantity of story information units used as the macro-structure level variable, they reported a high and significant positive correlation between story grammar scores and story length (indexed by number of words), a weak but significant positive correlation between story grammar scores and mean length of communication units, and also a weak but significant negative correlation between story grammar scores and the ratio of grammatical errors (as an index of grammaticality). They advocated that "although macrostructure and microstructure are two distinct underlying areas of narrative competence (Liles et al., 1995), children's performances at both levels are significantly associated and should be taken into consideration in narrative assessment" (Soodla and Kikas, 2011, pp: 231–232). Fernandez (2013) studied Spanish-speaking children and reported that second-order theory of mind scores and number of clauses in narrative production (as a measure of linguistic productivity and complexity) significantly predicted pragmatic language skills, where pragmatic language skill is an aggregate score involving not only the use of internal state terms and story grammar elements but also other measures such as use of performed evaluation devices and connectives in narratives. Mäkinen et al. (2014) studied Finnish-speaking children and reported that the number of different words (but not the number of communication units) predicted event content, at the macrostructural level, of their narrative production.

In studies involving bilingual children, Karlsen et al. (2016) examined predictors of narrative production in first-graders learning L2 Norwegian. Results showed that nonverbal cognitive abilities and home literacy support (indexed by number of children's books at home) predicted story macrostructure; while micro-aspects of narrative production were best predicted by L2 linguistic skills (vocabulary and grammar), home literacy support (indexed by number of children's books at home) and time spent in kindergarten. The study focused only on L2 and did not examine the associations in both L1 and L2 of these bilinguals. More recently, Bonifacci et al. (2018) examined the relationship between micro- and macro- structural competence in the narrative production of monolingual L1 and bilingual L2 Italian-speaking children. Regression analyses showed that MLU was a significant positive predictor of the number of macro-structural elements expressed in monolingual L1 Italian. The model was not significant for the bilingual L2 Italian group. Based on these findings the authors suggested that in monolinguals, narrative macrostructural competence is influenced by the syntactic complexity achieved in the target language; while for bilinguals macrostructural story quality appears to be scarcely influenced by the linguistic structure of the narrative production in L2. This study focused on comparing monolingual L1 versus bilingual L2 Italian and did not examine comparisons of L1 versus L2 in the same bilinguals.

To date there has been little information documenting the associations between macrostructure and microstructure in both languages of the same bilinguals. We do not know much about how the nature of relationship(s) between narrative macrostructure and microstructure might be similar or dissimilar between a bilingual child's L1 versus L2, or between the dominant versus weaker language. This investigation is conceptually important to the field, as it could contribute to our understanding of whether the relationships between macrostructure and microstructure competencies are affected by bilingual factors such as L1/L2 status, dominance patterns between the two languages, language proficiency of the two languages, typological distance between the two languages, and cross-linguistic influences between the two languages.

1.3. Current study

This study aims to add to the existing evidence based on the associations between narrative macrostructure and microstructure competence, in both L1 and L2 of the same bilingual children, bringing in data from a new language pair (Urdu-Cantonese) that is currently understudied. Urdu and Cantonese are typologically diverse languages with low typological proximity and little resemblance/overlap in form-function mappings between the two languages to facilitate positive transfer of L1 linguistic skills to L2. As such, similar patterns in macrostructure-microstructure relationships between two typologically distant languages could reflect the unique nature of particular macrostructure competencies. On the other hand, different patterns in macrostructure-microstructure relationships between two typologically distant languages could reflect the effect of bilingual factors such as L1/L2 status, dominance patterns between the two languages and/or language proficiency of the two languages. Investigating macrostructure-microstructure relationships in both languages of the same bilingual children offers a unique opportunity of a within-subjects design to examine the cross-linguistic manifestation of these possible relations and test these conceptual perspectives.

The study also capitalizes on the methodological and theoretical strengths of MAIN, using the newly adapted Urdu and Cantonese versions of MAIN to conduct dual language assessment (Gagarina et al., 2019a,b; Chan et al., 2020; Hamdani et al., 2020; Kan et al., 2020). Our research questions are:

1. How do the patterns of association between macrostructure and microstructure measures resemble and differ between these three macrostructure dimensions/components (story structure, structural complexity, and internal state terms)?
2. How do the patterns of association between macrostructure and microstructure measures resemble and differ between L1 and L2?

The current study features a group of bilingual ethnic minority children who acquire both languages in conditions of reduced input, a prominent acquisition challenge. These children acquire their heritage language (Urdu) as first and family language and acquire the majority and societal language (Cantonese) as a second or additional language when residing in Hong Kong. They mainly receive input in their first language at home, but not in society or school due to smaller number of speakers and the minority status of their heritage language. Moreover, these families often have restricted social contacts with native speakers of Cantonese, which means the amount of contact with the target language is also reduced. Lacking integration into the community and support from parents, many of whom do not speak Cantonese, these children also face the challenge of acquiring Cantonese under reduced input. They are also associated with lower SES family status (Huttenlocher et al., 2010), which ultimately may affect the quantity and quality of their language learning experiences, since in many studies higher family SES and parental (esp. maternal) level of education have been associated with a child's good language development (e.g., Dollaghan et al., 1999; Armon-Lotem et al., 2011). Examining the relationships between macrostructure and microstructure in these children provide new evidence to consider how these relationships are manifested in a unique acquisition context where these children develop their narrative competence under generally reduced and disadvantaged input conditions in both languages.

2. Methods

2.1. Participants

Twenty-four (13 females) bilingual Urdu-Cantonese children aged between 6 and 12 years old (M = 9.17 years, SD = 1.68 years) attending local primary schools grades one to six in Hong Kong participated. A parental questionnaire was completed to obtain background information on children's demographic data, developmental history and language environment. All participants were considered as typically-developing based on the following justifications: (i) no reported noticeable delay in major developmental milestones in L1, considering both the onset of first word and word-combination; (ii) no reported concerns regarding speech and language development from parents and teachers; and (iii) no suggestive evidence for intellectual disability based on their non-verbal reasoning performance assessed by Raven Progressive Matrices test (standard score, M = 91.5, SD = 12.2, Range = 73–125; Raven et al., 1996).

These children were born in Hong Kong, so their chronological age and length of residence is identical. They come from the Pakistani heritage community acquiring Urdu as their first, family and minority language since birth. They started to be exposed to Cantonese on a more regular and intensive basis since they started schooling around age 3 in local schools using Cantonese as the medium of instruction, acquiring Cantonese as their second, school and majority language.

2.2. Materials, tasks, and procedures

Oral narratives were elicited using Multilingual Assessment Instrument for Narratives (MAIN; Gagarina et al., 2019) adapted to Cantonese (Gagarina et al., 2019a; Chan et al., 2020) and Urdu (Gagarina et al., 2019b; Hamdani et al., 2020). Unlike other narrative assessment tools, MAIN is uniquely designed for dual language assessment in bilinguals. It contains four stories that are parallel in content and structure to assess macrostructure and microstructure abilities and allows systematic comparisons between the two languages of a bilingual child. Moreover, MAIN is cross-linguistically and cross-culturally robust, with over 80 language versions being used in research. The story scripts of these language versions follow the standardized adaptation process (Bohnacker and Gagarina, 2020) to ensure that macrostructural features are the same across languages, while microstructural features like number of words per story (+/−3), number of direct speech sentences are as similar as possible across stories and to the English version.

MAIN also has its theoretical appeal in studying narrative macrostructure. It incorporates ideas from story grammar theory (Mandler, 1979; Stein and Glenn, 1979), causal framework analysis (Trabasso and Nickels, 1992), and the binary story grammar decision tree (Westby, 2005) which consider not only the presence of story grammar elements, but also the causality involved between the main episodic components GAO, and the level of structural complexity and developmental level of narratives. Under a multi-dimensional approach in studying macrostructure, MAIN distinguishes 3 components of macrostructure: Story Structure (SS), Structural Complexity (SC) and Internal State Terms (IST). SS considers the story content organization in terms of counting the number of story grammar elements produced, aligning with the first dimension of evaluating the richness and coherence of the content structure a story. SC considers the complexity of combinations of the main components Goal-Attempt-Outcome in an episodic structure based on the binary decision tree (Westby, 2005), aligning with the second dimension of evaluating the level of structural complexity of a narrative. IST refer to words that express the internal states of a character generally referring to feelings and mental states such as intentions, thoughts, emotions, and reactions of characters in the story, aligning with the third dimension of evaluating the use of language to explicitly refer to the internal states of characters in a narrative production.

Each child completed two stories in Cantonese and another two in Urdu. The order of the language assessed was counterbalanced between participants, where half were assessed in Urdu first and in Cantonese second, while the other half in Cantonese first and in Urdu second. Following MAIN's instructions (Gagarina et al., 2019a,b), the stories Cat and Dog were administered in different languages, and Baby-Birds and Baby-Goats were also administered in different languages. The stories assigned to a particular language were also counterbalanced between participants, allowing the four possible story combinations (Cat-Baby Birds, Cat-Baby Goats, Dog-Baby Birds, Dog-Baby Goats) to be used evenly in equal number of times in both L1 and L2 across children as a group (see "counterbalancing procedures for research purposes" in Gagarina et al., 2019a,b). Moreover, each story was assessed twice, once in telling and then in retelling. Specifically, in telling, the child had to generate and tell a story based on the pictures to the experimenter. Then, in retelling, the child would listen to a pre-recorded model story along with the pictures, and then be expected to retell the story.

2.3. Macrostructure measures

Three macrostructure dimensions/components: story structure, structural complexity, and internal state terms were scored as response variables in both languages.

Story structure (SS). All four stories began with a setting (i.e., time, place), followed by three short episodes, each consisting of an initiating event, Goal, Attempt, Outcome, and a reaction. Each story produced was scored in terms of the number of story grammar elements verbalized. Each element scored for 1 point. Maximum 17 points for each story.

Structural complexity (SC). SC was measured using a 3-point weighting system adapted from Maviş et al. (2016). A sequence without Goal (i.e., Attempt-Outcome) would be given 1 point. An incomplete episode (single Goal, Goal-Attempt or Goal-Outcome) would be given 2 points. A complete episode (Goal-Attempt-Outcome) would be given 3 points. Maximum 9 points for each story.

Internal state terms (IST). The tokens of IST were counted following the MAIN manual. All instances of perceptual state terms (e.g., Cantonese: 睇; Urdu: سنا، دیکھا), physiological state terms (e.g., Cantonese: 肚餓; Urdu: پیاسا، بھوکا), consciousness terms (e.g., Cantonese: 瞓着; Urdu: جاگا، زندہ), emotion terms (e.g., Cantonese: 傷心; Urdu: اداس، خوش), mental verbs (e.g., Cantonese:決定; Urdu: سوچا، چاہتا), linguistic verbs or verbs of saying and telling (e.g., Cantonese: 講; Urdu: بلایا، چینخا) produced were counted in each story.

2.4. Microstructure measures

The following four measures were calculated for each story produced as predictor variables in both languages. Although measures of productivity such as total number of word tokens and number of communication units have been identified as having associations with macrostructure competence in the literature, they were not included in this study. This is because Poisson regression model adopted here (see section 2.5 for justifications) requires the measures to be independent as a pre-requisite. To ensure that the predictor variables are all independent, we kept MLCU but did not include the total number of word tokens and number of communication units because calculation of MLCU was derived from total number of word tokens divided by number of communication units.

Number of different words (NDW). NDW represents the number of different words without mazes, disregarding repeated word tokens. Since words are used in syntactic structures in narratives, NDW can be viewed as reflecting lexico-grammatical competence. NDW has been reported as having significant positive associations with macrostructure competence in Altman et al. (2016) and Mäkinen et al. (2014).

Mean length of Communication Units (MLCU). MLCU was computed by the total number of word tokens without mazes divided by the number of Communication Units. It is a typical measure of syntactic complexity and has been reported as having associations with macrostructure competence (Soodla and Kikas, 2011).

Proportion of grammatical Communication units (Gproportion). Gproportion, a measure of story grammaticality, was calculated by the number of grammatical Communication Units produced divided by the total number of Communication Units. It could be particularly interesting in a weaker L2 context when grammatical (in)competence may be sensitively captured by significantly fewer grammatical sentences. Grammaticality has been examined in Soodla and Kikas (2011), although they found only weak associations with macrostructure.

Proportion of correctly used connectives linking the major episodic components (Cproportion). Cproportion, a measure of narrative cohesion, was calculated by "the number of correctly used connectives divided by the total number of Goal-Attempt-Outcome (or any of the two, i.e., Goal-Attempt, Attempt-Outcome, Goal-Outcome) produced in a story sample." Connectives including additive, causal, sequential and adversative connectives that were used to connect any of the two or all three main episodic components (i.e., Goal and Attempt in Goal-Attempt, Attempt and Outcome in Attempt-Outcome, Goal and Outcome in Goal-Outcome, or Goal and Attempt and Outcome in Goal-Attempt-Outcome were counted as long as they were used correctly). The number of sequences (Attempt-Outcome), incomplete episodes (Goal-Attempt, Goal-Outcome), and complete episodes (Goal-Attempt-Outcome) produced were included in the calculation of the total number of Goal-Attempt-Outcome. Liles et al. (1995) reported that their index of cohesion was moderately related to narrative macrostructure, suggesting that some aspects of cohesion may facilitate a higher-order level of story organization.

2.5. Transcription, scoring and data analysis

The narrative samples were transcribed by a native speaker of the respective language and then cross-checked by one more native speaker to ensure accuracy. Independent scoring of macrostructure and microstructure were carried out by two native speakers of the respective language who were student speech therapists (Cantonese) or research assistants (Urdu) with relevant training. Discrepancies were resolved through discussion with the first author. The Urdu scorings were cross-checked by one more native speaker who is a speech therapist from Pakistan doing her PhD in Hong Kong (third author).

Poisson regression models were chosen because count variables were involved, and they followed a Poisson distribution. A count variable is defined as a variable reflecting the number of occurrence of certain events and it takes on positive discrete values such as 0, 1 and 2 (Coxe et al., 2009). For example, since SS refers to number of story elements expressed, and IST refers to number of internal state terms expressed, they are considered as count variables. Using the standard ordinary least squares (OLS) regression can be potentially problematic because it usually requires the random errors to follow a normal distribution N(0,σ2; Meloun and Militký, 2001). If a count variable is used as an outcome variable in OLS regression, and when the mean of the variable is low, OLS regression models are likely to produce biased results (Gardner et al., 1995).

In the first round of analyses, the data were analyzed with each of the four microstructural measures [Number of Different Words (NDW), Mean Length of Communication Units (MLCU), Proportion of Grammatical Communication Units (Gproportion), Proportion of correctly used Connectives linking the major episodic components (CProportion), Age, Elicitation Mode (telling vs. retelling), Language, and the two-way interaction terms between Language and each of the other predictors as predictor variables, and each of Story Structure (SS), Story Complexity (SC) and Internal State Term (IST) scores as a response variable in a model (i.e., one model for one response variable)]. The interaction terms with Language allow us to identify whether the effect of a predictor variable on a response variable of macrostructural competence was uniform or not across languages. Since we identified several significant two-way interactions with Language, in the second follow-up round of analyses, we therefore ran the analyses separately within each language. In this follow-up round of analyses, we conducted two sets of analyses. One set was simple bivariate correlations between each predictor variable and each macrostructure outcome variable within each language. Spearman's correlation coefficient was used for all the correlations except for Elicitation Mode, where point biserial correlation coefficient was used, as telling/retelling is a categorical variable. The second set of analyses were regression analyses within each language, which entered all predictor variables into a regression that would allow us to consider how a variable reflected its contribution, while taking into account the contribution of all other variables. As such, the predicted shared variance is distributed across all predictor variables. Specifically in these regression analyses, the data were analyzed with each of the four microstructural measures [Number of Different Words (NDW), Mean Length of Communication Units (MLCU)], Proportion of Grammatical Communication Units (Gproportion), Proportion of correctly used Connectives linking the major episodic components (CProportion), Age and the Elicitation Mode (telling vs. retelling) as predictor variables, and each of Story Structure (SS), Story Complexity (SC) and Internal State Term (IST) scores as a response variable in each model (i.e., one model for one response variable in a particular language).

The findings are considered significant with *p* values less than 0.05. Estimated rate ratio represents the expected value of increase (if the estimated coefficient of a variable is positive) or decrease (if the estimated coefficient of a variable is negative) of the assessed macrostructure dimension/component, if a participant were to increase a particular predictor variable by one unit, while holding all other variables in the model constant. For example, if the estimated rate ratio is 1.01 for a one-unit increase of a predictor variable [e.g., Number of Different Words (NDW)] in affecting scores of a response variable (e.g., Story Structure (SS)) and that the estimated coefficient is positive, this means that if the participants were to increase their NDW by one unit, their rate ratio for SS would be expected to increase by a factor of 1.01, while holding all other variables in the model constant. The higher the estimated rate ratio, the greater contribution the respective predictor variable has in the model.

3. Results

3.1. Language dominance

We also examined children's narrative skills in both languages. This gives background knowledge on which language (Urdu vs. Cantonese) could be the dominant language. Table 1 shows each of the seven measures comparing Urdu versus Cantonese. The results were generated by fitting a Poisson regression model for each measure as the dependent variable and language (Urdu vs. Cantonese) as the independent variable to examine if there are any significant differences between languages.

The following measures all consistently indicated that the Urdu scores were significantly or numerically higher than the Cantonese scores, suggesting that these children are largely dominant in their L1 Urdu: Story Structure (SS), Story Complexity (SC), Number of Different Words (NDW), Mean Length of Communication Units (MLCU), Proportion of Grammatical Communication Units (Gproportion), Proportion of correctly used Connectives linking the major episodic components (CProportion). This dominance pattern is consistent with information gathered from the parental questionnaires. Their parents reported in the questionnaires that these children spent more time in an Urdu-speaking environment than in a Cantonese-speaking environment. Specifically, when being asked "On average, how many % of hours per week does your child spend in each language environment (school + home + other environments all included) for Cantonese and for Urdu?," 22 out of 24 parents expressed a higher percentage of weekly exposure in an Urdu-speaking environment than in a Cantonese-speaking environment, with only 2 out of 24 parents expressed an equal percentage of weekly exposure to Urdu and Cantonese. This dominance pattern is also consistent with parental evaluations of their children's language proficiency of the two languages in the questionnaire. Specifically, when being asked "On a scale from 1 (poor) to 7 (excellent), please rate your child's ability to understand/speak spoken Cantonese and Urdu," 17 out of 24 parents gave a higher rating for Urdu than Cantonese in speaking and/or understanding, with only 7 out of 24 parents giving an equal rating for both languages in speaking and understanding. These 7 parents gave either one of the two highest ratings, i.e., a rating of 6 or 7, for both languages. It is also common that these ethnic minority parents are not proficient in Cantonese and therefore these families usually lack practices in supporting literacy in Chinese at home, although our parental questionnaire did not ask specifically about home practices in supporting literacy. There is some suggestive evidence from other responses in the questionnaire though. For instance, 20 out of 24 parents expressed that their child speaks more Urdu than Cantonese at home, suggesting lack of support for Cantonese from the family. The only measure for which Cantonese was stronger than Urdu was the children's Internal State Term (IST) scores. It is possible that IST, compared to Story Structure (SS) and Story Complexity (SC), is more related to the child's language-specific experience (see the introduction section on the unique linguistic nature of IST). This point will be elaborated further in the discussion section.

3.2. Macrostructure dimensions/components and their relationships with microstructure abilities

Tables 2, 3 present the simple bivariate correlation results in Urdu and Cantonese, respectively. The results showed a number of significant positive correlations between specific microstructural competencies such as Number of Different Words (NDW), Mean Length of Communication Units (MLCU), Age, Elicitation Mode and the outcome measures of macrostructural competencies in Story Structure (SS) and Story Complexity (SC), and Internal State Term (IST). Note that if the correlation efficient of Elicitation Mode is positive, it indicates that when the variable *x* takes on the value "1" (retelling), the outcome variable *y* tends to take on higher values compared to when the variable *x* takes on the value "0" (telling).

We next focus on reporting the significant positive predictors measuring microstructural competences of each macrostructure dimension/component in Urdu (L1) and then Cantonese (L2) in the regression analyses, which allow us to consider how a variable reflected its contribution, while taking into account the contribution of all other variables, with the corresponding value of ps, z values and rate ratios presented in Tables 4, 5, respectively. We then comment on Age, Elicitation Mode, and the significant negative predictors (with $p < 0.05$ but negative z-value) collectively across both languages toward the end of this section. Data came from all stories told and retold in Urdu or Cantonese.

In Urdu, findings from Table 4 revealed both similarities and differences between the three macrostructure components in terms of their significant positive predictors. Story Structure (SS) and Story Complexity (SC), and Internal State Term (IST) were similar in terms of having Number of Different Words (NDW) emerged consistently as a highly significant positive predictor of all three macrostructure

TABLE 1 The bilingual children's scores in Urdu versus Cantonese on the seven measures of macrostructure and microstructure abilities.

Nature of linguistic competence	Measures	Urdu *mean (SD) range*	Cantonese *mean (SD) range*	*z* value	*p* value
Macrostructure	SS	9.72 (2.49), 3–14	9.09 (2.79), 3–15	1.98	0.048*
	SC	4.29 (2.12), 0–9	3.92 (2.17), 0–9	1.82	0.068
	IST	6.40 (4.20), 0–24	7.38 (3.96), 1–21	−3.63	0.000***
Expressive lexical or lexico–grammatical	NDW	50.52 (14.07), 25–88	42.57 (14.38), 19–83	11.37	<0.0001***
Syntactic complexity	MLCU	7.75 (1.55), 3.35–13.89	7.41 (15.01), 3.42–12.23	1.19	0.234
Grammaticality	Gproportion	0.822 (0.128), 0.38–1	1.15 (0.351), 0.29–2	−3.26	0.001**
Discourse cohesion	Cproportion	0.367 (0.347), 0–2	0.328 (0.441), 0–2	0.644	0.52

*$p < 0.05$. **$p < 0.01$. ***$p < 0.001$; SS = Story Structure; SC = Story Complexity; IST = Internal State Terms; NDW = Number of Different Words; MLCU = Mean Length of Communication Units; Gproportion = Proportion of grammatical Communication Units; Cproportion = Proportion of correctly used connectives linking the major episodic components.

TABLE 2 Simple bivariate correlations between predictor variables and macrostructure outcome variables in Urdu.

Predictor/ macrostructure	NDW	MLCU	Gproportion	Cproportion	Elicitation mode (telling vs. retelling)	Age
SS	0.653***	0.547***	−0.014	−0.116	0.397***	0.261***
SC	0.517***	0.382***	0.003	−0.117	0.323***	0.212**
IST	0.663***	0.405***	−0.108	−0.065	0.501***	−0.018

**$p < 0.01$. ***$p < 0.001$; SS = Story Structure; SC = Story Complexity; IST = Internal State Terms; NDW = Number of Different Words; MLCU = Mean Length of Communication Units; Gproportion = Proportion of grammatical Communication Units; Cproportion = Proportion of correctly used connectives linking the major episodic components.

TABLE 3 Simple bivariate correlations between predictor variables and macrostructure outcome variables in Cantonese.

Predictor/ macrostructure	NDW	MLCU	Gproportion	Cproportion	Elicitation mode (telling vs. retelling)	Age
SS	0.711***	0.391***	0.046	0.015	0.422***	0.402***
SC	0.438***	0.253***	0.064	0.280***	0.269***	0.082
IST	0.694***	0.273***	0.070	−0.055	0.384***	0.216**

**$p < 0.01$. ***$p < 0.001$; SS = Story Structure; SC = Story Complexity; IST = Internal State Terms; NDW = Number of Different Words; MLCU = Mean Length of Communication Units; Gproportion = Proportion of grammatical Communication Units; Cproportion = Proportion of correctly used connectives linking the major episodic components.

dimensions/components. In addition, Story Structure (SS) and Story Complexity (SC) were relatively more similar in terms of having Mean Length of Communication Units (MLCU) as a significant positive predictor (in SS) or a close-to-significant positive predictor (in SC), with MLCU having the highest rate ratio among all predictors in both SS and SC. IST, on the other hand, differed from SS and SC, as NDW emerged as its only significant positive predictor.

In Cantonese, findings from Table 5 also showed both similarities and differences between the three macrostructure dimensions/ components in terms of the significant positive predictors. Story Structure (SS) and Internal State Term (IST) were similar in terms of having Number of Different Words (NDW) emerged as the only significant positive predictor among the four microstructural measures. In fact, NDW was consistently a highly significant positive predictor of all the three macrostructure dimensions/components. Story Complexity (SC), by contrast, differed from SS and IST, as it was related to an additional measure, Proportion of correctly used Connectives linking the major episodic components (CProportion), which had an even higher rate ratio than NDW (rate ratio of Cproportion = 1.38; rate ratio of NDW = 1.02).

Although it is reasonable to expect age-related improvements in macrostructural competence in both languages, when age was added as a predictor together with the other predictor variables in regression analyses, age did not emerge as the strongest (indexed by the highest rate ratio, or not even a significant positive) predictor relating to macrostructure competence in both languages. For instance, in L1 Urdu, although age was a significant positive predictor of Story Structure (SS) and Story Complexity (SC), its rate ratio was slightly lower than that of Mean Length of Communication Units (MLCU) and Number of Different Words (NDW). Moreover, was even a negative predictor (indicated by its negative z-value) of Internal State Term (IST). Similarly, in L2 Cantonese, although age was a significant positive predictor of SS, its rate ratio was slightly lower than that of NDW. Moreover, age was a non-significant predictor of IST, and was even a close-to-significant negative predictor (indicated by its negative z-value) of SC. This finding suggests that although age is often a cursory measure of length of exposure to a language (especially for L1 in acquisition studies), this relationship could be much less tight when L1 is a minority language and L2 a majority language in bilingual ethnic minority children. Rather, measures of quality and quantity of experience to each language are likely better candidate measures as predictors than age.

Regarding elicitation mode, as expected and consistent with previous studies (Pesco and Kay-Raining Bird, 2016), these children scored significantly higher in a number of macrostructure components in story retelling than telling (Internal State Term (IST) in Urdu: $z = -5.17$, $p < 0.001$; Story Structure (SS) in Cantonese: $z = -4.14$, $p < 0.001$; Story Complexity (SC) in Cantonese: $z = -2.29$, $p = 0.022$;

Internal State Term (IST) in Cantonese: $z = -4.57$, $p < 0.001$), with the benefit of a prior script.

A minor remark is that there were also two reported significant negative predictors among the microstructural measures, namely Proportion of Grammatical Communication Units (Gproportion) and Proportion of correctly used Connectives linking the major episodic components (Cproportion) in predicting Internal State Term (IST) in Urdu, as indicated by their negative z values (see Table 4). Conceptually it is unclear why there is a negative relationship between grammaticality and IST and between discourse cohesion and IST in Urdu. However, it was observed in this dataset that somehow those participants scoring higher in Gproportion and Cproportion happened to score lower in IST in Urdu. Future investigations examining how other measures of grammaticality and discourse cohesion correlate with IST will allow one to further evaluate the robustness of these findings, before attempting to give an explanation.

3.3. Cross-linguistic comparisons in how microstructure abilities predict each macrostructure dimension/component

There were both cross-linguistic similarities and differences attested. Regarding similarities, Number of Different Words (NDW) was consistently a highly significant positive predictor of these children's scores in Story Structure (SS) and Story Complexity (SC), and Internal State Term (IST) in both languages. Moreover, L1 and L2 were similar in IST in terms of having NDW emerged as the only significant positive predictor. Furthermore, L1 and L2 were similar in SC in terms of not only having NDW as a significant positive predictor but also having a grammatical skill-related microstructural measure as a positive predictor with a higher rate ratio, although the two languages also differed specifically with Mean Length of Communication Units (MLCU) emerged as the close-to-significant positive predictor in L1 Urdu, while Proportion of correctly used Connectives linking the major episodic components (CProportion) emerged as the significant positive predictor in L2 Cantonese. There was also a cross-linguistic difference attested in SS, as NDW was the only significant positive predictor emerged among all the four microstructural measures in L2 Cantonese, while in L1 Urdu MLCU and NDW emerged as important positive predictors.

4. Discussion

The current study examined whether and how macrostructure competence in each of the three components [Story Structure (SS) and Story Complexity (SC), and Internal State Term (IST)] and in both languages (L1 Urdu & L2 Cantonese) was (un)related to specific

TABLE 4 Predictor variables and their relations to SS, SC and IST in Urdu.

Macrostructure	Predictor variables	*z* value	*p* value	rate ratio
SS	NDW	4.10	< 0.001***	1.01
	MLCU	2.24	0.025*	1.04
	GProportion	−0.73	0.47	0.87
	CProportion	−1.15	0.25	0.92
	Age	2.84	0.0045**	1.00
	Elicitation Mode (Telling)	−0.95	0.34	0.95
SC	NDW	4.23	<0.001***	1.01
	MLCU	1.71	0.088	1.05
	Gproportion	−1.65	0.10	0.64
	CProportion	−1.87	0.06	0.81
	Age	2.87	0.0041**	1.01
	Elicitation Mode (Telling)	−1.28	0.20	0.90
IST	NDW	9.61	<0.001***	1.03
	MLCU	−1.96	0.05	0.96
	GProportion	−6.21	<0.001***	0.26
	CProportion	−2.60	0.009**	0.78
	Age	−2.16	0.03*	1.00
	Elicitation Mode (Telling)	−5.17	<0.001***	0.70

*$p < 0.05$. **$p < 0.01$. ***$p < 0.001$; SS = Story Structure; SC = Story Complexity; IST = Internal State Terms; NDW = Number of Different Words; MLCU = Mean Length of Communication Units; Gproportion = Proportion of grammatical Communication Units; Cproportion = Proportion of correctly used connectives linking the major episodic components; Elicitation Mode (Telling) = Elicitation Mode with Telling as reference level.

TABLE 5 Predictor variables and their relations to SS, SC and IST in Cantonese.

Macrostructure	Predictor variables	*z* value	*p* value	rate ratio
SS	NDW	5.97	<0.001***	1.01
	MLCU	−0.13	0.90	0.99
	GProportion	−1.41	0.16	0.91
	CProportion	0.41	0.69	1.02
	Age	1.99	0.047*	1.00
	Elicitation Mode (Telling)	−4.14	<0.001***	0.82
SC	NDW	6.02	<0.001***	1.02
	MLCU	0.36	0.72	1.01
	GProportion	−0.25	0.81	0.97
	CProportion	4.07	<0.001***	1.38
	Age	−1.92	0.055	1.00
	Elicitation Mode (Telling)	−2.29	0.022*	0.84
IST	NDW	12.49	<0.001***	1.03
	MLCU	−1.95	0.052	0.96
	GProportion	−1.56	0.12	0.88
	CProportion	−0.79	0.43	0.95
	Age	−0.97	0.33	1.00
	Elicitation Mode (Telling)	−4.57	<0.001***	0.78

*$p < 0.05$. **$p < 0.01$. ***$p < 0.001$; SS = Story Structure; SC = Story Complexity; IST = Internal State Terms; NDW = Number of Different Words; MLCU = Mean Length of Communication Units; Gproportion = Proportion of grammatical Communication Units; Cproportion = Proportion of correctly used connectives linking the major episodic components; Elicitation Mode (Telling) = Elicitation Mode with Telling as reference level.

microstructural linguistic abilities in a group of bilingual ethnic minority children, where Urdu is stronger than Cantonese for many measures.

One robust finding is that Number of Different Words (NDW; rather than age) showed up consistently as a highly significant positive predictor of all three macrostructural dimensions/components in both languages. This result aligns with Mäkinen et al. (2014) reporting NDW as a significant positive predictor of macrostructure measures. Moreover, this result aligns with Altman et al. (2016) reporting significant positive correlations between NDW and their macrostructural complexity measure and between NDW and the use of mental state terms in the narrative production of English-Hebrew bilinguals. The current finding is conceptually justifiable. Macrostructure contributes to the overall meaning of a story and the overall meaning of a story is conveyed through the semantics of the diverse words deployed. In order to express different story grammar elements [Story Structure (SS)], verbalize and combine the core components Goal-Attempt-Outcome to form complete episodes [Story Complexity (SC)], and express internal state terms within a narrative production [Internal State Term (IST)], children have to deploy the relevant words productively in a narrative context as a basis to support verbalization of these three macrostructural dimensions. The convergent evidence from the three macrostructural dimensions/components and from both languages attested corroborates this argument.

There were also partially different profiles in the ways specific microstructural skills related to the three macrostructural dimensions/components in L1 and L2. The pattern of results for each macrostructural dimension/component would therefore be discussed next. Specifically, Story Structure (SS) was related jointly to Number of Different Words (NDW) and Mean Length of Communication Units (MLCU) in the stronger L1, but related only to NDW in the weaker L2. This finding suggests that macrostructural content (indexed by SS) in the stronger L1, characterized by significantly more informative narratives (indexed by the significantly higher SS scores in L1 than L2), was jointly influenced by lexico-grammatical and syntactic competence. In order to include more relevant information units (indexed by more story grammar elements) in a story, this dimension of macrostructural competence (indexed by SS) needs to be supported by not only the ability to use diverse relevant lexical items (indexed by NDW) but also requires the syntactic ability to combine relevant lexical items to form larger information units (indexed by MLCU). On the other hand, macrostructural content (indexed by SS) in these children's weaker L2, characterized by significantly less informative narratives (indexed by significantly lower SS scores in L2 than L1), was related only to the ability to use diverse lexical items (indexed by NDW), and scarcely by the syntactic competence achieved in the target language. One may speculate that when it is about telling an informative story in a bilingual child's weak L2, having adequate, diverse, and relevant vocabularies and being able to deploy them plays a more pivotal role. Our bilingual L1 results align with the monolingual L1 results in Bonifacci et al. (2018). They reported that Mean Length of Utterances (MLU) was a significant positive predictor of the number of macro-structural elements expressed in the narrative production of monolingual L1 Italian children but not in bilingual L2 Italian. The current findings are similar to theirs suggesting that in children's L1, the quality of story macrostructure is influenced by the syntactic complexity (indexed by MLCU here, but indexed by MLU in Bonifacci et al., 2018) achieved in the same language. Whether this pattern of relationship is related to L1 status or proficiency in the dominant language is currently unclear, and will require further research to verify, for instance, comparing bilingual children with L1 as the weaker language.

Looking across both languages to compare Story Structure (SS) versus Story Complexity (SC), SC differed from SS in two respects. First, SC appears to be relatively more independent of general language proficiency. Unlike SS which showed significantly higher scores in L1 than L2 that aligned with the general language dominance pattern of these children, there was no significant difference in SC scores between L1 and L2 despite L1 being a stronger language in general. Conceptually, it is possible that the mental representation and knowledge of the core episodic structure Goal-Attempt-Outcome could be supported by transfer processes that are shared across the two languages, so SC is relatively more independent from linguistic proficiency in the target language, compared to SS. Second, SC was related not only to Number of Different Words (NDW) but jointly and even more related (indexed by a higher rate ratio) to a grammatical skill-related microstructural measure in both L1 and L2. This finding suggests that when one considers another macrostructural dimension in terms of the complexity of a story concept (indexed by SC), which taps into the ability to express and sequence the major components Goal-Attempt-Outcome as complete episodes in a narrative, this macrostructural competence needs to be supported by not only the ability to use diverse relevant vocabularies (indexed by NDW) but also requires some kind of syntactic competence in both L1 and L2. In L1, the syntactic competence to combine relevant words together to form larger information units (indexed by MLCU) was a close-to-significant predictor with the largest rate ratio. In L2, the syntactic competence to use cohesive devices (connectives) to connect the main episodic story grammar elements (indexed by Cproportion) emerged as a significant predictor with the largest rate ratio. This pattern of findings, manifested in both languages, might reflect the unique nature of SC. Recall SC measures children's macrostructure competence in combining the core episodic components. Mean Length of Communication Units (MLCU) implicates children's ability to combine and sequence relevant words to form longer information units; while Proportion of correctly used Connectives linking the major episodic components (CProportion) reflects children's ability to use appropriate cohesive devices like connectives to connect the semantic relations between the main episodic elements Goal-Attempt-Outcome expressed in a story. Functionally, SC, MLCU and Cproportion, by nature, all draw upon children's ability to connect and sequence some information/meaning units within a story. We speculate that this functional overlap observed between SC, MLCU and Cproportion might be relevant when attempting to make sense of the finding that SC was related to MLCU in children's L1 Urdu and Cproportion in children's L2 Cantonese, respectively. As for why MLCU showed up as the close-to-significant positive predictor in L1 but Cproportion showed up as the significant positive predictor in L2 is currently not entirely clear. Our findings showed that SC in a weak L2 context was unrelated to MLCU but more related to Cproportion in these bilinguals. Further research is needed to observe how robust this pattern of findings occurs in other bilinguals' weaker L2.

Compared with Story Structure (SS) and Story Complexity (SC), Internal State Term (IST) is likely most related to language-specific experience given the unique linguistic nature of IST. We observed that IST scores were significantly higher in L2 Cantonese than L1 Urdu, despite L1 being the stronger language in general. Similar findings have been reported by Altman et al. (2016) who reported bilingual English(L1)-Hebrew(L2) children using more mental state terms in their L2, despite 10 of the 19 children being L1 dominant and 9 out of 19 being balanced bilinguals. They attributed this finding to language-specific experiences such as L2 school curriculum and the type of language input in school setting that frequently used mental verbs in L2. Similarly, we speculate that the higher IST scores in L2 Cantonese observed might be due to language-specific experiences during which these children experienced frequent use of ISTs in L2 local school curriculum and setting that uses Cantonese as the medium of instruction. The current findings also revealed cross-linguistic similarities in terms of Number of Different Words (NDW) being the only significant positive predictor of IST scores in both languages. It is conceptually predictable that the production of IST requires lexico-grammatical competence of deploying lexical items such as metalinguistic and metacognitive verbs and emotion words.

A further remark concerns the measure of grammaticality, indexed by Proportion of Grammatical Communication Units (Gproportion). Unlike Number of Different Words (NDW) which emerged consistently as a highly significant positive predictor of Story Structure (SS), Story Complexity (SC) and Internal State Term (IST) in both languages, in contrast, Gproportion consistently did not show up as a significant positive predictor of SS, SC, and IST in both languages. This finding aligns with Soodla and Kikas (2011), in the sense that their measure of grammaticality also showed only weak association with children's macrostructural competence. One might speculate that the ability to produce grammatical communication units in narrative production does not appear to positively contribute to macrostructural competence in these bilinguals.

We make some further remarks about limitations of this study and suggestions for future research. There are likely large variabilities between participants within a relatively small sample size. Future studies with larger samples and a more restricted age range are needed to corroborate the current findings. Moreover, the aim was to examine the relationships between macrostructure and microstructure in these children, and as such we did not set out to include multiple age groups to examine age effects. Instead, since the children willing to participate in this study were of diverse age range, we included age as a predictor among other predictor variables in regression analyses to examine its relative contribution as a predictor of these children's macrostructural competence. Future research could assess different age groups to examine whether the relationships between macrostructural and microstructural competence might vary at different ages. Furthermore, to delimit the scope of investigation, the current study only examined a language pair of L1 and L2 that are diverse with low typological proximity and little resemblance/overlap in form-function mappings between the two languages to facilitate positive transfer. As such, cross-linguistic similarities between L1 and L2 in the relationships between macrostructure and microstructure competencies are more likely reflecting the unique nature of particular macrostructure competencies, as we currently hypothesize, rather than likely due to L1 linguistic skills influencing those in L2. On the other hand, if the language pair involves typologically close languages, then similar patterns in macrostructure-microstructure relationships between the two languages could be due to similarities between L1 and L2 facilitating L1-to-L2 positive transfer of linguistic skills and the unique nature of particular macrostructure competencies. In this case, we may expect even more robust cross-linguistic similarities due to the synergistic effects of both factors. Future research could examine and compare more language pairs (typologically similar vs. typologically diverse) to test these conceptual perspectives in a natural within-participants paradigm within the same bilingual children.

Regarding application values, the current findings and their interpretations also give deeper implications for educators and clinicians in assessment, pedagogy, and intervention. Given that macrostructure is related to specific lexical and grammatical skills at the microstructural level, and certain microstructural skills may be more important than the others in relating to a dimension/component of macrostructural competence depending on L1/L2/proficiency status, one should not assess macrostructure and microstructure as if they are disjoint abilities, but should consider the nature of relationships between them. These perspectives also enlighten pedagogy/intervention, motivating one to discover more about how to foster a child's lexical diversity along with building up her syntactic competence to support each of the three aspects of macrostructural competence in L1 versus L2. This line of inquiry is not restricted to the narrative genre and can be extended in future investigations to other academic discourse genres like exposition and argumentation, giving rise to a compositional construct of micro-properties of language that can predict competence in the macro-properties of language involving different genres of discourse in a child's social-communicative and academic developments.

5. Conclusion

We reported on the first empirical study of Urdu-Cantonese bilingual children's narrative abilities, bringing in data from a new and typologically distant language pair that is currently understudied. We examined macrostructural competence and its relation to specific microstructural linguistic skills in both languages of the same bilingual children, which, to our knowledge, has been under- or un-documented in the current published literature. We found that the significant predictor variables which were related to macrostructure competence were similar and partially different across SS, SC and IST. We discussed the findings by considering the unique nature of each macrostructure dimension/component and how each dimension/component might be supported by or related to specific microstructural linguistic skills. The take-home message is that while the cross-linguistic similarities observed provides convergent evidence in support of the unique nature of a particular macrostructure component, the cross-linguistic differences attested suggest that the possible relations between macrostructural and microstructural competence could vary between languages of a bilingual child that might be attributed to differences in language proficiency and/or L1/L2 status. Future studies assessing different groups of bilinguals with variations in their dominance profiles between L1 and L2 are necessary to tease these factors apart.

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.

Ethics statement

The studies involving human participants were reviewed and approved by the Institutional Review Board for human subjects ethics at the Hong Kong Polytechnic University (reference number:

HSEARS20190813001-01). Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Author contributions

AC designed the study and interpreted the data in consultation with the other coauthors and recruited the participants, supervised native-speaker experimenters and research personnel in data collection, coding, reliability checks and analyses, and wrote a first draft of the paper. SC ran the statistical analyses. Subsequently all authors worked on refining and revising the text. All authors contributed to the article and approved the submitted version.

Funding

This research was supported by a research grant (YW4G; P0014049; PI: AC), awarded by The Hong Kong Polytechnic University. SH is supported by a Hong Kong PhD Fellowship Scheme (HKPFS) awarded by the Hong Kong Research Grants Council.

Acknowledgments

We thank our student speech therapists Barbie Chui, Joyce Lo, Carmen Lai and Vickie Wong for their assistance with data collection and coding. Special thanks to Ute Bohnacker, Natalia Gagarina and the reviewers for their constructive comments on this manuscript.

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.

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

Altman, C., Armon-Lotem, S., Fichman, S., and Walters, J. (2016). Macrostructure, microstructure, and mental state terms in the narratives of English-Hebrew bilingual preschool children with and without specific language impairment. *Appl. Psycholinguist.* 37, 165–193. doi: 10.1017/S0142716415000466

Applebee, A. N. (1978). *The child's concept of a story*. Chicago: Chicago University Press.

Armon-Lotem, S., Walters, J., and Gagarina, N. (2011). The impact of internal and external factors on linguistic performance in the home language and in L2 among Russian–Hebrew and Russian–German preschool children. *Linguistic Approaches to Bilingualism* 1, 291–317. doi: 10.1075/lab.1.3.04arm

Auza, A., Harmon, M. T., and Murata, C. (2018). Retelling stories: grammatical and lexical measures for identifying monolingual Spanish speaking children with specific language impairment (SLI). *J. Commun. Disord.* 71, 52–60. doi: 10.1016/j.jcomdis.2017.12.001

Bartsch, K., and Wellman, H. M. (1995). *Children talk about the mind*. New York, NY: Oxford University Press.

Bedore, L. M., Peña, E. D., Gillam, R. B., and Ho, T. H. (2010). Language sample measures and language ability in Spanish-English bilingual kindergarteners. *J. Commun. Disord.* 43, 498–510. doi: 10.1016/j.jcomdis.2010.05.002

Berman, R. A., and Slobin, D. I. (1994). *Relating events in narrative: A crosslinguistic developmental study*. Hove: Erlbaum.

Bishop, D. V. M., and Donlan, C. (2005). The role of syntax in encoding and recall of pictorial narratives: evidence from specific language impairment. *Br. J. Dev. Psychol.* 23, 25–46. doi: 10.1348/026151004X20685

Bohnacker, U., and Gagarina, N. (2020). "Introduction to MAIN–revised, how to use the instrument and adapt it to further languages" in *ZAS papers in linguistics*, (Berlin: ZAS). vol. 64, xiii–xxi.

Bonifacci, P., Barbieri, M., Tomassini, M., and Roch, M. (2018). In few words: linguistic gap but adequate narrative structure in preschool bilingual children. *J. Child Lang.* 45, 120–147. doi: 10.1017/S0305000917000149

Burns, F. A., de Villiers, P. A., Pearson, B. Z., and Champion, T. B. (2012). Dialect-neutral indices of narrative cohesion and evaluation. *Lang. Speech Hear. Serv. Sch.* 43, 132–152. doi: 10.1044/0161-1461(2011/10-0101)

Chan, A., Cheng, K., Kan, R., Wong, A. M.-Y., Fung, R., Wong, J., et al. (2020). "The multilingual assessment instrument for narratives (MAIN): adding Cantonese to MAIN" in *ZAS papers in linguistics*, vol. *64* (Berlin: ZAS), 23–29.

Coxe, S., West, S. G., and Aiken, L. S. (2009). The analysis of count data: a gentle introduction to Poisson regression and its alternatives. *J. Pers. Assess.* 91, 121–136. doi: 10.1080/00223890802634175

Dollaghan, C. A., Campbell, T. F., Paradise, J. L., Feldman, H. M., Janosky, J. E., Pitcairn, D. N., et al. (1999). Maternal education and measures of early speech and language. *J. Speech Lang. Hear. Res.* 42, 1432–1443. doi: 10.1044/jslhr.4206.1432

Eisenberg, S. L., and Guo, L. (2013). Differentiating children with and without language impairment based on grammaticality. *Lang. Speech Hear. Serv. Sch.* 44, 20–31. doi: 10.1044/0161-1461(2012/11-0089)

Fernandez, C. (2013). Mindful storytellers: emerging pragmatics and theory of mind development. *First Lang.* 33, 20–46. doi: 10.1177/0142723711422633

Fiestas, C. E., and Peña, E. (2004). Narrative discourse in bilingual children: language and task effects. *Lang. Speech Hear. Serv. Sch.* 35, 155–168. doi: 10.1044/0161-1461(2004/016)

Gagarina, N., Gey, S., and Sürmeli, N. (2019c). Identifying early preschool bilinguals at risk of DLD: a composite profile of narrative and sentence repetition skills. *ZAS Papers in Linguistics* 62, 168–189. doi: 10.21248/zaspil.62.2019.448

Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiūnienė, I., et al. (2015). Assessment of narrative abilities in bilingual children. In S. Armon-Lotem, JongJ. L. de and N. Meir (Ed.), *Assessing multilingual children: Disentangling bilingualism from language impairment* (pp. 243–276). UK: Multilingual Matters.

Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Bohnacker, U., et al. (2019). MAIN: multilingual assessment instrument for narratives-revised. *ZAS Papers in Linguistics* 63.

Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Bohnacker, U., et al. (2019a). "MAIN: multilingual assessment instrument for narratives-revised" in *ZAS papers in linguistics (ZASPiL)*. eds. A. Chan, K. Cheng, R. Kan, A. M.-Y. Wong, R. Fung and J. Wong et al., vol. *64* (Berlin: ZAS)

Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Bohnacker, U., et al. (2019b). "MAIN: Multilingual assessment instrument for narratives-revised" in *ZAS papers in linguistics (ZASPiL)*. eds. S. Z. Hamdani, R. Kan, A. Chan and N. Gagarina, vol. *64* (Berlin: Zentrum für Allgemeine Sprachwissenschaft (ZAS))

Gagarina, N., Klop, D., Tsimpli, I., and Walters, J. (2016). Narrative abilities in bilingual children. *Appl. Psycholinguist.* 37, 11–17. doi: 10.1017/S0142716415000399

Gardner, W., Mulvey, E. P., and Shaw, E. C. (1995). Regression analyses of counts and rates: Poisson, overdispersedover dispersed Poisson, and negative binomial models. *Psychol. Bull.* 118, 392–404. doi: 10.1037/0033-2909.118.3.392

Halliday, M. A. K, and Hasan, R. (1976). *Cohesion in English*. London: Longman.

Hamdani, S., Kan, R., Chan, A., and Gagarina, N. (2020). The multilingual assessment instrument for narratives (MAIN): adding Urdu to MAIN. *ZAS Papers in Linguistics* 64, 257–261. doi: 10.21248/zaspil.64.2020.580

Heilmann, J., Miller, J. F., and Nockerts, A. (2010). Sensitivity of narrative organization measures using narrative retells produced by young school-age children. *Lang. Test.* 27, 603–626. doi: 10.1177/0265532209355669

Huttenlocher, J., Waterfall, H., Vasilyeva, M., Vevea, J., and Hedges, L. V. (2010). Sources of variability in children's language growth. *Cogn. Psychol.* 61, 343–365. doi: 10.1016/j.cogpsych.2010.08.002

Justice, L. M., Bowles, R. P., Kaderavek, J. N., Ukrainetz, T. A., Eisenberg, S. L., and Gillam, R. B. (2006). The index of narrative microstructure: a clinical tool for analyzing school-aged children's narrative performances. *Am. J. Speech Lang. Pathol.* 15, 177–191. doi: 10.1044/1058-0360(2006/017)

Kan, R. T. Y., Chan, A., and Gagarina, N. (2020). Investigating Children's narrative abilities in a Chinese and multilingual context: Cantonese, mandarin, Kam and Urdu adaptations of the multilingual assessment instrument for narratives (MAIN). *Front. Psychol.: Language Sciences.* 11:573780. doi: 10.3389/fpsyg.2020.573780

Karlsen, J., Geva, E., and Lyster, S. (2016). Cognitive, linguistic, and contextual factors in Norwegian second language learner's narrative production. *Appl. Psycholinguist.* 37, 1117–1145. doi: 10.1017/S014271641500051X

Kupersmitt, J. R., Yifat, R., and Kulka, B. (2014). The development of coherence and cohesion in monolingual and sequential bilingual children's narratives. *Narrat. Inq.* 24, 40–76. doi: 10.1075/ni.24.1.03kup

Labov, W. (1972). *Language in the inner city: Studies of black English vernacular*. Philadelphia, PA: University of Pennsylvania Press.

Liles, B., Duffy, R., Merritt, D., and Purcell, S. (1995). Measurement of narrative discourse ability in children with language disorders. *J. Speech Hear. Res.* 38, 415–425. doi: 10.1044/jshr.3802.415

Loban, W. (1976). *Language development: Kindergarten through grade twelve*. Urbana, IL: National Council of Teachers of English

Mäkinen, L., Loukusa, S., Nieminen, L., Leinonen, E., and Kunnari, S. (2014). The development of narrative productivity, syntactic complexity, referential cohesion and event content in four– to eight–year–old Finnish children. *First Lang.* 34, 24–42. doi: 10.1177/0142723713511000

Mandler, G. (1979). "Organization and repetition: organization principles with special reference to rote learning" in *Perspectives on memory research*. ed. L. Nilsson (Hillsdale, NJ: Erlbaum), 293–327.

Maviş, İ., Tuncer, M., and Gagarina, N. (2016). Macrostructure components in narrations of Turkish–German bilingual children. *Appl. Psycholinguist.* 37, 69–89. doi: 10.1017/S0142716415000429

Meloun, M., and Militký, J. (2001). Detection of single influential points in OLS regression model building. *Anal. Chim. Acta* 439, 169–191. doi: 10.1016/S0003-2670(01)01040-6

Miller, P. H., and Aloise, P. A. (1989). Young children's understanding of the psychological causes of behaviour: a review. *Child Dev.* 60, 257–285. doi: 10.2307/1130975

O'Neill, D. K., Pearce, M. J., and Pick, J. L. (2004). Preschool children's narratives and performance on the Peabody Individualized Achievement Test–Revised: Evidence of a relation between early narrative and later mathematical ability. *First Language*. 24, 149–183.

Pesco, D., and Kay-Raining Bird, E. (2016). Perspectives on bilingual children's narratives elicitedrefd with the multilingual assessment instrument for narratives. *Appl. Psycholinguist.* 37, 1–9. doi: 10.1017/S0142716415000387

Raven, J. C., Court, J. H., and Raven, J. (1996). *Raven's standard progressive matrices*. Oxford, England: Oxford Psychologists Press.

Reilly, J., Losh, M., Bellugi, U., and Wulfeck, B. (2004). "Frog, where are you?" narratives in children with specific language impairment, early focal brain injury, and Williams syndrome. *Brain Lang.* 88, 229–247. doi: 10.1016/S0093-934X(03)00101-9

Renfrew, C. (1997). *Bus story test: A test of narrative speech*. Milton Keynes, UK: Speechmark.

Rodina, Y. (2017). Narrative abilities of preschool bilingual Norwegian–Russian children. *Int. J. Biling.* 21, 617–635. doi: 10.1177/1367006916643528

Sheng, L., Shi, H., Wang, D., Hao, Y., and Zheng, L. (2020). Narrative production in mandarin–speaking children: effects of language ability and elicitation method. *J. Speech Lang. Hear. Res.* 63, 774–792. doi: 10.1044/2019_JSLHR-19-00087

Silliman, E. R., Bahr, R. H., Brea, M. R., Hnath-Chisolm, T., and Mahecha, N. R. (2002). Spanish and English proficiency in the linguistic encoding of mental states in narrative retellings. *Linguist. Educ.* 13, 199–234. doi: 10.1016/S0898-5898(01)00062-6

Soodla, P., and Kikas, E. (2011). "Oral narratives of 6–7 years old Estonian children" in *Global perspectives in early childhood education: Diversity, challenges and possibilities*. eds. M. Veisson, E. Hujala, M. Waniganayake, P. Smith and E. Kikas (Frankfurt am Main, Germany: Peter Lang), 217–235.

Stein, N. L. (1988). "The development of children's storytelling skill" in *Child language: A reader*. eds. M. B. Franklin and S. Barten (New York: Oxford University Press), 262–279.

Stein, N. L., and Albro, E. R. (1997). "Building complexity and coherence: Children's use of goal structured knowledge in telling stories" in *Narrative development: Six approaches*. ed. M. Bamberg (Mahwah, NJ: Erlbaum), 5–44.

Stein, N., and Glenn, C. (1979). "An analysis of story comprehension in elementary school children" in *New directions in discourse processing*. ed. R. O. Freedle (Norwood, NJ: Ablex Publishing), 53–120.

Torng, P., and Sah, W. (2020). Narrative abilities of mandarin-speaking children with and without specific language impairment: macrostructure and microstructure. *Clinical Linguistics & Phonetic* 34, 453–478. doi: 10.1080/02699206.2019.1655097

Trabasso, T., and Nickels, M. (1992). The development of goal plans of action in the narration of a picture story. *Discourse Process.* 15, 249–275. doi: 10.1080/01638539209544812

Uccelli, P., and Paéz, M. M. (2007). Narrative and vocabulary development of bilingual children from kindergarten to first grade: developmental changes and associations among English and Spanish skills. *Lang. Speech Hear. Serv. Sch.* 38, 225–236. doi: 10.1044/0161-1461(2007/024)

Westby, C. E. (2005). "Assessing and facilitating text comprehension problems" in *Language and Reading Disabilities*. eds. H. Catts and A. Kamhi (Boston: Allyn & Bacon), 157–232.

Westerveld, M. F., and Gillon, G. T. (2010). Profiling oral narrative ability in young school–aged children. *Int. J. Speech Lang. Pathol.* 12, 178–189. doi: 10.3109/17549500903194125

Westerveld, M. F., and Roberts, J. M. A. (2017). The oral narrative comprehension and production abilities of verbal preschoolers on the autism spectrum. *Lang. Speech Hear. Serv. Sch.* 48, 260–272. doi: 10.1044/2017_LSHSS-17-0003

Zwitserlood, R., van Weerdenburg, M., Verhoeven, L., and Wijnen, F. (2015). Development of morphosyntactic accuracy and grammatical complexity in Dutch school-age children with SLI. *J. Speech Lang. Hear. Res.* 58, 891–905. doi: 10.1044/2015_JSLHR-L-14-0015

frontiers | Frontiers in Communication

TYPE Original Research
PUBLISHED 09 March 2023
DOI 10.3389/fcomm.2023.1111584

Check for updates

OPEN ACCESS

EDITED BY
Gary Morgan,
Fundació per a la Universitat Oberta de Catalunya, Spain

REVIEWED BY
Eliseo Diez-Itza,
University of Oviedo, Spain
Verónica Martínez,
University of Oviedo, Spain

*CORRESPONDENCE
Tessel Boerma
✉ t.d.boerma@uu.nl

SPECIALTY SECTION
This article was submitted to Language Sciences, a section of the journal Frontiers in Communication

RECEIVED 29 November 2022
ACCEPTED 15 February 2023
PUBLISHED 09 March 2023

CITATION
Boerma T, Everaert E, Vlieger D, Steggink M, Selten I, Houben M, Vorstman J, Gerrits E and Wijnen F (2023) Grammatical skills of Dutch children with 22q11.2 Deletion Syndrome in comparison with children with Developmental Language Disorder: Evidence from spontaneous language and standardized assessment. *Front. Commun.* 8:1111584. doi: 10.3389/fcomm.2023.1111584

Grammatical skills of Dutch children with 22q11.2 Deletion Syndrome in comparison with children with Developmental Language Disorder: Evidence from spontaneous language and standardized assessment

Tessel Boerma[1,2]*, Emma Everaert[1,2], Dinte Vlieger[1], Maaike Steggink[1], Iris Selten[1,2], Michiel Houben[2], Jacob Vorstman[3,4], Ellen Gerrits[1,5] and Frank Wijnen[1]

[1]Department of Languages, Literature and Communication, Institute for Language Sciences, Utrecht University, Utrecht, Netherlands, [2]Department of Pediatrics, Wilhelmina Children's Hospital, University Medical Center Utrecht, Utrecht, Netherlands, [3]Program in Genetics and Genome Biology, Department of Psychiatry, Research Institute, Hospital for Sick Children, Toronto, ON, Canada, [4]Department of Psychiatry, University of Toronto, Toronto, ON, Canada, [5]Research Group Speech and Language Therapy–Participation is Communication, HU University of Applied Sciences, Utrecht, Netherlands

Background: Virtually all children with 22q11.2 Deletion Syndrome (22q11DS) experience language difficulties, next to other physical and psychological problems. However, the grammatical skills of children with 22q11DS are relatively unexplored, particularly in naturalistic settings. The present research filled this gap, including two studies with different age groups in which standardized assessment was complemented with spontaneous language analysis. In both studies, we compared children with 22q11DS to children with Developmental Language Disorder (DLD), for whom the origin of language difficulties is unknown.

Methods: The first study included 187 preschool children ($n = 44$ with 22q11DS, $n = 65$ with DLD, $n = 78$ typically developing; TD). Standardized assessment consisted of grammar and vocabulary measures in both expressive and receptive modality. Spontaneous language during a play session was analyzed for a matched subsample ($n = 27$ per group). The second study included 29 school-aged children ($n = 14$ with 22q11DS, $n = 15$ with DLD). We administered standardized tests of receptive vocabulary and expressive grammar, and elicited spontaneous language with a conversation and narrative task. In both studies, spontaneous language measures indexed grammatical accuracy and complexity.

Results: Spontaneous language analysis in both studies did not reveal significant differences between the children with 22q11DS and peers with DLD. The preschool study showed that these groups produced less complex and more erroneous utterances than TD children, who also outperformed both groups on the standardized measures, with the largest differences in expressive grammar. The children with 22q11DS scored lower on the receptive language tests than the children with DLD, but no differences emerged on the expressive language tests.

Discussion: Expressive grammar is weak in both children with 22q11DS and children with DLD. Skills in this domain did not differ between the groups, despite clear differences in etiology and cognitive capacities. This was found irrespective of age and assessment method, and highlights the view that there are multiple routes to (impaired) grammar development. Future research should investigate if interventions targeting expressive grammar in DLD also benefit children with 22q11DS. Moreover, our findings indicate that the receptive language deficits in children with 22q11DS exceed those observed in DLD, and warrant special attention.

KEYWORDS

22q11.2 Deletion Syndrome, Developmental Language Disorder, spontaneous language, standardized language assessment, grammar, school-age, preschool

1. Introduction

The 22q11.2 Deletion Syndrome (22q11DS) is a genetic condition, which leads to multiple physical and psychological problems, including congenital heart defect and low intellectual functioning (McDonald-McGinn et al., 2015). Although phenotypic expression is heterogeneous, speech and/or language problems are reported in 95% of the children with 22q11DS (Solot et al., 2019), making this one of the most common features of the syndrome. The language problems in children with 22q11DS have, however, almost exclusively been described with standardized tests. Very few studies have analyzed children's spontaneous language, even though this is a more ecologically valid way to evaluate language development and can be used to set therapy goals (Klatte et al., 2022). The current study aimed to fill this gap.

In addition, we compared the language abilities of children with 22q11DS to children with Developmental Language Disorder (DLD). Similar to children with 22q11DS, children with DLD have severe difficulties with learning language. However, their language difficulties exist in the absence of the challenging physical and cognitive conditions that we see in 22q11DS. As of yet, there are no direct, large-scale comparative studies of children with 22q11DS and children with DLD. Such comparisons are meaningful to determine whether interventions for children with DLD may also be suited for children with 22q11DS. Moreover, given the etiological differences between the groups, it can enhance our understanding of the mechanisms underlying language impairment. We therefore conducted two studies, comparing the spontaneous language of both preschool and school-aged children with 22q11DS to peers with DLD. Moreover, we analyzed the results of a number of standardized language tests. In the study with preschool children, we also included a typically developing (TD) control group. In both studies, we focused on the domain of grammar, as this is a hallmark deficit in DLD, while relatively unexplored in 22q11DS.

1.1. 22q11.2 Deletion Syndrome

22q11DS is caused by a microdeletion on the long arm ('q') of chromosome 22, with the name thus referring to its genetic cause. The syndrome was previously also called Velo-Cardio-Facial, DiGeorge or Shprintzen syndrome, but we now know that these conditions are all due to the same genetic deletion: 22q11DS (McDonald-McGinn et al., 2015). It is the most frequently occurring genetic syndrome after Down syndrome, with an incidence of 1 in 2148 live births (Blagojevic et al., 2021). Despite the relatively uniform etiology, individuals with 22q11DS differ greatly in symptom expression. Over 180 manifestations have been associated with the syndrome (McDonald-McGinn et al., 2015). Congenital heart defects are the most common physical symptom, estimated to occur in up to 75% of the population. Palatal abnormalities, such as cleft palate and velopharyngeal insufficiency, are also frequently observed. In addition, cognitive and psychiatric problems are part of the syndrome. Many individuals with 22q11DS have borderline intellectual functioning or mild intellectual disability (Fiksinski et al., 2022). Moreover, 22q11DS is associated with elevated rates of psychopathology, including attention deficit hyperactivity disorder, autism spectrum disorder, anxiety disorder and psychotic disorder (Schneider et al., 2014).

1.2. Language impairment in children with 22q11DS

Next to the symptoms mentioned above, speech-language problems are observed in virtually all children with 22q11DS (Solot et al., 2019) and do not appear to be related to other manifestations of the syndrome, such as congenital heart defect and palatal abnormalities (Gerdes et al., 1999; Solot et al., 2001). In early childhood, it is reported that the first words and sentences emerge relatively late (e.g., Gerdes et al., 1999; Solot et al., 2000; Roizen et al., 2007), with some children even remaining nonverbal until the age of 4 years (Solot et al., 2001). During the preschool age, both expressive and receptive language abilities of children with 22q11DS are significantly weaker in comparison to TD children, as indicated by lower scores on standardized language tests (Gerdes et al., 1999, 2001; Solot et al., 2001; Everaert et al., 2022). A recent study (Everaert et al., 2022), using the same preschool sample as the current study, for example showed that Dutch children with 22q11DS between 3 and 6.5 years old scored, on average, 2 standard deviations below the normed mean on a composite measure of expressive language. For receptive language, this was 1.5 standard deviations below the normed mean. The significant difference in the severity of the expressive and receptive language impairment is in line with what is reported in other research with preschoolers (Gerdes et al., 1999; Solot et al., 2001). Next to composite measures, Everaert et al. (2022) also examined subtest outcomes of the standardized assessment and observed pervasive difficulties across language domains, with the lowest scores on expressive morphosyntactic skills. With the exception of Scherer et al. (1999), who showed low lexical diversity in the spontaneous language of 4 children with 22q11DS between 0;6 and 2;6 years old, an investigation of the spontaneous language of preschool children with 22q11DS has not yet been undertaken.

Research on school-age children with 22q11DS also used standardized language assessment and indicates that language impairment in 22q11DS is persistent, both in production and comprehension (Moss et al., 1999; Solot et al., 2001; Glaser et al., 2002; Rakonjac et al., 2016; Van den Heuvel et al., 2018). Language impairment even goes beyond what is expected based on children's level of intellectual functioning (Glaser et al., 2002; Persson et al., 2006; Van den Heuvel et al., 2018), in agreement with what is found for preschoolers (Gerdes et al., 1999; Scherer et al., 1999). However, in contrast to preschool children, school-age children with 22q11DS are reported to have weaker receptive than expressive language and relatively strong expressive morphosyntactic abilities (Glaser et al., 2002; Van den Heuvel et al., 2018). These contrasting findings may reflect unique developmental trends for different language modalities and domains, although more research is needed to confirm this.

Next to reporting standardized test scores, a number of studies with school-age children with 22q11DS have examined children's language profile in more detail. Van den Heuvel et al. (2018) conducted a fine-grained error analysis of two standardized tests of expressive syntax. Difficulties interpreting and using contextual cues were found to characterize the errors of their 6–13-year-old participants with 22q11DS on these tasks. In addition, three studies reported weak narrative abilities of children with 22q11DS at the macrolevel, gauging story structure and information transfer (Persson et al., 2006; Van den Heuvel et al., 2017; Selten et al., 2021). Persson et al. (2006) also analyzed the microstructural narrative production abilities of their 19 participants between 5 and 8 years old. Grammatical errors were not highly prevalent in the narrative samples, but low grammatical complexity, as indicated by short sentences and few subordinate clauses, was found to be characteristic of the stories that these children told. Van den Heuvel et al. (2017) also reported a reduced sentence length of their 6–13-year-old participants with 22q11DS in comparison with TD peers.

1.3. 22q11DS and Developmental Language Disorder

Given the severe language impairment of children with 22q11DS, which cannot be (fully) explained by cognitive or physical features of the syndrome, it is not surprising that parallels have been drawn with children with DLD. DLD is a neurodevelopmental disorder which primarily affects the ability to learn a native language (Bishop et al., 2017), estimated to occur in 3–7% of the child population (Tomblin et al., 1997; Norbury et al., 2016; Calder et al., 2022). The language difficulties of children with DLD cannot be explained by an obvious cause, such as a biomedical condition, hearing impairment, or intellectual disability. Instead, DLD is thought to arise from the interaction between multiple genetic and environmental risk factors (Bishop, 2009). These risk factors may differ from child to child, making the etiology of DLD heterogeneous. On the phenotypic level, diverse language problems in all language domains can be observed (for an overview, see Leonard, 2014; Gerrits et al., 2017). However, morphosyntactic difficulties, in Germanic languages particularly those related to verbs, are seen as a hallmark deficit and have been proposed as clinical markers that support the identification of DLD (see Leonard, 2014). Such difficulties can be observed in performance on standardized tests or other elicitation probes (e.g., Riches, 2012; Krok and Leonard, 2015; Boerma et al., 2017), but are also often shown in children's spontaneous language. Low grammatical accuracy and complexity in the spontaneous language of Dutch children with DLD is for example reflected by frequent tense and agreement errors, difficulties with argument structure, the overuse of root infinitives, a short sentence length, and the use of few complex sentences (e.g., Bol and Kuiken, 1988; De Jong, 1999; Wexler et al., 2004; Verhoeven et al., 2011; Zwitserlood et al., 2015).

As DLD per definition precludes a known biomedical condition, children with 22q11DS cannot be diagnosed with DLD. Instead, they may have a so-called 'language disorder associated with X' (Bishop et al., 2017). Despite the different labels, there appears to be substantial clinical overlap between the groups. Children with 22q11DS are often seen and treated by the same professionals that provide treatment for children with DLD (Boerma et al., 2022). It is, however, unclear whether the two groups can be differentiated based on their language profile. Previous research comparing children with DLD and children with 22q11DS is scarce. In their discussion section, Persson et al. (2006) indirectly compared the results from their 22q11DS sample with the results from a different study including children with DLD. They observed similarities across the two groups with respect to sentence length and the production of subordinate clauses, but noticed differences in grammatical accuracy, with lower accuracy for the children with DLD compared to the children with 22q11DS. Three studies directly compared children in the two groups. Kambanaros and Grohmann (2017) conducted a longitudinal case study of a boy with 22q11DS, testing him at age 6 and age 10, and compared him to children with DLD. At the age of 6, the boy produced longer sentences relative to peers with DLD, but at age 10 he scored worse on the comprehension of subject relative clauses. Other measures, including a wide range of standardized tests and experimental tasks, did not differentiate the boy from the children with DLD, neither at age 6 nor at age 10. In addition, Selten et al. (2021), using the same school-aged sample as the current study, examined narrative comprehension and production at the macrolevel of 6–10-year old children with 22q11DS and children with DLD. They did not find a significant difference on any of the narrative measures between the two groups. Using fMRI data from the same children, Van Steensel et al. (2021) even reported comparably reduced brain activation during language processing in both groups.

1.4. The current study

Previous research showed that language impairment is a common feature of 22q11DS. Children with 22q11DS experience severe language difficulties across all language domains and in both receptive as well as expressive modality. However, our knowledge of the language profile of children with 22q11DS is almost exclusively based on standardized test performance. While such tests give important information on whether language abilities are age-appropriate, they also have a number of limitations. For example, standardized language assessment does not provide insight into grammatical production skills in real-life situations, some aspects of grammar are difficult to reliably test in a standardized way, and some children may not comply with the necessary behavioral restrictions of standardized testing (Costanza-Smith, 2010; Doedens and Meteyard, 2022; Klatte et al., 2022). The latter may also hold for young children with 22q11DS, as indicated by the task completion rates reported in the study of Everaert et al. (2022). Ideally, standardized language assessment is complemented with the analysis of spontaneous language, which is ecologically valid, can be used with all children, and is considered to be the gold standard for setting therapy goals in the domain of grammar (Heilmann, 2010; Price et al., 2010).

The current study therefore investigated the spontaneous language of children with 22q11DS, aiming to further our knowledge on the syndrome's language profile. In view of the contrasting findings of previous work between preschool and school-age children, we conducted a study with each age group. We complemented spontaneous language analysis with standardized measures and, in the study with preschool children, included a TD control group. In addition, in both studies, we compared the children with 22q11DS to age-matched peers with DLD. This is the first large-scale comparison of a group of children with language problems associated with 22q11DS, a known biomedical condition accompanied by physical and cognitive challenges, and a group of children experiencing language difficulties that are not associated with such challenges. An open question is whether those two groups can be differentiated at the phenotypic level, which may have important implications for both our understanding of the required conditions for language acquisition as well as for clinical care. We focused on grammar, as weaknesses in this domain are characteristic of DLD. At the same time, relatively little is known about the grammatical skills of children with 22q11DS, especially in naturalistic settings.

Based on previous research (Persson et al., 2006; Kambanaros and Grohmann, 2017), we expected that the grammatical complexity of children with 22q11DS and children with DLD would be comparably low. Moreover, grammatical errors could be more prevalent in the group of children with DLD in comparison with the children with 22q11DS, although the evidence base for this prediction is very limited. For the preschool children, we predicted that both children with 22q11DS and children with

TABLE 1 Demographic characteristics of the preschool participants.

		N	Sex	Age in months		Intellectual functioning[a]		Parental education[b]		CELF core language[c]	
			Girls/Boys	Mean (*SD*)	Range	Mean (*SD*)	Range	Mean (*SD*)	Range	Mean (*SD*)	Range
Full sample	22q11DS	44	19/25	58.4 (12.4)	37–77	80.0 (12.0)	50–103	6.4 (1.8)	2–9	70.8 (12.2)	55–102
	DLD	65	13/52	56.7 (9.9)	36–74	97.7 (12.9)	69–124	6.3 (1.6)	3.5–9	76.9 (12.4)	55–107
	TD	78	44/34	55.5 (11.0)	36–78	106.4 (13.0)	81–139	7.8 (1.3)	3.5–9	106.3 (12.8)	85–133
Subsample	22q11DS	27	11/16	54.7 (11.3)	37–73	81.9 (11.4)	50–103	6.9 (1.9)	2–9	73.8 (12.8)	55–102
	DLD	27	11/16	54.6 (11.5)	37–74	97.0 (13.8)	70–124	6.3 (1.6)	3.5–9	78.3 (10.3)	60–94
	TD	27	11/16	54.4 (11.6)	37–75	104.6 (11.9)	84–131	7.7 (1.3)	5–9	101.3 (8.3)	87–120

22q11DS, 22q11.2 Deletion Syndrome; DLD, Developmental Language Disorder; TD, Typical Development; CELF, Clinical Evaluation of Language Fundamentals.
[a]This information is based on a wide variety of standardized, age-appropriate measures (M = 100, SD = 15). In the full sample, scores were missing for one TD child and two children with 22q11DS. In the subsample, this was the case for one child with 22q11DS.
[b]Parental education is the average education level of both parents, measured on a nine-point-scale (1 = no education, 9 = university degree). In the full sample, information was missing for one TD child and two children with DLD. In the subsample, this was the case for one child with DLD.
[c]This score of global language ability is a standardized composite (M = 100, SD = 15) of three language tests from the CELF-Preschool-2-NL. In the full sample, scores were missing for eight children with 22q11DS and two children with DLD. In the subsample, this was the case for four children with 22q11DS.

DLD would perform below TD peers on all measures, although grammatical accuracy of the children with 22q11DS could be on par with the control group. Finally, although we expected roughly similar results in the preschool and school-age study, we reckoned with the possibility that school-age children with 22q11DS would have relatively stronger grammatical skills than preschoolers, given the previous contrasting findings on expressive morphosyntactic abilities in these age groups (preschool: Everaert et al., 2022; school-age: Glaser et al., 2002; Van den Heuvel et al., 2018).

2. Study 1: Preschool

2.1. Methods

2.1.1. Participants

The children in the preschool study participated in a prospective cohort study ("3T project") which examined development in the domains of behavior, cognition and language. Participants were recruited between November 2018 and November 2019. All children were between 3 and 6.5 years of age, grew up monolingually, and had no hearing impairment. The latter two criteria were verified through a telephone interview with parents. The first group, children with 22q11DS (see Everaert et al., 2022), had a genetically confirmed diagnosis of 22q11DS. They were recruited via the 22q11DS expertise center at University Medical Center Utrecht in the Netherlands and via the Dutch patient support association. The second group, children with DLD, had been diagnosed with DLD before and independent of the 3T project by licensed professionals. They obtained an overall score of 2 standard deviations (*SD*) below the mean on a standardized language test battery or a score of 1.5 SD below the mean on two out of four language domains which were tested with at least two measures (for the full protocol, see Stichting Siméa, 2017). Moreover, next to the absence of hearing impairment, they had a non-verbal intelligence of 70 or above. The children with DLD were recruited via organizations that provide care and education services for children with communication difficulties, including Royal Kentalis, Royal Auris, VierTaal and NSDSK. At the time of the study, they all received speech-language therapy at day care or school. Finally, the third group, TD children, did not have documented developmental delays and no family history of language disorders or dyslexia. They were recruited via regular day care centers or elementary schools. Three TD children were excluded, because they obtained a score of more than 1 SD below the mean on standardized language assessment that was administered for the purpose of the 3T project. The final sample included 44 children with 22q11DS, 65 children with DLD and 78 TD children. The demographic characteristics of this sample are presented in Table 1. For a description of the prevalence of physical symptoms in our 22q11DS sample and the percentage of children receiving speech-language therapy, we refer to Everaert et al. (2022).

The three groups of children did not differ in age in months [F (2,184) $= 0.97$, $p = 0.38$, $\eta_p^2 = 0.01$]. However, there were significant differences in sex [χ^2 (2, $N = 187$) $= 19.6$, $p < 0.001$, $V = 0.32$], with relatively more boys in the group with DLD than in the other two groups (in line with what is known on DLD; Tomblin et al., 1997, but see Calder et al., 2022). Intellectual functioning, obtained from medical/school records or assessment by the current researchers, also differed significantly between the groups [F (2,181) $= 58.04$, $p < 0.001$, $\eta_p^2 = 0.39$]. The TD children obtained the highest scores, followed by the children with DLD and, finally, the children with 22q11DS (all $p < 0.001$). The average education level of both parents, measured with an online questionnaire, was also higher for the TD children in comparison with the 22q11DS and DLD groups [H (2) $= 38.0$, $p < 0.001$, $\eta^2 = 0.20$], but did not differ significantly between the latter two groups. The same pattern was observed for global language ability [F (2,174) $= 142.2$, $p < 0.001$, $\eta_p^2 = 0.62$], assessed with the Core Language Index Score of the CELF-Preschool-2-NL (Wiig et al., 2012).

As can be observed in Table 1, a subsample of 27 children in each of the three groups was selected to allow for individual matching on age in months and sex, making the groups as comparable as possible [age in months: F (2,78) $= 0.005$, $p = 0.995$, $\eta_p^2 < 0.01$; sex: χ^2 (2, $N = 81$) $= 0.00$, $p = 1.00$, $V = 0.00$]. Spontaneous language was analyzed for this subsample. A child with 22q11DS was matched to a child with DLD and a TD child from the same sex who were at most 3 months older or younger. Moreover, only TD children were selected who scored in the average range (between 85 and 115) on the Core Language Index. For one matched TD child, the quality of the language sample recording appeared to be too poor. We therefore had to replace this child with another, who did have the right sex and age but who scored above average on global language ability (i.e., 120). Similar to the full sample, the TD children in the subsample obtained higher core language scores than children in the other two groups [F (2,74) $= 50.8$, $p < 0.001$, $\eta_p^2 = 0.58$], which, in turn, did not differ from each other. We did not match on intellectual functioning, as differences between the groups are inherent [F (2,77) $= 22.8$, $p < 0.001$, $\eta_p^2 = 0.37$]. In the subsample, intellectual functioning of the children with DLD and TD children was not significantly different anymore ($p = 0.082$), and was higher than the intellectual functioning of the children with 22q11DS (all $p < 0.001$). Finally, parental education differences between the three groups remained significant [H (2) $= 9.5$, $p = 0.009$, $\eta^2 = 0.10$]. This effect was driven by differences between the DLD and TD groups ($p = 0.003$).

2.1.2. Instruments

2.1.2.1. Standardized language measures

Standardized language measures were used to assess children's abilities in the domains of expressive and receptive grammar. To determine whether grammatical skills are a relative strength or weakness, we also included measures of expressive and receptive vocabulary. Scores of the children with 22q11DS on these tests have been reported in Everaert et al. (2022).

Subtests of the Preschool version of the Clinical Evaluation of Language Fundamentals, CELF-Preschool-2-NL (Wiig et al., 2012), evaluated expressive grammar, receptive grammar and expressive vocabulary. All subtests were administered following the official manual and have a normed mean of 10 ($SD = 3$). Expressive grammar was measured with two subtests, on word level and

TABLE 2 Main outcome parameters of the spontaneous language samples.

	Parameter	Description
Grammatical accuracy	% T-units correct	Number of error-free T-units divided by the total number of T-units.
	% Verb-related errors[a]	Number of verb-related errors divided by the total number of clauses.
	% Non-verb-related errors[b]	Number of non-verb-related errors divided by the total number of T-units.
Grammatical complexity	MLU	Number of words divided by the total number of T-units.
	MLU 5	Number of words divided by the total number of T-units in the 5 longest T-units.
	% Clauses with a verb	Number of utterances containing a verb divided by the total number of clauses.
	% Complex utterances[c]	Number of complex utterances divided by the total number of T-units.

MLU, Mean Length of Utterance.

[a]Verb-related errors include argument omissions, subject-verb agreement errors, tense errors, root infinitives, verb-second placement errors, overgeneralizations, past participle errors, verb omissions and other verb-related errors which could not be further categorized. Examples can be found in the Appendix.

[b]Non-verb-related errors include determiner errors, errors with adjectival inflection, preposition errors, pronoun errors, errors with conjunction, plural errors, errors with the pronominal/adverbial "er" [there], word order errors (not related to verb-second placement), and other non-verb-related errors which could not be further categorized. Examples can be found in the Appendix.

[c]Complex utterances include subordinate clauses, clauses with conjunction reduction, direct speech, and infinitival clauses. Examples can be found in the Appendix.

on sentence level. During the subtest Word Structure, children saw one or two pictures and were asked to complete a sentence uttered by the researcher, thereby eliciting the production of verbs, adjectives, plurals, pronouns and diminutives. The second subtest of expressive grammar was Recalling Sentences, which is a sentence repetition task with items that increase in length and complexity. This type of task is considered to test syntactic skills (Polišenská et al., 2015). Receptive grammar was measured with the subtest Sentence Structure. Children saw four pictures and were asked to point to the picture that best matched a sentence uttered by the researcher. The test assesses children's understanding of different grammatical structures, including passives, relative clauses, negation and prepositional phrases. Finally, expressive vocabulary was evaluated with the Expressive Vocabulary subtest. Children saw a picture of an object or action and had to label the picture.

Receptive vocabulary skills were assessed with the Peabody Picture Vocabulary Test (PPVT-III-NL; Schlichting, 2005). The test was administered in accordance with the official manual and quotient scores with a mean of 100 (SD = 15) are reported. Children saw four pictures and heard a target word. They were asked to point to the picture which corresponded to the target word.

2.1.2.2. Spontaneous language samples

Spontaneous language of children was collected during a play session of ∼15–20 min. The play break followed a standardized protocol and was divided in three parts. In the first part, children played alone with a fixed set of toys, including the Playmobil city life petting zoo set and a number of plastic fruits/vegetables. After a few minutes, or sooner if the child did not speak during this part, the researcher brought a tractor and joined the child. In this second part, the child and researcher played together, but the child remained in charge of what was happening. The researcher was instructed to follow the child, only taking initiative when the child had clear difficulty playing with the toys. After around 10 min, the final part of the play break began, in which both the child and researcher colored with crayons. If the child did not speak much, the researcher would ask open-ended questions.

2.1.3. Procedure

The 3T project was approved by the Medical Research Ethics review board of the University Medical Center Utrecht (CCMO registry nr. NL63223.041.17). Parents of participating children signed an informed consent form. The researchers who worked with the children had a background in linguistics or psychology and were trained using a standardized protocol. Children were individually tested in a quiet room at day care or school. Standardized language tests, cognitive tasks and the play break were administered in a fixed order during two sessions of ∼45 min each. The two test sessions were on separate days and were always administered by same researcher. The play break was in the second session. This was video-recorded with a GoPro HERO camera and, for adequate audio recordings, a Samson Go Mic portable USB microphone was used. The standardized tests for expressive language were recorded with the same USB microphone and also scored by a second researcher. Discrepancies were discussed and solved by consensus.

The language samples of the 27 children in each of the three groups were transcribed according to the Codes for the Human Analysis of Transcripts (CHAT) conventions (part of CHILDES; MacWhinney, 2000), by trained researchers with a background in linguistics. The T-unit was used as the basic unit of analysis, defined as a main clause with subordinate clauses attached to it (Hunt, 1970). Quality checks were done by the first and senior author to guarantee that the conventions were accurately followed. Moreover, the transcripts were annotated on a separate tier for grammatical accuracy and complexity (see Data analysis). For sake of reliability, the annotations of nine transcriptions (three of each group; 11%) were compared with annotations from a second researcher. Annotation agreement was reached in 94.6% of the T-units.

2.1.4. Data analysis

The analyses were performed in Computerized Language Analysis Software (CLAN, part of CHILDES; MacWhinney, 2000) and SPSS version 28 (IBM Corp, 2021). Univariate ANOVA's were done to compare the three groups on the five standardized language measures. As the groups significantly differed in SES and sex, while these differences are not inherent to the groups, we also conducted univariate ANCOVA's. The inclusion of the covariates SES and sex did not change the results. Intellectual functioning differences are inherent to the groups and intellectual functioning was therefore not included as a covariate in the analyses (Miller and Chapman, 2001; Dennis et al., 2009). All analyses were done for the full sample as well as the subsample. Results for the subsample did not differ from the results of the full sample and are therefore not reported. As an additional analysis, we conducted paired samples *t*-tests in the DLD and 22q11DS groups to investigate whether there was a discrepancy between expressive grammar (measured with subtests "word structure" and "recalling sentences") and the other language domains. For this analysis, quotient scores of the receptive vocabulary task were transformed to CELF-scores.

The analyses of the spontaneous language samples focused on grammatical accuracy and grammatical complexity, and were based on the work of Zwitserlood et al. (2015). The main outcome parameters of both categories are presented in Table 2 (see the Appendix for examples of errors and complex utterance categories). All outcome parameters exclude interjections and communicators (e.g., "uh," "yes," "no"; on average 19% of the total number of a child's utterances), onomatopoeia (2%), unintelligible utterances (6%), as well as incomplete sentences due to trailing off and interruption (2%). Furthermore, the outcome parameters are corrected for length of the included language sample, as this differed per child. That is, all outcome parameters are calculated as proportions, taking into account the total number of T-units (or, in some specific cases, the total number of clauses). Sample length, calculated as the total number of T-units after exclusions, did not significantly differ between the three groups of children (22q11DS: $M = 108$, $SD = 51$; DLD: $M = 130$, $SD = 61$; TD: $M = 122$, $SD = 61$; F (2, 78) $= 1.02$, $p = 0.37$, $\eta^2_p = 0.025$).

Next to the outcome parameters presented in Table 2, we also report on a number of specific verb-related errors (part of the main parameter "% verb-related errors"), as these errors are known to occur frequently in the spontaneous language of Dutch children with DLD. These specific verb-related errors include (1) the number of subject-verb agreement errors relative to the total subject-verb agreement attempts, (2) the number of past tense errors relative to the total number of T-units requiring a past tense, (3) the number of root infinitives relative to the number of T-units containing a verb, (4) the omission of an argument (subject, object or other) relative to the number of T-units containing a verb. Comparable to the analyses with the standardized language measures, univariate AN(C)OVA's were done to compare the three groups on all main outcome parameters for grammatical accuracy and grammatical complexity. The inclusion of SES as covariate did not change the results. For the specific verb-related errors and for the main outcome parameter "% complex utterances," we conducted non-parametric tests (Kruskall Wallis H test and, for *post-hoc* comparisons, Mann Whitney U test), as inspection of the data showed violations of the assumptions of normality and equality of error variances. Effect sizes were interpreted following Cohen (1988).

2.2. Results

2.2.1. Standardized language measures

The performance of the three groups of children (full sample) on the standardized tests of grammar and vocabulary is presented in Table 3. The results showed significant group effects on all five measures. For receptive grammar [F (2,180) $= 68.6$, $p < 0.001$, $\eta^2_p = 0.43$], all groups differed significantly from each other (all $p < 0.001$), with the highest scores for the TD children and the lowest scores for the children with 22q11DS. The TD children also obtained the highest scores on both subtests of expressive grammar [word level: F (2,175) $= 116.9$, $p < 0.001$, $\eta^2_p = 0.57$; sentence level: F (2,173) $= 135.3$, $p < 0.001$, $\eta^2_p = 0.61$], but there were no differences between the group of children with 22q11DS and the group of children with DLD on these measures (all $p = 1.00$). Receptive vocabulary showed similar results as receptive grammar [F (2,182) $= 64.3$, $p < 0.001$, $\eta^2_p = 0.41$], with significant differences between all groups (TD>DLD>22q11DS; all $p < 0.001$). Finally, performance on expressive vocabulary [F (2,177) $= 88.6$, $p < 0.001$, $\eta^2_p = 0.50$] was best for the TD children in comparison to the other two groups. Scores of the children with 22q11DS and the children with DLD did not differ significantly ($p = 0.09$).

Comparing the average scores per group across language domains, we see low performance of children with 22q11DS on all measures. For both the children with 22q11DS and the

TABLE 3 Performance of the three groups of preschool children on the standardized language measures[a].

	22q11DS			DLD			TD		
	N^b	Mean (*SD*)	Range	*N*	Mean (*SD*)	Range	*N*	Mean (*SD*)	Range
Receptive grammar	40	5.7 (2.6)	1–10	65	8.1 (3.0)	1–14	78	11.5 (2.5)	7–18
Expressive grammar: word	36	4.3 (3.1)	1–12	64	4.4 (2.5)	1–11	78	10.8 (2.9)	4–17
Expressive grammar: sentence	35	4.8 (2.3)	1–11	64	4.5 (1.7)	1–9	77	10.0 (2.4)	5–15
Receptive vocabulary	42	83.7 (14.0)	55–114	65	96.0 (10.5)	72–120	78	108.9 (11.6)	82–144
Expressive vocabulary	39	5.2 (2.3)	1–10	63	6.3 (2.4)	1–11	78	10.6 (2.4)	6–16

22q11DS, 22q11.2 Deletion Syndrome; DLD, Developmental Language Disorder; TD, Typical Development.
[a]Sentence Structure, Word Structure, Recalling Sentences and Expressive Vocabulary of the CELF-Preschool-2-NL (M = 10, SD = 3) were used to measure receptive grammar, expressive grammar: word, expressive grammar: sentence and expressive vocabulary, respectively. The PPVT-III-NL (M = 100, SD = 15) was used to test receptive vocabulary.
[b]Not all children, particularly children with 22q11DS, were able to complete all tests due to poor task compliance and limited language production (for an elaborate discussion of the task completion rates of the group of children with 22q11DS, see Everaert et al., 2022).

children with DLD, the lowest mean scores are on the two subtests of expressive grammar (close to −2 *SD* below the mean). For the children with DLD, a larger discrepancy between expressive grammar and the other domains are observed than for the children with 22q11DS. Paired samples *t*-tests between the two expressive grammar subtests on the one hand and the other standardized measures on the other hand showed significant differences across the board in the DLD group (all $p < 0.001$), with effect sizes ranging from 0.79 to 1.73. In the 22q11DS group, significant differences were also observed ($p < 0.05$), with the exception of "recalling sentences" in comparison with "active vocabulary" ($p = 0.20$) and "recalling sentences" in comparison with "sentence comprehension" ($p = 0.053$). Effect sizes ranged from 0.22 to 0.98.

2.2.2. Spontaneous language samples

For each of the three groups, the means and standard deviations on all outcome measures for grammatical accuracy and grammatical complexity are presented in Table 4.

2.2.2.1. Grammatical accuracy

Grammatical accuracy was subdivided into three main outcome parameters and four specific verb-related errors. The relative number of error-free T-units is a broad measure of grammatical accuracy, for which a significant effect of Group was observed [F (2,78) $= 18.0$, $p < 0.001$, $\eta_p^2 = 0.32$]. TD children produced relatively more error-free T-units than children with 22q11DS and children with DLD (both $p < 0.001$). No significant differences emerged between the latter two groups ($p = 1.00$). The same pattern was found for the other two main outcome parameters. That is, there were significant effects of Group on both verb-related errors [F (2,78) $= 19.4$, $p < 0.001$, $\eta_p^2 = 0.33$] and non-verb-related errors [F (2,78) $= 12.9$, $p < 0.001$, $\eta_p^2 = 0.25$]. In comparison with the other two groups, TD children produced relatively less verb-related (both $p < 0.001$) and non-verb-related (22q11DS: $p = 0.007$; DLD: $p < 0.001$) errors. The groups of children with 22q11DS and children with DLD did not differ significantly from each other on either parameter (verb-related: $p = 1.00$; non-verb-related: $p = 0.20$).

Results from the specific verb-related errors showed one very extreme outlier in the 22q11DS group on the proportion of subject-verb agreement errors (scoring 100%). This child was very young (3;1 years old) and produced a limited number of utterances. We excluded this outlier from the analyses, although results with and without the outlier remained the same. The analyses demonstrated significant group effects on the proportion of subject-verb agreement errors [H (2) $= 9.3$, $p = 0.009$, $\eta^2 = 0.10$], root infinitives [H (2) $= 12.4$, $p = 0.002$, $\eta^2 = 0.13$] and argument omissions [H (2) $= 27.7$, $p < 0.001$, $\eta^2 = 0.33$]. On all three error categories, TD children scored lower, and thus produced less errors, than children with 22q11DS (subject-verb agreement errors: $U = 183.0$, $z = -3.0$, $p = 0.003$, $r = 0.41$; root infinitives: $U = 266.0$, $z = -2.3$, $p = 0.02$, $r = 0.31$; argument omissions: $U = 118.5$, $z = -4.3$, $p < 0.001$, $r = 0.58$) and children with DLD (subject-verb agreement errors: $U = 243.0$, $z = -2.1$, $p = 0.04$, $r = 0.29$; root infinitives: $U = 187.0$, $z = -3.6$, $p = 0.02$, $r = 0.49$; argument omissions: $U = 87.0$, $z = -4.8$, $p < 0.001$, $r = 0.65$). There were no significant differences between the children with 22q11DS and the children with DLD on these three specific verb-related errors (subject-verb agreement errors: $p = 0.48$); root infinitives: $p = 0.22$; argument omissions: $p = 0.72$). With respect to the number of past tense errors, no significant group effect emerged ($p = 0.80$), likely due to the relatively infrequent use of past tense contexts.

2.2.2.2. Grammatical complexity

Grammatical complexity was subdivided into four main outcome parameters. Results from the analyses on Mean Length of Utterance (MLU) showed significant group effects on both MLU [F (2,78) $= 13.1$, $p < 0.001$, $\eta_p^2 = 0.25$] and MLU 5 [F (2,78) $= 10.5$, $p < 0.001$, $\eta_p^2 = 0.21$]. TD children produced longer sentences than children with 22q11DS and children with DLD (all $p < 0.001$),

TABLE 4 Outcomes of the three groups of preschool children on the spontaneous language measures.

		22q11DS		DLD		TD	
		Mean (*SD*)	Range	Mean (*SD*)	Range	Mean (*SD*)	Range
Grammatical accuracy	% T-units correct	70.9 (9.6)	52–95	69.3 (7.5)	52–83	82.1 (8.4)	60–100
	% Verb-related errors	22.0 (9.7)	4–41	22.2 (8.0)	8–41	10.0 (6.7)	0–29
	% Non-verb-related errors	16.8 (6.4)	2–29	20.1 (7.4)	8–34	11.1 (5.7)	0–23
	Specific verb-related errors:						
	% Subject-verb agreement[a]	8.0 (5.8)	0–23	8.3 (8.4)	0–29	3.6 (3.6)	0–12
	% Past tense	6.3 (20.0)	0–100	6.2 (18.6)	0–67	3.6 (8.8)	0–33
	% Root infinitives	1.9 (3.8)	0–16	1.9 (3.2)	0–15	0.23 (0.74)	0–3
	% Argument omissions	17.5 (11.5)	0–47	19.3 (11.8)	2–46	5.8 (5.7)	0–25
Grammatical complexity	MLU	3.0 (0.94)	1–5	3.0 (0.73)	2–5	3.9 (0.62)	3–5
	MLU 5	6.6 (2.5)	2–12	6.5 (1.8)	4–11	9.1 (2.6)	4–16
	% Clauses with a verb[b]	54.8 (12.6)	25–76	53.6 (13.9)	21–78	65.6 (10.4)	42–78
	% Complex utterances	1.2 (2.0)	0–8	1.5 (1.7)	0–6	4.1 (3.1)	0–10

22q11DS, 22q11.2 Deletion Syndrome; DLD, Developmental Language Disorder; TD, Typical Development.
[a]One very extreme outlier in the 22q11DS group was excluded (see main text).
[b]One very extreme outlier in the 22q11DS group was excluded (see main text).

whereas the latter two groups did not differ in their MLU and MLU 5 (all $p = 1.00$). Another index of grammatical complexity was the proportion of utterances containing a verb. There was one very extreme outlier in the 22q11DS group from a young child (3;4 years old; scoring 1.8%) which was excluded from the analyses; results with and without the outlier remained the same. A significant effect of Group emerged on the proportion of utterances containing a verb [F (2,77) = 7.7, $p < 0.001$, $\eta_p^2 = 0.17$], with TD children producing relatively more utterances with a verb than the groups of children with 22q11DS and children with DLD (all $p < 0.001$), who did not differ ($p = 1.00$). Finally, the same pattern appeared from the proportion of complex sentences [H (2) = 18.2, $p = 0.002$, $\eta^2 = 0.21$]. There were no significant differences between the children with 22q11DS and the children with DLD ($p = 0.25$), who produced less complex sentences than their TD peers (22q11DS: $U = 147.5$, $z = -3.8$, $p < 0.001$, $r = 0.52$; DLD: $U = 174.5$, $z = -3.3$, $p < 0.001$, $r = 0.45$).

3. Study 2: School-age

3.1. Methods

3.1.1. Participants

The children in the school-age study participated in a project on language processing and activation in the brain (see Selten et al., 2021; Van Steensel et al., 2021). Participants were recruited between November 2017 and July 2018. The 6–10-year-old participants included 14 children with a genetically confirmed diagnosis of 22q11DS and 15 children with an official diagnosis of DLD (for a description of the DLD criteria and protocol used in the Netherlands, see 2.1.1.). All children had either a verbal or nonverbal intellectual functioning level of 70 or above. Moreover, they did not have hearing loss of more than 35 decibel, as determined by pure tone audiometry, nor a diagnosis of autism spectrum disorder. Finally, due to an fMRI scan which was also part of the research protocol (Van Steensel et al., 2021), children were excluded if they had metal objects in their bodies or if they experienced severe anxiety in the scanner. Recruitment procedures were similar to the study with preschool children. Demographic characteristics of the two groups of children are presented in Table 5. The two groups did not differ on age in months [t (27) = 0.79, $p = 0.44$, $d = 0.29$] and sex [χ^2 (1, N = 29) = 0.042, $p = 0.84$, $V = 0.04$]. As expected, significant differences in intellectual functioning were observed [t (1,20.2) = 6.57, $p < 0.001$, $d = 2.48$], with higher levels of the children with DLD relative to the children with 22q11DS.

3.1.2. Instruments

3.1.2.1. Standardized language measures

We included one standardized measure of expressive grammar and, as a reference, one standardized measure of receptive vocabulary, which were both administered in line with the official manuals. Results from these measures have been reported as background measures in the study of Selten et al. (2021). Similar to the study with preschool children, expressive grammar was tested with a sentence repetition task. The Recalling Sentences subtest of the school-aged version of the CELF, the CELF-IV-NL (Kort et al., 2008), required children to repeat sentences of increasing length and complexity. The normed scores have a mean of 10 (SD = 3). Receptive vocabulary was assessed with the PPVT-III-NL (see 2.1.2.1.).

3.1.2.2. Spontaneous language samples

Spontaneous language of children was collected with a narrative task which was preceded by a conversation between the researcher and the participating child. We used the Multilingual Assessment Instrument for Narratives (MAIN) (Gagarina et al., 2012; for the Dutch version, see Blom et al., 2020) to elicit semi-spontaneous language. The MAIN targets narrative abilities of 3- to 10-year-old children and consists of four comparable stories, all matched to six full-color picture sequences. In the current research, the stories "Cat" and "Baby Birds" were used. The children first saw the picture sequence belonging to "Cat." The researcher told the story and asked the child ten comprehension questions. Subsequently, children saw the picture sequence belonging to "Baby Birds" and were asked to generate their own story, which was, again, followed by ten comprehension questions. The MAIN can be used to analyze children's understanding and production of story structure (i.e., narrative abilities at the macrolevel; see Selten et al., 2021), but can also be used to examine microstructural narrative skills, including grammatical accuracy and complexity. For the current study, we used the narrative generated by the children, thus excluding children's answers to the comprehension questions, and complemented this with spontaneous language from a preceding conversation. This allowed us to elicit more utterances and to more reliably investigate grammatical skills. The conversation between the researcher and child was about day-to-day topics, such as birthdays, vacations and hobbies.

3.1.3. Procedure

Ethical approval was obtained from the Medical Research Ethics review board of the University Medical Center Utrecht

TABLE 5 Demographic characteristics of the school-aged participants.

		Sex	Age in months		Intellectual functioning[a]	
	N	Girls/Boys	Mean (SD)	Range	Mean (SD)	Range
22q11DS	14	6/8	104.2 (19.1)	80–131	74.0 (8.6)	64–94
DLD	15	7/8	98.4 (20.5)	74–131	105.4 (15.7)	86–136

22q11DS, 22q11.2 Deletion Syndrome; DLD, Developmental Language Disorder.

[a]This information is based on a wide variety of standardized, age-appropriate measures (M = 100, SD = 15), obtained from medical/school records or via own administration. There was missing data for one child with DLD.

(CCMO registry nr. NL62366.041.17). Parents of participants gave written informed consent. The researchers who worked with the children were the same as those who worked with the preschool children. The individual test session of ∼1 h took place in a quiet room at the University Medical Center Utrecht. Language tests were administered in a fixed order. Spontaneous language as well as the standardized test for expressive grammar were recorded with a Samson Go Mic portable USB microphone. With respect to the transcriptions and annotations of the spontaneous language samples, procedures were similar to what has been previously described for the preschool children (see 2.1.3.). A total of 10% of the annotations, randomly selected from three participants with 22q11DS and three participants with DLD, were compared with annotations from a second researcher. Annotation agreement was reached in 91.5% of T-units.

3.1.4. Data analysis

Similar to the preschool study, the analyses were performed in Computerized Language Analysis Software (CLAN; MacWhinney, 2000) and SPSS version 28 (IBM Corp, 2021). Independent samples t-tests were done to compare the children with 22q11DS and the children with DLD on the two standardized language measures. Moreover, a paired samples *t*-test was done to investigate whether there was a discrepancy between expressive grammar (measured with the subtest "recalling sentences") and other language domains (in this case, receptive vocabulary). The data-analysis approach of the spontaneous language of the school-age children corresponded to the approach of the study with preschoolers (see 2.1.4.). The mean percentage of excluded utterances was 17% for interjections/communicators, 1% for onomatopoeia, 4% for unintelligible utterances, and 3% for incomplete sentences. Sample length, calculated as the total number of T-units after exclusions, did not significantly differ between the two groups of children (22q11DS: $M = 69$, $SD = 28$; DLD: $M = 80$, $SD = 26$; t (27) = 1.09, $p = 0.29$, $d = 0.41$). Independent samples *t*-tests compared scores of the two groups on the main outcome parameters for grammatical accuracy and complexity (Table 2), as well as on the four specific verb-related error categories. As the groups in the school-age study were small, we provided the full statistics for both significant and non-significant results. Effect sizes were interpreted following Cohen (1988).

3.2. Results

3.2.1. Standardized language measures

The mean scores of the children with 22q11DS and the children with DLD on the expressive grammar test were 5.1 ($SD = 2.2$, range = 1–8) and 3.9 ($SD = 2.0$, range = 1–7), respectively. These scores were not significantly different from each other [t (27) = 1.6, $p = 0.13$, $d = 0.58$]. On the receptive vocabulary test, the children with 22q11DS scored, on average, 83.1 ($SD = 13.7$, range = 66–110). The children with DLD had a mean score of 93.2 ($SD = 13.6$, range = 72–117), which fell just short of significance relative to the children with 22q11DS [t (26) = 2.0, $p = 0.06$, $d = 0.74$]. Comparable to the results from the preschool children, the weakest mean scores for both groups were found on expressive grammar. The discrepancy between the expressive grammar and receptive vocabulary scores was larger for the children with DLD than for the children with 22q11DS, as shown by the results of the paired samples *t*-tests. A significant difference emerged between expressive grammar and receptive vocabulary in the DLD group [t (14) = 7.0, $p < 0.001$, $d = 1.81$], whereas this difference did not reach significance in the 22q11DS group [t (12) = 1.0, $p = 0.08$, $d = 0.52$).

3.2.2. Spontaneous language samples

For each of the two groups, the means and standard deviations on all outcome measures for grammatical accuracy and grammatical complexity are presented in Table 6.

TABLE 6 Outcomes of the two groups of school-aged children on the spontaneous language measures.

		22q11DS		DLD	
		Mean (*SD*)	Range	Mean (*SD*)	Range
Grammatical accuracy	% T-units correct	71.0 (8.8)	58–87	71.4 (8.5)	60–84
	% Verb-related errors	17.5 (8.7)	8–35	16.8 (5.8)	9–30
	% Non-verb-related errors	19.8 (7.2)	9–33	20.0 (7.7)	8–37
	Specific verb-related errors:				
	% Subject-verb agreement	5.7 (5.1)	0–17	4.8 (3.3)	0–13
	% Past tense[a]	9.1 (10.0)	0–33	11.7 (14.3)	0–43
	% Root infinitives	0.35 (0.74)	0–2	0.17 (0.46)	0–2
	% Argument omissions	4.9 (3.7)	0–12	7.1 (4.7)	2–17
Grammatical complexity	MLU	5.3 (0.67)	5–7	5.1 (1.2)	4–7
	MLU 5	11.3 (2.1)	8–15	11.6 (3.3)	8–18
	% Clauses with a verb	70.5 (8.0)	49–87	70.4 (11.2)	51–81
	% Complex utterances	8.7 (4.9)	2–21	8.7 (8.3)	0–28

22q11DS, 22q11.2 Deletion Syndrome; DLD, Developmental Language Disorder; TD, Typical Development.
[a]One very extreme outlier in the 22q11DS group was excluded (see main text).

3.2.2.1. Grammatical accuracy

Again, grammatical accuracy was subdivided into three main outcome parameters and four specific verb-related errors. On all three main outcome parameters, no significant differences emerged between the children with 22q11DS and the children with DLD [error-free T-units: t (27) = 0.13, p = 0.90, d = 0.05; verb-related errors: t (27) = 0.28, p = 0.78, d = 0.10; non-verb related errors: t (27) = 0.06, p = 0.95, d = 0.02]. Effect sizes were all small. Results from the specific verb-related errors showed one very extreme outlier in the 22q11DS group on the proportion past tense errors (scoring 100% due to one incorrect past tense attempt). We excluded this outlier from the analyses, although results with and without the outlier remained the same. The analyses demonstrated that the groups did not differ significantly on the proportion of subject-verb agreement errors [t (27) = 0.58, p = 0.57, d = 0.22], past tense errors [t (26) = 0.54, p = 0.59, d = 0.21], and argument omissions [t (27) = 1.4, p = 0.17, d = 0.52]. The effect sizes were all small, except for the proportion of argument omissions for which a medium effect size was found. The proportion of root infinitives was very small in both groups, so no statistical analyses were performed for this category.

3.2.2.2. Grammatical complexity

Grammatical complexity was subdivided into four main outcome parameters. The same pattern was observed for all complexity parameters. The children with 22q11DS and the children with DLD did not differ on MLU [t (22.7) = 0.55, p = 0.59, d = 0.20], MLU 5 [t (27) = 0.23, p = 0.82, d = 0.09], the proportion of clauses containing a verb [t (23.0) = 0.04, p = 0.99, d = 0.02] and the proportion of complex sentences [t (23.0) = 0.02, p = 0.99, d = 0.01). Effect sizes were all small.

4. Discussion

Language impairment is characteristic of children with 22q11.2 Deletion Syndrome (22q11DS; Solot et al., 2019), next to other physical and psychological symptoms such as congenital heart defect and low intellectual functioning (McDonald-McGinn et al., 2015). However, the language difficulties of children with 22q11DS have almost exclusively been described with standardized language tests, while the analysis of spontaneous language is more ecologically valid and the preferred method for setting therapy goals in the domain of grammar (Klatte et al., 2022). We aimed to contribute to a more complete overview of the language profile of preschool and school-age children with 22q11DS, conducting two studies in which we complemented standardized language testing with the analysis of spontaneous language. In both studies, we compared children with 22q11DS to age-matched children with Developmental Language Disorder (DLD), who also experience severe language difficulties but for whom the cause is unknown. We focused on children's grammatical skills, as these are typically weak in children with DLD (Leonard, 2014) while relatively unexplored in children with 22q11DS.

4.1. The language profile of children with 22q11DS

The standardized test results from both the study with preschool children and school-age children confirm that language impairment is common in children with 22q11DS (e.g., Van den Heuvel et al., 2018; Solot et al., 2019; Everaert et al., 2022). Although there was substantial variation within our 22q11DS samples, the mean scores on the standardized subtests were all more than 1 Standard Deviation (SD) below what is expected based on chronological age. In both the preschool and school-age study, the lowest scores were found on the subtests for expressive grammar, with mean scores between 1.7 and nearly 2 *SD* below the mean. Although this contrasts with previous research on school-age children with 22q11DS (Glaser et al., 2002; Van den Heuvel et al., 2018), which reported a relative weakness in receptive grammar and semantics, differences between the mean subtest scores were small and strong conclusions about relative strengths and weaknesses in the language profile of children with 22q11DS can therefore not be drawn (see also Everaert et al., 2022). In addition, the results from the two studies that we conducted with different age groups do not give reason to assume unique developmental trends for different language domains or modalities in 22q11DS, as was previously suggested (for a discussion, see Van den Heuvel et al., 2018). Although direct comparisons between the age groups should be interpreted with caution, mean norm scores on the two standardized tests that were included in both studies were comparable between the preschool and school-age children with 22q11DS and thus do not point to a developmental shift in the language profile.

The spontaneous language analysis in the preschool study, which included a typically developing (TD) control group, confirmed the findings from the standardized assessments. Hence, the current study shows that language impairment in 22q11DS is also characterized by weak language performance in real-life situations. During play, our 3–6-year-old participants with 22q11DS produced shorter and less complex utterances than their age-matched TD peers. They also made more grammatical errors in both verb- and non-verb-related categories. The low complexity of the spontaneous language that we observed in the children with 22q11DS corresponds to previous results from a narrative and a perspective-taking task (Persson et al., 2006; Van den Heuvel et al., 2017). However, the results from the current study diverge from Persson et al. (2006) with respect to grammatical accuracy. Their 5–8-year-old participants with 22q11DS produced substantially fewer utterances with grammatical errors than both the preschool and school-age participants with 22q11DS of the current study. This could possibly be explained by a relatively short utterance length of the participants of Persson et al. (2006), which, in turn, could result in fewer grammatical errors. However, Persson et al. (2006) used a narrative task to elicit spontaneous language, which is associated with longer utterances and more errors than elicitation methods such as play or conversation that were used in the current study (e.g., Wetherell et al., 2007). A reverse pattern of findings would have therefore been easier to understand. Note that if we compare our findings to Zwitserlood et al. (2015), a Dutch study which also elicited spontaneous language with a narrative task, we do see differences in the expected direction. The participants of Zwitserlood et al. (2015) produced relatively longer/more complex utterances and made relatively more errors than the participants of the current study, in line with results from research comparing different elicitation methods (e.g., Wetherell et al., 2007).

4.2. Comparing children with 22q11DS to children with DLD

The comparisons of the children with 22q11DS to children with DLD pointed toward differences in their respective receptive language skills and similarities in their expressive language abilities. The preschool children with 22q11DS were outperformed by the children with DLD on the standardized receptive language tests of grammar and vocabulary. A trend in the same direction was observed in the school-age study, which only included one receptive language measure (i.e., receptive vocabulary). We did not find significant differences between the children with 22q11DS and children with DLD on the expressive language tests, in either age group. Like the children with 22q11DS, the children with DLD also scored lowest on the subtests measuring expressive grammar, which was to be expected based on what is known about DLD (e.g., Leonard, 2014). A clear discrepancy between the expressive grammar subtest scores and the scores on the other tested domains was only found in the children with DLD.

The analysis of spontaneous language also revealed that expressive grammar is vulnerable in both 22q11DS and DLD. We did not find evidence for a difference on any of the main outcome parameters gauging grammatical accuracy and complexity between children with 22q11DS and peers with DLD, irrespective of age group. Moreover, the frequency of specific verb-related errors which are known to characterize the spontaneous language of Dutch children with DLD (e.g., De Jong, 1999; Zwitserlood et al., 2015) also did not differ between the groups. In fact, mean scores of the two groups were remarkably close together on many of the outcome variables. This largely confirms the findings from the three previous studies that directly compared children with 22q11DS to children with DLD and also reported substantial overlap between the groups (Kambanaros and Grohmann, 2017; Selten et al., 2021; Van Steensel et al., 2021). Of note, although we were not able to include a TD control group in the school-age study, the overlap in expressive language performance between 22q11DS and DLD suggests that school-aged children with 22q11DS are likely to struggle with language production in naturalistic settings. This confirms the findings in the preschool study.

4.3. Implications, limitations and future directions

Our findings highlight the necessity to regularly assess and monitor the language development of children with 22q11DS as part of routine clinical care, as recommended by Solot et al. (2019). Given the broad linguistic weaknesses of children with 22q11DS, but also the large individual differences in the severity of these weaknesses, routine assessments from a young age onward are necessary to support early interventions, and, in turn, mitigate the ramifications of language impairment and improve outcomes. Research can contribute to these goals by providing more knowledge on these individual differences and the factors that are associated with those differences (e.g., intellectual functioning, SES, physical symptoms, etc.), which was beyond the scope of the current research. In addition, future research can provide more insight into the developmental trajectory of the language skills of children with 22q11DS. Although our results suggest comparably severe weaknesses in both preschool and school-age groups, a limitation of the current research is the lack of a TD control group in the school-age study as well as the small sample size in this age group. Moreover, the cross-sectional nature of our research does not allow us to draw conclusions about children's developmental trajectories. There is a strong need for longitudinal research on the language impairment of children with 22q11DS in comparison to TD peers, particularly as previous work suggested an increasing severity of receptive language impairment with age (Van den Heuvel et al., 2018) and in light of the observation that intellectual functioning declines during childhood and adolescence in 22q11DS (e.g., Fiksinski et al., 2022).

The current study showed substantial overlap between children with 22q11DS and children with DLD in terms of expressive grammatical skills, as evidenced by both standardized language assessment and spontaneous language analysis. Given inherent differences between children with 22q11DS and children with DLD, this overlap has important theoretical implications. Neither the large differences in intellectual functioning and co-occurring physical symptoms, nor the presence or absence of a known genetic condition, seems to result in differences in the expressive grammatical skills of these two groups of children. Our findings thereby correspond to other studies that showed more commonalities than differences in the grammatical skills of etiologically diverse groups of children (e.g., Bloom and Lahey, 1978; Bol and Kuiken, 1990; Laws and Bishop, 2004; Bol and Kasparian, 2009), and support the consensus among professionals on this topic (Bishop et al., 2016). It appears that there are multiple routes toward impaired grammar development with similar, or even virtually identical, phenotypic characteristics. The shared phenotypic characteristics of children's expressive grammar could be hypothesized to reflect, at least in part, simplification processes that are typical for earlier stages of development. In other words, if acquiring or using grammatical rules is, for whatever reason, difficult, there are common ways to make it easier. The current study was, however, not set up to test this hypothesis and was limited by the use of standardized tests and spontaneous language samples. Comparative research on language impairment in etiologically diverse groups, preferably with experimental designs (see e.g., Perovic et al., 2013), is needed to understand the observed commonalities and differences in children's language profiles.

As mentioned, the current study did not only find similarities in the language profiles of children with 22q11DS and children with DLD. Receptive language difficulties were more severe in children with 22q11DS, showing that, despite overlap, different disorders have their own profile of relative strengths and weaknesses (e.g., Rice et al., 2005; Fidler et al., 2007). Given the poor prognosis of children with receptive language problems (e.g., Snowling et al., 2006; Zambrana et al., 2014) and the uncertainty about the effectiveness of therapy in this group (Law et al., 2003), special attention to these problems in children with 22q11DS is warranted in both research and clinical care. A possible avenue for future research would be to compare children with 22q11DS to a subgroup of children with DLD who both have expressive and receptive language problems. This can provide further insight into the mechanisms underlying (impaired) language development, for example enhancing our knowledge on the relation between low intellectual functioning and receptive language problems. It is also of clinical relevance, as children with 22q11DS and children with DLD often get language support in similar services, such as speech-language therapy and special education (see Boerma et al., 2022). The overlap in expressive grammar of the two groups of children may offer professionals working with children with 22q11DS a starting point for setting therapy goals in the domain of grammar. Moreover, it may even suggest that expressive grammar interventions targeting children with DLD also benefit children with 22q11DS. Although studies directly investigating the effectiveness of interventions in 22q11DS are a crucial next step, a subgroup comparison with children with DLD who have both expressive and receptive language problems could furthermore inform professionals about the usefulness of receptive language interventions with children with 22q11DS.

4.4. Conclusion

The current study is the first to investigate grammatical accuracy and complexity in the spontaneous language of children with 22q11DS. Complementing spontaneous language analysis with standardized testing in preschool and school-aged children, we showed weak expressive grammar in both naturalistic as well as standardized test settings, thereby contributing to a more complete description of the language profile of children with 22q11DS. The expressive grammatical skills of the children with 22q11DS did not differ from those of children with DLD, despite clear differences between the two groups in the presence or absence of known etiology and accompanying cognitive and physical challenges. This overlap indicates that expressive grammar may be a shared and significant vulnerability across different populations that can further our knowledge of the mechanisms underlying language acquisition and that can improve clinical care for children such as those with 22q11DS. The observed weaker receptive language skills of the children with 22q11DS compared to the children with DLD show that different disorders are associated with a unique language profile of strengths and weaknesses. It is an open question whether the differences in receptive language are related to factors which inherently differentiate the 22q11DS and DLD groups.

Data availability statement

The datasets presented in this article are not readily available because the datasets generated and/or analyzed for the current study are not publicly available due to GDPR compliance as well as legal and ethical limitations. A limited amount of data can be shared by the corresponding author upon reasonable request. Requests to access the datasets should be directed to t.d.boerma@uu.nl.

Ethics statement

The studies involving human participants were reviewed and approved by Medical Research Ethics Review Board of the University Medical Center Utrecht. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Author contributions

TB and FW: conceptualization. TB, EE, DV, MS, and IS: data curation. TB, EE, and IS: investigation and project administration. TB: formal analysis and writing–original draft. FW: funding acquisition and supervision. TB, EE, DV, MS, IS, EG, MH, JV, and FW: writing–review and editing. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Netherlands Organization for Scientific Research (NWO) [project number 360-89-080] and Utrecht University's Strategic Research Theme Dynamics of Youth.

Acknowledgments

We would like to thank all children, parents, professionals, and schools for their participation. We thank Fenna Duijnkerke and Marieke Huls for their support with data collection, and Mariska van Steensel for her role in the EPISODE study. Finally, we acknowledge the support of Stichting Steun 22Q11, Royal Auris, Royal Kentalis, VierTaal, and NSDSK.

Conflict of interest

JV has served as a consultant for NoBias Therapeutics.

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.

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

Bishop, D. V. (2009). Genes, cognition, and communication: insights from neurodevelopmental disorders. *Ann. N. Y. Acad. Sci.* 1156, 1–18. doi: 10.1111/j.1749-6632.2009.04419.x

Bishop, D. V., Snowling, M. J., Thompson, P. A., Greenhalgh, T., and Catalise Consortium. (2016). CATALISE: a multinational and multidisciplinary Delphi consensus study. Identifying language impairments in children. *PLoS ONE* 11, e0158753. doi: 10.1371/journal.pone.0158753

Bishop, D. V., Snowling, M. J., Thompson, P. A., Greenhalgh, T., and Catalise-2 Consortium. (2017). 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

Blagojevic, C., Heung, T., Theriault, M., Tomita-Mitchell, A., Chakraborty, P., Kernohan, K., et al. (2021). Estimate of the contemporary live-birth prevalence of recurrent 22q11.2 deletions: a cross-sectional analysis from population-based newborn screening. *Can. Med. Assoc. Open Access J.* 9, E802–E809. doi: 10.9778/cmajo.20200294

Blom, E., Boerma, T., and de Jong, J. (2020). Multilingual assessment instrument for narratives (MAIN) adapted for use in Dutch. *ZAS Papers Ling.* 64, 51–56. doi: 10.21248/zaspil.64.2020.557

Bloom, L., and Lahey, M. (1978). *Language Development and Language Disorders*. New York, NY: Wiley.

Boerma, T., Selten, I., Everaert, E., Gerrits, E., and Wijnen, F. (2022). *Onbekend maakt onbegrepen: meer weten over het 22q11.2 deletiesyndroom. Van Horen Zeggen*. Available online at: https://vhz-online.nl/onbekend-maakt-onbegrepen-meer-weten-over-het-22q11-2-deletiesyndroom (accessed November 10, 2022).

Boerma, T., Wijnen, F., Leseman, P., and Blom, E. (2017). Grammatical morphology in monolingual and bilingual children with and without language impairment: the case of Dutch plurals and past participles. *J. Speech Lang. Hear. Res.* 60, 2064–2080. doi: 10.1044/2017_JSLHR-L-16-0351

Bol, G., and Kuiken, F. (1990). Grammatical analysis of developmental language disorders: a study of the morphosyntax of children with specific language disorders, with hearing impairment and with Down's syndrome. *Clin. Linguist. Phon.* 4, 77–86. doi: 10.3109/02699209008985472

Bol, G. W., and Kasparian, K. (2009). The production of pronouns in Dutch children with developmental language disorders: a comparison between children with SLI, hearing impairment, and Down's syndrome. *Clin. Linguist. Phon.* 23, 631–646. doi: 10.1080/02699200902995677

Bol, G. W., and Kuiken, F. (1988). *Grammaticale analyse van taalontwikkelingsstoornissen*. Utrecht: Elinkwijk (Doctoral Dissertation, University of Amsterdam).

Calder, S. D., Brennan-Jones, C. G., Robinson, M., Whitehouse, A., and Hill, E. (2022). The prevalence of and potential risk factors for developmental language disorder at 10 years in the raine study. *J. Paediatr. Child Health* 58, 2044–2050. doi: 10.1111/jpc.16149

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences, 2nd ed*. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Costanza-Smith, A. (2010). The clinical utility of language samples. *Perspect. Lang. Learn. Educ.* 17, 9–15. doi: 10.1044/lle17.1.9

De Jong, J. (1999). *Specific Language Impairment in Dutch: Inflectional Morphology and Argument Structure*. Doctoral Dissertation, University of Groningen, The Netherlands.

Dennis, M., Francis, D. J., Cirino, P. T., Schachar, R., Barnes, M. A., and Fletcher, J. M. (2009). Why IQ is not a covariate in cognitive studies of neurodevelopmental disorders. *J. Int. Neuropsychol. Soc.* 15, 331–343. doi: 10.1017/S1355617709090481

Doedens, W. J., and Meteyard, L. (2022). What is functional communication? a theoretical framework for real-world communication applied to aphasia rehabilitation. *Neuropsychol. Rev.* 32, 937–73. doi: 10.1007/s11065-021-09531-2

Everaert, E., Selten, I., Boerma, T., Houben, M., Vorstman, J., de Wilde, H., et al. (2022). The language profile of preschool children with 22q11.2 deletion syndrome and the relationship with speech intelligibility. *Am. J. Speech-Lang. Pathol.* 32, 128–144. doi: 10.1044/2022_AJSLP-21-00328

Fidler, D. J., Philofsky, A., and Hepburn, S. L. (2007). Language phenotypes and intervention planning: bridging research and practice. *Ment. Retard. Dev. Disabil. Res. Rev.* 13, 47–57. doi: 10.1002/mrdd.20132

Fiksinski, A. M., Bearden, C. E., Bassett, A. S., Kahn, R. S., Zinkstok, J. R., Hooper, S. R., et al. (2022). A normative chart for cognitive development in a genetically selected population. *Neuropsychopharmacology* 47, 1379–1386. doi: 10.1038/s41386-021-00988-6

Gagarina, N. V., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiuniene, I., et al. (2012). MAIN: multilingual assessment instrument for narratives. *ZAS Papers Ling.* 56, 155–155. doi: 10.21248/zaspil.56.2019.414

Gerdes, M., Solot, C., Wang, P. P., McDonald-McGinn, D. M., and Zackai, E. H. (2001). Taking advantage of early diagnosis: preschool children with the 22q11.2 deletion. *Genet. Med.* 3, 40–44. doi: 10.1097/00125817-200101000-00009

Gerdes, M., Solot, C., Wang, P. P., Moss, E., LaRossa, D., Randall, P., et al. (1999). Cognitive and behavior profile of preschool children with chromosome 22q11.2 deletion. *Am. J. Med. Genet.* 85, 127–133.

Gerrits, E., Beers, M., Bruinsma, G., and Singer, I. (2017). *Handboek taalontwikkelingsstoornissen*. Bussum: Coutinho.

Glaser, B., Mumme, D. L., Blasey, C., Morris, M. A., Dahoun, S. P., Antonarakis, S. E., et al. (2002). Language skills in children with velocardiofacial syndrome (deletion 22q11.2). *J. Pediatr.* 140, 753–758. doi: 10.1067/mpd.2002.124774

Heilmann, J. (2010). Myths and realities of language sample analysis. *Perspect. Lang. Learn. Educ.* 17, 4–8. doi: 10.1044/lle17.1.4

Hunt, K. W. (1970). Syntactic maturity in school children and adults. *Monograph. Soc. Res. Child Develop.* 35, 1–61.

IBM Corp (2021). *IBM SPSS Statistics for Windows*. Version 28.0. Armonk: IBM Corp.

Kambanaros, M., and Grohmann, K. K. (2017). Linguistic and nonverbal abilities over time in a child case of 22q11 deletion syndrome. *Biolinguistics* 11, 57–82. doi: 10.5964/bioling.9077

Klatte, I. S., Van Heugten, V., Zwitserlood, R., and Gerrits, E. (2022). Language sample analysis in clinical practice: speech-language pathologists' barriers, facilitators, and needs. *Lang. Speech Hear. Serv. Sch.* 53, 1–16. doi: 10.1044/2021_LSHSS-21-00026

Kort, W., Compaan, E., Schittekatte, M., and Dekker, P. (2008). *Clinical Evaluation of Language Fundamentals*. Fourth edition: Dutch version. Amsterdam: Pearson Assessment and Information B.V.

Krok, W. C., and Leonard, L. B. (2015). Past tense production in children with and without specific language impairment across Germanic languages: a meta-analysis. *J. Speech Lang. Hear. Res.* 58, 1326–1340. doi: 10.1044/2015_JSLHR-L-14-0348

Law, J., Garrett, Z., and Nye, C. (2003). Speech and language therapy interventions for children with primary speech and language delay or disorder. *Cochrane Database Syst. Rev.* 2003, CD004110. doi: 10.1002/14651858.CD004110

Laws, G., and Bishop, D. V. (2004). Verbal deficits in Down's syndrome and specific language impairment: a comparison. *Int. J. Lang. Commun. Disord.* 39, 423–451. doi: 10.1080/13682820410001681207

Leonard, L. B. (2014). *Children With Specific Language Impairment*. Cambridge, CA: MIT Press.

MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk*. 3rd Edition. Mahwah, NJ: Lawrence Erlbaum Associates.

McDonald-McGinn, D. M., Sullivan, K. E., Marino, B., Philip, N., Swillen, A., Vorstman, J. A. S., et al. (2015). 22q11.2 deletion syndrome. *Nat. Rev. Dis. Primers* 1, 15071. doi: 10.1038/nrdp.2015.71

Miller, G. A., and Chapman, J. P. (2001). Misunderstanding analysis of covariance. *J. Abnorm. Psychol.* 110, 40. doi: 10.1037/0021-843X.110.1.40

Moss, E. M., Batshaw, M. L., Solot, C. B., Gerdes, M., McDonald-McGinn, D. M., Driscoll, D. A., et al. (1999). Psychoeducational profile of the 22q11.2 microdeletion: a complex pattern. *J. Pediatr.* 134, 193–198. doi: 10.1016/S0022-3476(99)70415-4

Norbury, C. F., Gooch, D., Wray, C., Baird, G., Charman, T., Simonoff, E., et al. (2016). The impact of nonverbal ability on prevalence and clinical presentation of language disorder: evidence from a population study. *J. Child Psychol. Psychiatry* 57, 1247–1257. doi: 10.1111/jcpp.12573

Perovic, A., Modyanova, N., and Wexler, K. (2013). Comparison of grammar in neurodevelopmental disorders: the case of binding in Williams syndrome and autism with and without language impairment. *Lang. Acquis.* 20, 133–154. doi: 10.1080/10489223.2013.766742

Persson, C., Niklasson, L., Óskarsdóttir, S., Johansson, S., Jönsson, R., and Söderpalm, E. (2006). Language skills in 5–8-year-old children with 22q11 deletion syndrome. *Int. J. Lang. Commun. Disorders* 41, 313–333. doi: 10.1080/13682820500361497

Polišenská, K., Chiat, S., and Roy, P. (2015). "Sentence repetition: what does the task measure?." *Int. J. Lang. Commun. Disord.* 50, 106–118. doi: 10.1111/1460-6984.12126

Price, L. H., Hendricks, S., and Cook, C. (2010). Incorporating computer-aided language sample analysis into clinical practice. *Lang. Speech Hear. Serv. Schools* 41, 206–222. doi: 10.1044/0161-1461(2009/08-0054)

Rakonjac, M., Cuturilo, G., Stevanovic, M., Jelicic, L., Subotic, M., Jovanovic, I., et al. (2016). Differences in speech and language abilities between children with 22q11.2 deletion syndrome and children with phenotypic features of 22q11.2 deletion syndrome but without microdeletion. *Res. Dev. Disabil.* 55, 322–329. doi: 10.1016/j.ridd.2016.05.006

Rice, M. L., Warren, S. F., and Betz, S. K. (2005). Language symptoms of developmental language disorders: an overview of autism, down syndrome, fragile X, specific language impairment, and Williams syndrome. *Appl. Psycholinguist.* 26, 7–27. doi: 10.1017/S0142716405050034

Riches, N. G. (2012). Sentence repetition in children with specific language impairment: an investigation of underlying mechanisms. *Int. J. Lang. Commun. Disord.* 47, 499–510. doi: 10.1111/j.1460-6984.2012.00158.x

Roizen, N. J., Antshel, K. M., Fremont, W., AbdulSabur, N., Higgins, A. M., Shprintzen, R. J., et al. (2007). 22q11.2DS deletion syndrome: developmental milestones in infants and toddlers. *J. Develop. Behav. Pediatrics* 28, 119–124. doi: 10.1097/01.DBP.0000267554.96081.12

Scherer, N. J., D'Antonio, L. L., and Kalbfleisch, J. H. (1999). Early speech and language development in children with velocardiofacial syndrome. *Am. J. Med. Genetics* 88, 714–723.

Schlichting, L. (2005). *Peabody Picture Vocabulary Test-III-NL*. Pearson Benelux BV.

Schneider, M., Debbané, M., Bassett, A. S., Chow, E. W., Fung, W. L. A., Van Den Bree, M. B., et al. (2014). Psychiatric disorders from childhood to adulthood in 22q11. 2 deletion syndrome: results from the International Consortium on Brain and Behavior in 22q11. 2 Deletion Syndrome. *Am J Psychiatry* 171, 627–639. doi: 10.1176/appi.ajp.2013.13070864

Selten, I. S., Boerma, T. D., Everaert, E., Vansteensel, M. J., Vorstman, J. A. S., and Wijnen, F. N. K. (2021). Narrative comprehension and production abilities of children with 22q11.2 deletion syndrome. *Res. Dev. Disabil.* 199, 104109. doi: 10.1016/j.ridd.2021.104109

Snowling, M. J., Bishop, D. V. M., Stothard, S. E., Chipchase, B., and Kaplan, C. (2006). Psychosocial outcomes at 15 years of children with a preschool history of speech-language impairment. *J. Child Psychol. Psychiatry* 47, 759–765. doi: 10.1111/j.1469-7610.2006.01631.x

Solot, C. B., Gerdes, M., Kirschner, R. E., McDonald-McGinn, D. M., Moss, E., Woodin, M., et al. (2001). Communication issues in 22q11.2 deletion syndrome: children at risk. *Genet. Med.* 3, 67–71. doi: 10.1097/00125817-200101000-00015

Solot, C. B., Knightly, C., Handler, S. D., Gerdes, M., Mcdonald-Mcginn, D. M., Moss, E., et al. (2000). Communication disorders in the 22q11.2 microdeletion syndrome. *J. Commun. Disord.* 33, 187–204. doi: 10.1016/S0021-9924(00)00018-6

Solot, C. B., Sell, D., Mayne, A., Baylis, A. L., Persson, C., Jackson, O., et al. (2019). Speech-language disorders in 22q11.2 deletion syndrome: best practices for diagnosis and management. *Am. J. Speech Lang. Pathology* 28, 984–999. doi: 10.1044/2019_AJSLP-16-0147

Stichting Siméa (2017). *Richtlijn toelaatbaarheid*. Available online at: https://simea.nl/media/richtlijnen/simea-brochure-richtlijn-toelaatbaarheid-20170901.pdf (accessed November 10, 2022).

Tomblin, J. B., Records, N. L., Buckwalter, P., Zhang, X., Smith, E., and O'Brien, M. (1997). Prevalence of specific language impairment in kindergarten children. *J. Speech Lang. Hear. Res.* 40, 1245–1260. doi: 10.1044/jslhr.4006.1245

Van den Heuvel, E., Manders, E., Swillen, A., and Zink, I. (2018). Atypical language characteristics and trajectories in children with 22q11.2 deletion syndrome. *J. Commun. Disord.* 75, 37–56. doi: 10.1016/j.jcomdis.2018.06.001

Van den Heuvel, E., Reuterskiöld, C., Solot, C., Manders, E., Swillen, A., and Zink, I. (2017). Referential communication abilities in children with 22q11. 2 deletion syndrome. *Int. J. Speech-Lang. Pathol.* 19, 490–502. doi: 10.1080/17549507.2016.1221456

Van Steensel, M. J., Selten, I. S., Charbonnier, L., Berezutskaya, J., Raemaekers, M. A., Ramsey, N. F., et al. (2021). Reduced brain activation during spoken language processing in children with developmental language disorder and children with 22q11.2 deletion syndrome. *Neuropsychologia* 158, 107907. doi: 10.1016/j.neuropsychologia.2021.107907

Verhoeven, L., Steenge, J., and van Balkom, H. (2011). Verb morphology as clinical marker of specific language impairment: evidence from first and second language learners. *Res. Dev. Disabil.* 32, 1186–1193. doi: 10.1016/j.ridd.2011.01.001

Wetherell, D., Botting, N., and Conti-Ramsden, G. (2007). Narrative in adolescent specific language impairment (SLI): a comparison with peers across two different narrative genres. *Int. J. Lang. Commun. Disord.* 42, 583–605. doi: 10.1080/13682820601056228

Wexler, K., Schaeffer, J., and Bol, G. (2004). Verbal syntax and morphology in typically developing Dutch children and children with SLI: how developmental data can play an important role in morphological theory. *Syntax* 7, 148–198. doi: 10.1111/j.1467-9612.2004.00006.x

Wiig, E. H., Secord, W. A., and Semel, E. (2012). *CELF Preschool-2-NL: Clinical Evaluation of Language Fundamentals: Preschool-Nederlandstalige versie*. Handleiding. (J. De Jong, Trans). Pearson Benelux BV.

Zambrana, I. M., Pons, F., Eadie, P., and Ystrom, E. (2014). Trajectories of language delay from age 3 to 5: persistence, recovery and late onset. *Int. J. Lang. Commun. Disorders* 49, 304–316. doi: 10.1111/1460-6984.12073

Zwitserlood, R., van Weerdenburg, M., Verhoeven, L., and Wijnen, F. (2015). Development of morphosyntactic accuracy and grammatical complexity in Dutch school-age children with SLI. *J. Speech Lang. Hear. Res.* 58, 891–905. doi: 10.1044/2015_JSLHR-L-14-0015

Appendix

Examples of the error and complex utterance categories as coded in the spontaneous language analysis.

Parameter	Category	Examples
% Verb-related errors[a]	Argument omissions	∗*nu moet nog even wachten* [now have to wait] correct: nu moet je nog even wachten [now you have to wait]
	Subject-verb agreement errors	∗*ik komt* [I comes] correct: ik kom [I come]
	Tense errors	∗*toen krijg ik een verrekijker* [then I get a binocular] correct: toen kreeg ik een verrekijker [then I got a binocular]
	Root infinitives[b]	∗*jij varken tekenen* [you pig draw] correct: jij tekent (een/het) varken [you draw (a/the) pig]
	Verb-second placement errors	∗*waar deze moet?* [where this goes?] correct: waar moet deze? [where does this go?]
	Overgeneralizations	∗*hij vliegde weg* [he flied away] correct: hij vloog weg [he flew away]
	Past participle errors	∗*ik heb die voor mama maakt* [I have that for mama makes] correct: ik heb die voor mama gemaakt [I have made that for mama]
	Verb omissions	∗*deze niet rood* [this one not red] correct: deze is niet rood [this one is not red]
	Other verb-related errors	∗*hij moet deze dichtmaak* [he has to this one close] correct: hij moet deze dichtmaken [he has to close this one]
% Non-verb-related errors[a]	Determiner errors	∗*mag ik naar traktor kijken?* [can I look at tractor?] correct: mag ik naar de traktor kijken? [can I look at the tractor?]
	Adjectival inflection errors	∗*een grote ding* [a big thing] correct: een groot ding [a big thing]
	Preposition errors	∗*ook één jou* [also one you] Correct: ook één voor jou [also one for you]
	Pronoun errors	∗*naar mij huis* [to me house] correct: naar mijn huis [to my house]
	Conjunction errors	∗*hij kan open dicht* [he can open close] correct: hij kan open en dicht [he can open and close]
	Plural errors	∗*ik heb drie bos* [I have three forest] correct: ik heb drie bossen [I have three forests]
	Errors with 'er' [there]	∗*de dieren passen niet in* [the animals do not fit in] correct: de dieren passen er niet in [the animals do not fit in there]
	Word order errors	∗*het lijkt een hond op* [it looks a dog like] correct: het lijkt op een hond [it looks like a dog]
	Other non-verb-related errors	∗*ik wil die als jij* [I want that as you] correct: ik wil diezelfde als jij [I want the same as you]
% Complex utterances	Subordinate clauses	ik dacht dat ik een spelletje ging doen [I thought I was going to play a game]
	Conjunction reduction	de kat is bang en de hond boos [the cat is afraid and the dog angry]
	Direct speech	hij zegt: "ik wil slapen" [he says: "I want to sleep"]
	Infinitival clauses	kan je even helpen om dit aan elkaar te maken [can you help to tie this together]

[a]Categories may include different types of errors. For example, argument omissions include both subject and object omissions. With the exception of a number of categories that specify the type of error in the name (e.g., verb omissions), error categories can include omission and substitution errors (and in rare cases also addition errors). The given examples illustrate just one type of error within a specific error category.
[b]Root infinitives are clauses in which an infinitive is used as main predicate, although a finite verb is expected. In Dutch, the latter can only be determined with certainty when an overt subject is expressed. Therefore, this category only includes non-finite clauses with an overt subject. Previous research may have used less stringent operationalizations of root infinitives.

frontiers | Frontiers in Communication

TYPE Original Research
PUBLISHED 13 July 2023
DOI 10.3389/fcomm.2023.1153665

Check for updates

OPEN ACCESS

EDITED BY
Anja Gampe,
University of Duisburg-Essen, Germany

REVIEWED BY
Aylin C. Küntay,
Koç University, Türkiye
Antje Endesfelder Quick,
Leipzig University, Germany

*CORRESPONDENCE
Mariam Komeili
✉ mariam.komeili@atcspeec.com
Theodoros Marinis
✉ t.marinis@uni-konstanz.de

RECEIVED 29 January 2023
ACCEPTED 05 June 2023
PUBLISHED 13 July 2023

CITATION
Komeili M, Tavakoli P and Marinis T (2023) Using multiple measures of language dominance and proficiency in Farsi-English bilingual children. *Front. Commun.* 8:1153665.
doi: 10.3389/fcomm.2023.1153665

Using multiple measures of language dominance and proficiency in Farsi-English bilingual children

Mariam Komeili[1]*, Parvaneh Tavakoli[2] and Theodoros Marinis[3]*

[1]School of Psychology and Language Studies, University of Reading, Reading, United Kingdom, [2]Department of English Language and Applied Linguistics, University of Reading, Reading, United Kingdom, [3]Department of Linguistics, University of Konstanz, Konstanz, Germany

This paper aims to identify effective means of measuring dominance and proficiency in bilingual children. Thirty-seven Farsi-English Heritage language speaking children from 6;1 to 11;6 were assessed on their vocabulary, morphosyntax, and narrative microstructure skills in both languages to address whether there is a difference between their proficiency in Farsi as a heritage and English as a majority language, how the scores on the vocabulary, morphosyntax, and narrative microstructure tasks relate to one another, and based on the results of each task in both languages if any of the children are at risk for a Developmental Language Disorder. Vocabulary was assessed using the LITMUS-Cross-Linguistic Lexical Task (CLT), morphosyntax using the LITMUS-Sentence Repetition (SR) tasks, and Narrative microstructure using the LITMUS-Multilingual Assessment Instrument for Narratives (MAIN). Individual language proficiency was identified via an in-depth profile analysis for each participant who looked at their performance on all experimental tasks in both languages. The data demonstrated that on the vocabulary and narrative tasks the participants were more dominant in English than in Farsi, while on the sentence repetition task there were no significant differences between the two languages. Correlation analyses showed that vocabulary scores were strongly correlated to the sentence repetition scores and the microstructure scores. The English and Farsi sentence repetition scores also correlated moderately with the microstructure scores within each language. Profile analysis showed that no child within the study scored <1.5 or 2 standard deviations below the mean on more than two tasks in both languages. However, interesting patterns emerged indicating that some participants had a greater proficiency in one language vs. the other language. The results from this study showed that measuring language within a single domain (e.g., morphosyntax) is not enough to identify a bilingual child's language dominance and/or proficiency. Instead, an in-depth profile analysis and language assessments across various language domains need to be done in order to appropriately measure language dominance and proficiency. Consequently, this study supports the importance of measuring language across multiple domains in studies of bilingual children. The clinical significance of appropriately identifying language dominance and proficiency was also shown, as such information would allow clinicians to make more appropriate clinical decisions.

KEYWORDS

dominance, proficiency, bilingualism, heritage language, majority language, vocabulary, morphosyntax, narratives

1. Introduction

Bilingualism refers to the knowledge and command of two or more languages to different degrees (Montrul, 2008). Over the years researchers have identified multiple types of bilinguals, each with varying degrees of language input and output. Simultaneous bilingual children are exposed to two languages generally from birth or within the 1st year, while sequential bilinguals can either be defined as early or late sequential bilinguals (Unsworth, 2015; Montrul, 2016). The former refers to those who first acquire one language and then develop a second one before formal literacy in the first language has set in and late sequential bilinguals are defined as those in which the second language develops after literacy in the first language has begun, usually at around the age of 5 or 6 (Ortega, 2020). Over the past two decades the literature has begun to focus on another type of bilingual speaker known as *Heritage speakers* (*HSs*; Benmamoun et al., 2013). HSs are considered second or third generation immigrants where the minority language spoken in the home by their parents is the heritage language (HL) and the majority language (ML) is the language spoken at school and/or in the greater society. The ML generally eventually develops into the dominant language (Montrul, 2016). While our knowledge of bilingualism has increased significantly, questions remain as to how to test the language abilities of a bilingual child's two languages as well as how to identify language dominance and proficiency. It is crucial to be able to test children in both their languages and identify their dominance and proficiency for both educational and clinical/diagnostic purposes (Kohnert, 2010). The present study addresses dominance and proficiency using several measures of language skills that target expressive and receptive vocabulary, as well as morphosyntax and narrative microstructure in primary school Farsi-English bilingual children living in Canada. The study also aims to identify if any of the participants are at risk for developing a Developmental Language Disorder (DLD) by assessing the children' language skills in both their languages (Kohnert, 2010).

1.1. Dominance and proficiency

Questions about language dominance and proficiency and how these are defined have been discussed at length (Silva-Corvalán and Treffers-Daller's, 2015; Treffers-Daller's and Korybski, 2015). Language proficiency is based on the child's overall abilities in a language, while language dominance is measured based on how proficient a child is in a particular language in comparison to another one considering external factors such as exposure and use (Silva-Corvalán and Treffers-Daller's, 2015; Treffers-Daller's and Korybski, 2015). It is common to find that bilinguals are more dominant or proficient in one language. Even when acquiring both languages from birth simultaneously, children are generally more dominant in one vs. the other language. While dominance and proficiency are very often used as ways to distinguish between different acquisition patterns in children and to examine cross-linguistic interactions, there are conflicting views on their validity as useful explanatory constructs. The reason for this is because dominance and proficiency change constantly throughout the lifespan as a function of the amount of input and use of one's languages. A child's dominance level often shifts when they start school and start using the school language (Kohnert and Bates, 2002; Silva-Corvalán and Treffers-Daller's, 2015; Montrul, 2016). For example, Kohnert and Bates (2002) did a cross-sectional study of school age Spanish-English children and found that when they started learning English at school, their knowledge of English took over Spanish at around the age of 11; however, the production of English took over Spanish at around the age of 14 (Kohnert et al., 1998). What is important to note, is that the shift in dominance is a gradual process and occurs in different domains at different times (Bedore et al., 2012). Regardless, of the conflicting views, it is important to measure language dominance and proficiency in bilinguals, as this would allow for greater cross study comparisons (Luk and Bialystok, 2013; Silva-Corvalán and Treffers-Daller's, 2015; Treffers-Daller's and Korybski, 2015). How are dominance and proficiency measured?

Self-ratings by adults or parental ratings for children are common means of measuring language dominance and proficiency (Bedore et al., 2012). However, parental ratings are sometimes inaccurate because language proficiency is constantly changing and a parent might not interact with a child in both their languages in order to provide accurate data. Bedore et al. (2012) argue that more direct measures of language knowledge (e.g., analysis of language performance elicited by an oral narrative) would provide a more objective measure of language dominance and proficiency. In their study, they measured language dominance in a group of 1,192 Spanish-English 5 year old children to address if different measures of language experience and ability would lead to the same classification of language dominance. Therefore, they measured dominance via a parental questionnaire, as well as an English semantics and morphosyntax assessment and found that the measure used does matter. While the tests for semantics and morphosyntax were able to classify the children into different dominance categories, the test of semantics proved to be a stronger measure. A study by Meir (2018) also suggests that numerous quantitative measures can be used to identify the language dominance of a bilingual child. These include measures of mean length of utterance (MLU), directionality of code-switching, parental ratings, exposure patterns, vocabulary, and/or morphosyntax. It is important to note that a single measure is not sufficient to capture the multidimensionality of dominance and proficiency. Among scholars in the field, it is becoming increasingly prevalent to use more than one measure to identify language dominance and proficiency. Meir and Armon-Lotem (2017) and Meir (2018) measured language dominance in a group with Russian as a HL and Hebrew as the ML of the society. In both studies, they identified language dominance in the two groups through scores of language proficiency. Language proficiency in HL Russian children was measured using a battery of tasks which provided data on object naming, production of case, and verb inflection. Language proficiency scores for ML Hebrew were obtained via tests for expressive vocabulary, sentence repetition, sentence comprehension, expression, pronunciation, and storytelling.

10.3389/fcomm.2023.1153665

Although it has been shown that it is important to use multiple measures to measure language dominance and proficiency, many studies have used only lexical measures as a means to operationalize these concepts (e.g., Lambert et al., 1959; Fishman and Cooper, 1969; Cromdal, 1999; Bialystok et al., 2008; Reyes and Azuara, 2008; Silva-Corvalán and Treffers-Daller's, 2015; Treffers-Daller's and Korybski, 2015). Vocabulary is considered a good measure of language dominance for a variety of reasons. Firstly, vocabulary and grammar are known to be strongly related (Bates and Goodman, 1999; Tomasello, 2000). Secondly, lexical knowledge influences performance on online tasks (Bialystok et al., 2008). Finally, lexical knowledge is a significant prerequisite of academic achievement in both monolingual and bilingual children (Treffers-Daller's and Korybski, 2015).

The present study uses a range of measures to address dominance and proficiency in bilingual children. It also provides insights into the relationship between these measures within each language in a group of school–aged bilingual children who have Farsi as their HL and English as their ML. Such knowledge is important for clinicians, such as speech and language therapists, in terms of diagnosing a language disorder. Despite the critical role of measurement of dominance and proficiency in the identification of language disorders in bilingual populations, there is relatively limited research investigating these constructs separately in multiple domains across languages. This study aims to add to the literature by indicating that accurate measures of dominance and proficiency in bilingual children are significant for researchers and clinicians alike.

1.2. Language acquisition in the heritage language

HSs acquire their HL naturalistically from the home environment. HSs can be simultaneous bilinguals, acquiring both their HL and the ML in early childhood, or sequential bilinguals, acquiring the HL first before acquiring the ML when they start school (Rothman et al., 2016). Ultimately, the ML ends up becoming the stronger or more dominant language in adulthood (Montrul, 2016). Regardless of language dominance, language proficiency levels in the HL of the children varies greatly (Montrul, 2016). For example, some HSs can be highly fluent and literate in their HL, while others have very little expressive ability and are only able to understand the HL.

Due to the variability in their proficiencies, the grammatical systems of HSs vary significantly. However, less is known about their vocabulary skills. The vocabulary skills of HSs have not been studied as much as their morphosyntactic skills. This is problematic, as attrition affects the lexicon earlier and more significantly than it does so for morphosyntax (Gharibi and Boers, 2016). It is important to measure the vocabulary skills of HSs, as research shows that vocabulary and grammar are significantly correlated (Gharibi and Boers, 2016; Montrul, 2016; Hamann and Abed Ibrahim, 2017). In addition, research states that the vocabulary abilities of HSs tend to be lower than that of homeland speakers (Hoff and Core, 2013; Montrul, 2016). An example of this can be found in a study by Gharibi and Boers (2016). To investigate the vocabulary skills of both simultaneous and sequential HSs in comparison to their monolingual counterparts, they studied two groups of children: (1) Thirty Farsi-English bilinguals living in New Zealand and (2) Thirty monolingual Farsi children living in Iran. All participants were administered a receptive and a productive vocabulary task. The study showed that overall the monolinguals outperformed the bilinguals in both tasks. However, the gap for the simultaneous bilingual group was much greater. A study by Hamann and Abed Ibrahim (2017) also demonstrated that bilingual children lagged behind monolingual children in terms of receptive and productive vocabulary. Apart from vocabulary, HSs have often been found to fall behind their monolingual counterparts in certain areas of grammar (Montrul, 2008; Cabo and Rothman, 2012; Benmamoun et al., 2013; Thordardottir and Brandeker, 2013; Hamann and Abed Ibrahim, 2017). According to Benmamoun et al. (2013) HSs tend to keep the basic and core principles of grammatical systems (e.g., noun-verb distinctions). However, the aspects of syntax that involve higher levels of grammar (e.g., complex syntax) are often under-developed in HSs. Therefore, it is important to measure a child's language proficiency using language tasks which are able to provide information on a range of linguistic domains. Measuring language proficiency across language domains would provide a more accurate measure of a child's true language skills in each of their languages.

1.3. The relationship between vocabulary and morphosyntax

There is a vast amount of research looking at the relationship between vocabulary and grammar in both monolingual and bilingual children. However, less research has looked at different language domains in the same group of bilingual children (Jia et al., 2002; Chondrogianni and Marinis, 2011; Thordardottir and Brandeker, 2013). Jia et al. (2002) measured the association between the development of the L1 and L2 in Mandarin-English speaking young adults living in the US. The participants provided self-ratings of language proficiency for each language and were given tests to measure their vocabulary and morphosyntactic skills in both languages. The results showed that the participants who had better overall performance in the ML generally underperformed in the HL. Chondrogianni and Marinis (2011) looked at vocabulary and morphosyntax in the same group of bilingual school-age children with Turkish as their HL and English as their ML. The results showed that in their ML the children performed better on the tasks that targeted general comprehension of grammar and production of tense marking morphology, but they had a lower accuracy on the comprehension of single word vocabulary and (complex) morphosyntax (e.g., articles, passives, and wh-questions). This study was one of the few to examine different language domains at the same time in the same population of bilingual children, but did not measure the children's HL. Thordardottir and Brandeker (2013) compared the vocabulary and morphosyntax of a group of 5-year-old English and French monolingual children as well as a group of English-French bilingual children with varying degrees of exposure. The children were assessed via a parental questionnaire as well as on non-word repetition, Sentence Repetition (SR), and vocabulary tasks. The monolingual children were tested in their native language while the bilingual children were tested in both English and French on all tasks. The study showed that SR scores were positively correlated with the vocabulary scores within languages, but no correlations were found between the two domains across languages.

The literature looking at language dominance and proficiency with respect to narrative microstructure has shown that children generally performed better in the majority language in comparison to the heritage language (Bohnacker, 2016; Méndez et al., 2018). The literature has also shown that children are required to reach a threshold level of vocabulary in order to be able to produce appropriate story narratives. Méndez et al. (2018) demonstrated that lexical abilities are highly associated to complexity measures in a bilingual child's languages during narrative retell tasks. Méndez et al. (2018) also indicated that vocabulary is a strong predictor of narrative skills in the majority language but not in the minority language.

TABLE 1 Demographic information on the study's participants.

	Mean	Std. deviation	Min–max
Age at Testing (months)	103.81	14.03	73–139
Length of Exposure English (months)	63.27	25.51	8–109
Length of Exposure Farsi (months)	99.91	15.23	64–125
Total Use of English (%)	29.59	18.48	0–90
Total Use of Farsi (%)	74.45	19.68	20–100
Richness score English (score out of 18)	10.27	1.85	6–14
Richness score Farsi (score out of 18)	6.10	2.39	0–13
Total Parental Education (raw score in years)	17.87	2.62	13.50–23

1.4. The present study

The present study adds to the research on dominance and proficiency by focusing on the relationship between language domains—vocabulary, morphosyntax, and oral narratives in bilingual children with typical development. Assessing lexical and morphosyntactic skills in both languages will allow direct information to be obtained about language dominance and proficiency from multiple domains. In addition, evaluating domains separately in the same group of children will allow for differentiation between the two languages and a greater understanding of how each one develops individually.

The study has the following research questions:

1) Is there a difference between the children's proficiency in Farsi as a HL and English as a ML as measured by vocabulary, morphosyntax, and narrative microstructure?
2) How do the scores on the vocabulary, morphosyntax, and narrative microstructure tasks relate to one another?
3) Based on the results of each task in both languages, are any of the children at risk for a Developmental Language Disorder?

2. Methodology

2.1. Participants

A total of 38 heritage Farsi and majority English school aged children between the ages of 6;1 to 11;6 participated in the study. All children were living in Toronto, Canada at the time of testing and attended mainstream English language schools during the week and a Farsi Saturday school for 8 h each week. During their Farsi school they were taught Farsi reading, writing, and math and were required to speak Farsi throughout the day. Thirteen of the 38 children were also exposed to French to some degree, but exposure to this language was only a few hours at school. All children were exposed to Farsi before the age of 2 and all children had at least 2 years of exposure to English. One family was not able to provide demographic information and therefore results are only presented for 37 participants.

The Questionnaire for Parents of Bilingual Children (PABIQ; Tuller, 2015) was used to collect background and demographic data on the participants of the study. Table 1 above provides demographic information in terms of Age at Testing (AaT), Length of Exposure (LoE), Total Use of each language in the home, Richness of each language obtained in the home, and Parental Education. Note that Total use is measured based on how much of a language the children were using with family members in the home, while Language Richness is based more on the level of engagement in watching TV, reading books, and storytelling, as well as the children's interaction with friends and family in the community.

When multiple languages are being used in a home, their use is often not balanced. This is shown in the study population by comparing the scores of Total Use in the home and Language Richness score. The results demonstrated that for the measure of Total Use in the home the children were more dominant in Farsi ($M = 74.45$) than English ($M = 29.59$). In contrast, the Language Richness score showed that the children were more dominant in English ($M = 10.27$) than Farsi ($M = 6.10$). In terms of socio-economic status based on years of education, both parents of all children had attended college or university, thus, putting them on middle to upper socio-economic status.

2.2. Tasks

The lexical task used in the study was the Cross-Linguistic Lexical Task (CLT; Haman et al., 2015). The CLT is made up of four tests—noun production, verb production, noun comprehension, verb comprehension and therefore allows for an overall measure of receptive and expressive vocabulary. In the current study, the British CLT (Haman et al., 2015) and the Farsi CLT (Talabi, 2018) were used. Both the British and Farsi CLTs have 32 items per test. The Farsi CLT was originally designed with pictures considered appropriate for Farsi children living in Iran but some pictures were deemed unethical (i.e., gun, knife etc.) for heritage Farsi speakers in western communities. Therefore, one picture per sub-test was removed and a final adapted Farsi version with 31 items per sub test was used in the study. Subtests in both the English and Farsi CLT were counterbalanced during administration. The CLT tasks took about 15–20 min to complete. Administration of the CLT involved having the participants see the pictures on a computer screen and responses were transcribed during the task administration. Two practice items were presented at the start of each task to familiarize the participants with the tasks.

In addition to the lexical tasks, two sentence repetition (SR) tasks, each with 30 items, were used: the English LITMUS-SR-30 (Marinis and Armon-Lotem, 2015) and the Farsi LITMUS-SR-30 (Komeili et al., 2020). Both tasks were administered via a paper version in which the items were presented orally. This is considered a more clinically valid and realistic means of assessment, as it allows for greater rapport with the children. The participants were required to listen to the sentences and repeat them back verbatim. At the start of the SR tasks two practice sentences were provided to help the children understand how the task worked. All participants' responses were audio recorded and transcribed at a later date. The SR tasks were scored for accuracy: children were given a score of 1 if the sentence was repeated exactly as they were said and a score of 0 if one or more errors were made. Interrater reliability by a trained rater was done for both tasks and was found to be 90.7% for the LITMUS-SR English and 91.1% for the LITMUS-SR Farsi. Each task took about 15–20 min to complete.

The last tasks used in this study were the English and Farsi LITMUS-MAIN narrative tasks (Gagarina et al., 2012, 2019). The participants were assessed on their microstructure skills during a retelling task. Participants were told one story in Farsi and one in English (counterbalancing took place across sessions and participants) and were then asked to retell the story in their own words while looking at the pictures. During re-telling the children were not shown the entire picture strip and instead were shown the pictures in sections (i.e., first pictures 1 and 2, then pictures 3 and 4 and then pictures 5 and 6) and were asked to re-tell the story as it progressed. Prompts were provided when needed to obtain further information from the participants. Responses were audio recorded for later transcription.

The LITMUS-MAIN does not provide a specific scoring outline for microstructure. However, guidelines are given which suggest that a measure of complexity and length should be considered (Norris and Ortega, 2009). We chose to employ an analysis method which is widely used in second language acquisition research to analyze spoken data of second language learners. The framework, known as the CAF method (Housen and Kuiken, 2009), measures complexity, fluency, and accuracy of language performance and is believed to be a reliable indicator of proficiency. To obtain measures of complexity, accuracy, and fluency, the children's narratives were transcribed and broken down into Analysis of Speech Units (AS-Units). An AS unit is a syntactic unit similar to the C-Unit which is made up of the main clause and all of its subordinate clauses (Foster et al., 2000). The AS unit is considered an "idea unit" and clausal boundaries are identified based on intonation and pauses. In the present study two scores for each CAF component were identified. The complexity score was based on two valid measures—ratio of subordination (RS) and number of words per clause (NWC). These are known as valid measures of complexity in the literature (Norris and Ortega, 2009) and were also part of those suggested by the MAIN guidelines. The RS was obtained by dividing the total number of clauses by the total number of AS units, while the NWC was obtained by dividing the total number of words by the total number of clauses. There were also two measures for accuracy—percentage of error free clauses (EFC) and percentage of correct verbs (TVC). For the first measure the total number of EFC was divided by the total number of clauses and a percentage was taken. Any error, for example phonological, lexical, grammatical, and morphological, impeding communication were considered. In addition, using a word from a different language (i.e., code switched) in a clause during the re-telling was also considered an error in this context, as the task instructions asked the children to complete each task in a single language (i.e., in monolingual mode). The second accuracy measure of TVC was obtained by taking the number of correct verbs as a proportion of the total number of verbs in the retelling. A correct verb was a verb with no semantic, morphosyntactic, or ordering errors. These measures are reported to be reliable indices of accuracy tapping into aspects of accuracy at global and local levels (Norris and Ortega, 2009; Tavakoli, 2018). Following the literature in this area (Tavakoli and Wright, 2020), fluency was measured in terms of speed, number of syllables per minute (SPM), and breakdown, number of filled pauses (NFP) per minute. Obtaining a measure of fluency could potentially demonstrate if there is any relationship between the children's vocabulary skills and their overall fluency when re-telling a narrative. The transcription, coding and scoring for both the English and Farsi LITMUS-MAIN followed the same procedure. Inter-rater reliability for transcription, coding, and scoring were obtained for 20% (seven narratives) of the data for all three components separately. Inter-rater reliability for complexity, accuracy and fluency in the Farsi transcription was 96.96, 93.86, and 98.9%, respectively and 98.35, 96.24, and 96.59% in the English transcriptions, respectively.

3. Results

3.1. Proficiency as measured by vocabulary

Descriptive data were first obtained and then a set of statistical analyses were conducted. The mean percentage and the standard deviations for the noun and verb comprehension and production tasks in English are illustrated in Figure 1A while those for the Farsi task are presented in Figure 1B. Percentages were used in the ANOVAs that address research question 1 while the raw scores were used in the correlation analyses for research questions 2 and 3.[1]

A repeated measures ANOVA was conducted with percentage scores to identify if there were any main effects of language,

1 In addition to the analyses presented, we conducted ANOVAs and partial correlations with age as an factor because the age range of the children was large. The results of these analyses were similar to the ones presented here.

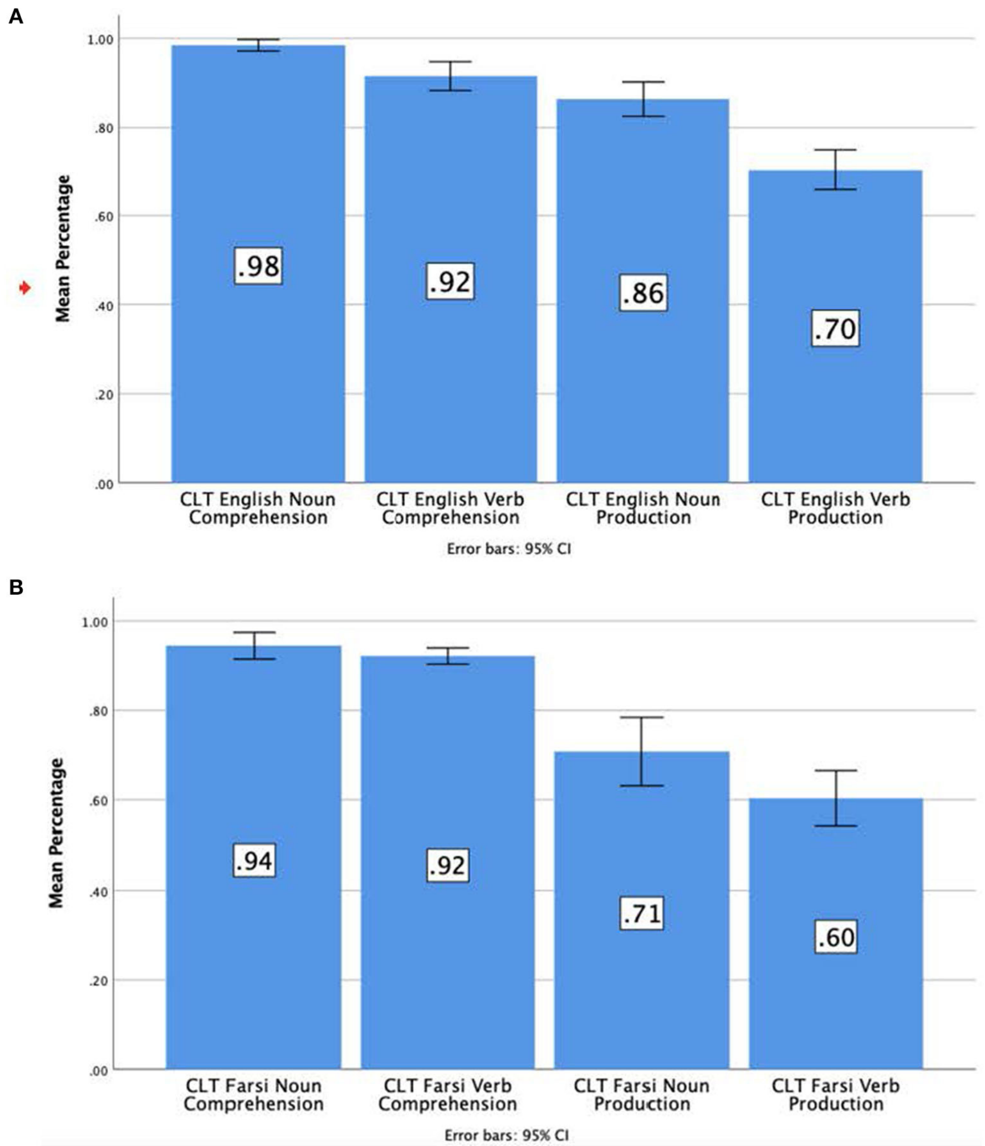


FIGURE 1
(A, B) Means of CLT percentage scores for the English and Farsi noun and verb comprehension and production tasks.

modality, or word type and if there were any interactions between these factors. The 2 × 2 × 2 repeated factors ANOVA with the factors Language (English and Farsi), Modality (comprehension and production), and Word type (noun and verb) showed significant main effects for all three factors: Language: $F_{(1,36)} = 6.666$ $p = 0.14$ $\eta^2 = 0.156$; Modality: $F_{(1,36)} = 258.956$, $p < 0.001$, $\eta^2 = 0.878$; and Word Type: $F_{(1,36)} = 108.395$, $p < 0.001$, $\eta^2 = 0.751$. Significant interactions were further observed between Language and Modality $F_{(1,36)} = 11.936$, $p = 0.001$, $\eta^2 = 0.249$; Language and Word Type $F_{(1,36)} = 10.284$ $p = 0.03$, $\eta^2 = 0.222$; and Modality and Word Type $F_{(1,36)} = 52.920$, $p < 0.001$, $\eta^2 = 0.595$. However, the three way interaction between Language, Modality and Word Type was not significant $F_{(1,36)} = 0.113$, $p = 0.738$, $\eta^2 = 0.003$. *Post-hoc* analyses in the form of pairwise comparisons were completed to identify where differences lay. The *post-hoc* analyses showed that participants performed significantly better in English than in Farsi in the production of both nouns and verbs [nouns: $t_{(36)} = 3.234$, $p = 0.003$; verbs: $t_{(36)} = 2.441$, $p = 0.02$]. However, in terms of comprehension they performed significantly better in English than in Farsi only in nouns [nouns: $t_{(36)} = 2.391$, $p = 0.022$; verbs: $t_{(36)} = -0.341$, $p = 0.725$]. These results show that the participants were more proficient in English than in Farsi both in terms of expressive and receptive vocabulary. However, the greater proficiency was mainly due to expressive language, as significant differences in production between the languages were found for both nouns and verbs while differences in comprehension were only found between the English and Farsi nouns. For this reason, in the correlation analyses between tasks in Section 3.4 we used for the CLT tasks a composite score of expressive vocabulary together for nouns and verbs.

3.2. Proficiency as measured by SR tasks

In order to further investigate the language proficiency of the children in terms of their morpho-syntactic abilities, analyses were conducted on the English and Farsi LITMUS-SR tasks. Descriptive data for the LITMUS-SR tasks are presented in Table 2.

A repeated measures ANOVA on the accuracy scores of the sentence repetition task with the factor Language with two levels (English and Farsi) did not show a significant difference between the two languages $F_{(1,36)} = 0.070$, $p = 0.793$, $\eta^2 = 0.002$.

3.3. Proficiency as measured by narrative microstructure

The descriptive statistics of the measures of narrative microstructure are presented in Tables 3–5 for each of the different microstructure components: complexity, fluency, and accuracy each for English and Farsi.

A set of six repeated measures ANOVAs were done on the LITMUS-MAIN microstructure measures, one per score with the factor Language (English and Farsi). There was a main effect of language with the scores in the English task significantly higher than in the Farsi task in the dependent Variables WPC, $F_{(1,36)} = 128.468$, $p = 0.001$, $\eta^2 = 0.781$; EFC, $F_{(1,36)} = 6.064$, $p = 0.019$, $\eta^2 = .144$; and TVC, $F_{(1,36)} = 5.448$, $p = 0.025$, $\eta^2 = 0.131$. The number of filled pauses was significantly higher in Farsi than English: NFP, $F_{(1,36)} = 17.031$, $p < 0.001$, $\eta^2 = 0.321$. Non-significant differences were found for RS, $F_{(1,36)} = 1.556$, $p = 0.220$, $\eta^2 = 0.041$ and for SPM, $F_{(1,36)} = 0.284$, $p = 0.597$, $\eta^2 = 0.008$.

3.4. Correlations between tasks

To address research question 2 and investigate how the results from the vocabulary, morphosyntax, and narrative microstructure tasks relate to one another, Pearson correlation analyses were conducted between the results from the three tasks in both English and Farsi and between the two languages.

3.5. Correlation between vocabulary and morphosyntax

The results of the correlation analyses between the CLT expressive scores and the SR scores, shown in Table 6 and Figures 2, 3 below, indicate a strong correlation between vocabulary and morphosyntax within each language.

TABLE 2 Descriptive data for English and Farsi LITMUS-SR tasks.

	Mean	Std. deviation	Min−max
SR English accuracy score	20.24	7.27	0–29
SR Farsi accuracy score	19.81	7.68	3–29

TABLE 6 Correlations between the CLT tasks and the SR scores in English and Farsi.

	CLT Exp Farsi	CLT Exp English	SR Eng	SR Farsi
CLT Exp Farsi	1	−0.300		0.724**
CLT Exp English		1	0.743**	

**Correlation is significant at the 0.01 level (2-tailed).

TABLE 3 Descriptive statistics for the microstructure complexity measures.

	Ratio of subordination (RS)			Number of words per clause (WPC)		
	Mean	Standard deviation	Range	Mean	Standard deviation	Range
English	2.05	0.43	1.20–3.00	6.02	0.58	4.86–7.21
Farsi	2.19	0.61	1–4	4.41	0.71	3.86–5.02

TABLE 4 Descriptive statistics for the microstructure fluency measures.

	Syllables per minute (SPM)			Number of filled pauses (NFP)		
	Mean	Standard deviation	Range	Mean	Standard deviation	Range
English	121.70	27.38	69–172	1.66	2.58	0–12
Farsi	118.61	37.70	32.82–201	3.77	3.57	0–12

TABLE 5 Descriptive statistics for the microstructure accuracy measures.

	Percentage of error free clauses (EFC)			Total verbs correct (TVC)		
	Mean	Standard deviation	Range	Mean	Standard deviation	Range
English	89.23	14.04	35.70–100	96.65	4.75	81.81–100
Farsi	79.56	18.21	28.57–100	91.00	12.71	42.85–100

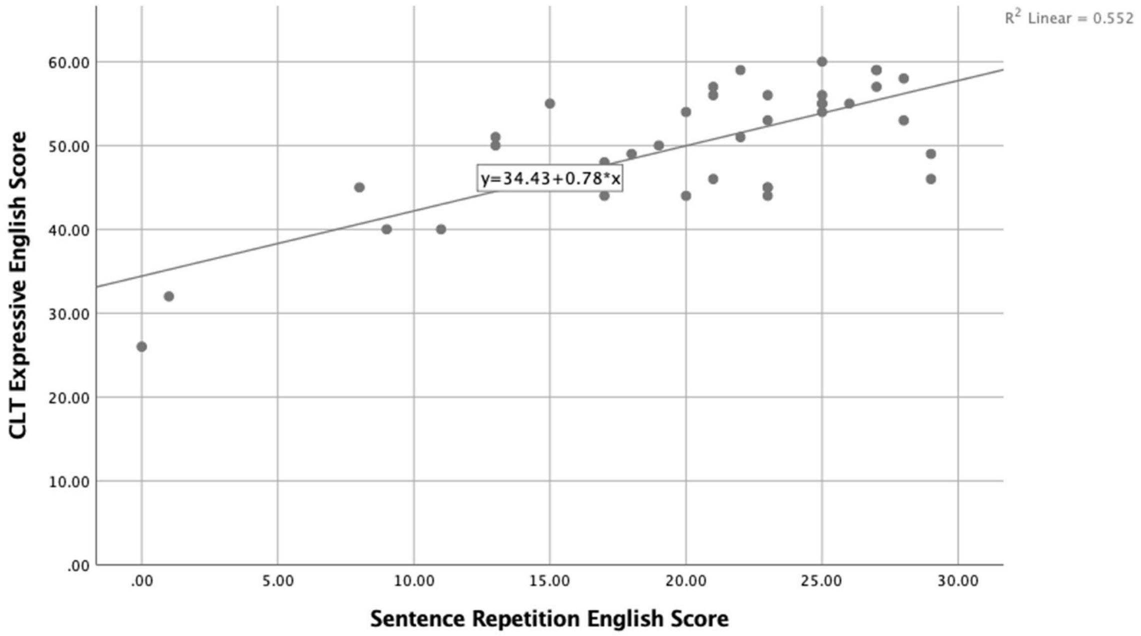


FIGURE 2
The relationship between the CLT Expressive English scores and SR English scores.

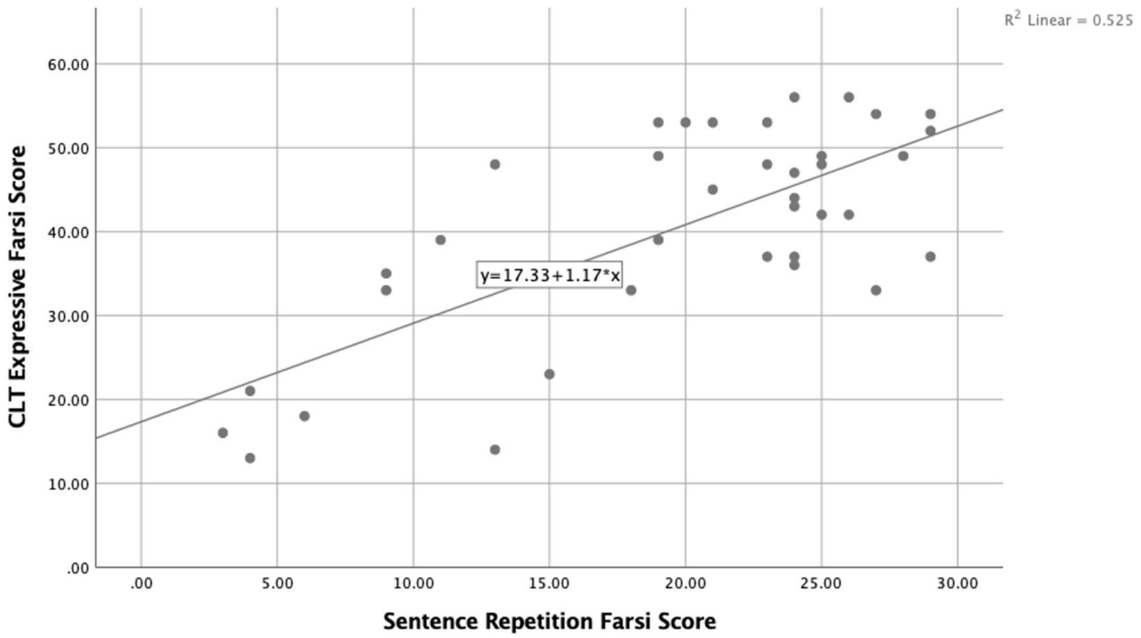


FIGURE 3
The relationship between the CLT Expressive Farsi scores and the Farsi SR score.

3.6. Correlations between vocabulary and narrative microstructure

To identify associations between expressive vocabulary and narrative microstructure scores within each language, two separate Pearson Correlations were done between the CLT Expressive English scores and the English microstructure scores and between the CLT Expressive Farsi scores and the Farsi microstructure scores. The results are shown in Tables 7, 8, respectively.

In both languages, the CLT Expressive English scores are moderately correlated with the fluency score SPM and moderately to highly correlated with the accuracy scores of TVC and EFC, but they are not associated with the complexity scores RS and WPC or with the fluency score NFP.

3.7. Correlation between SR and narrative microstructure

To identify how the scores on the SR tasks relate to those on the narrative microstructure tasks in each language respectively, two separate correlational analyses were conducted, one for each language. The results of these analysis can be seen in Tables 9, 10 below.

The analyses demonstrated that the English SR task correlated with all three aspects of the microstructure—complexity (RS), accuracy (EFC and TVC) and fluency (SPM)—with the scores measuring morphosyntax (EFC and TVC) having the strongest correlations. The Farsi SR scores had a highly significant correlation with the fluency measure of SPM score and a moderately high significant correlation with the accuracy measures of EFC and TVC, but no significant correlations were found between SR and the complexity scores.

TABLE 7 Correlation analysis between CLT English and the microstructure scores in English.

		Complexity		Fluency		Accuracy	
	CLT English Exp	RS	WPC	SPM	NFP	EFC	TVC
CLT English Exp	1	0.219	−0.028	0.576**	−0.193	0.740**	0.537**
RS		1	−0.168	0.376*	−0.080	0.168	0.187
WPC			1	−0.165	0.024	0.170	0.016
SPM				1	−0.162	0.433*	0.267
NFP					1	−0.177	0.163
EFC						1	0.398*
TVC							1

*Correlation is significant at the 0.05 level (2-tailed).
**Correlation is significant at the 0.01 level (2-tailed).

TABLE 8 Correlation analysis between CLT Farsi and the microstructure scores in Farsi.

		Complexity		Fluency		Accuracy	
		CLT Exp Farsi	RS	WPC	SPM	NFP	EFC
CLT Farsi Exp	1	0.315	0.254	0.608**	−0.246	0.658**	0.683**
RS		1	−0.171	0.255	0.033	0.081	0.171
WPC			1	0.033	−0.138	0.107	0.069
SPM				1	−0.310	0.549**	0.422**
NFP					1	−0.165	−0.330*
EFC						1	0.825**
TVC							1

*Correlation is significant at the 0.05 level (2-tailed).
**Correlation is significant at the 0.01 level (2-tailed).

3.8. Profile analysis

To address the third research question and identify whether any of the children are at risk for a Developmental Language Disorder (DLD), we looked at the participants' individual performance along all tasks in English and Farsi, as shown in Tables 11, 12, respectively. These results allow us to demonstrate the individual differences in our participants. The cells which are not highlighted show scores that are within 1.5 SD of the mean of the group. The scores highlighted in orange represent scores which were 1.5 SD below the group mean, while cells highlighted in red represent scores which were 2SD below the group mean.

The tables indicate that participants 8, 19, and 20 have low proficiency in English, as they were performing −1.5 and/or −2SD below the mean on most of the English tasks but do not have low scores in the Farsi tasks. On the other hand, participants 4, 25, 33, 35, and 36 have low proficiency in Farsi as they were performing −1.5 and/or −2SD below the mean on the majority of the Farsi tasks but their scores on the English tasks are within 1.5 SD. It is evident that no child in this study demonstrated low scores in both their languages. This indicates that none of the children appear to be at risk for a DLD.

4. Discussion

This section discusses how the results relate to the current literature. Each of the three research questions will be presented and discussed in separate sections. The limitations of the study as well as potential future research are discussed at the end of the discussion section.

4.1. Discussion of research question 1: is there a difference between the children's proficiency in Farsi as a HL and English as a ML as measured by vocabulary, morphosyntax, and narrative microstructure?

Language dominance and proficiency levels in bilingual children vary greatly (Montrul, 2008, 2016; Bedore et al., 2012) who are often more dominant in one language vs. another (Carroll, 2017). It is possible to have one HS who has high receptive and expressive language abilities in their HL, while another has very little expressive ability and is only able to understand the HL. Therefore, when discussing the language abilities of bilingual children, one must consider the degree of proficiency in both the HL and ML (Montrul, 2008, 2016; Bedore et al., 2012; Luk and Bialystok, 2013; Silva-Corvalán and Treffers-Daller's, 2015; Treffers-Daller's and Korybski, 2015). One of the most common measures of language proficiency and dominance

TABLE 9 Correlational analysis for the English SR scores and the English microstructure scores.

		Complexity		Fluency		Accuracy	
	SR	RS	WPC	SPM	NFP	EFC	TVC
SR	1	0.517**	−0.008	0.502**	−0.238	0.703**	0.520**
RS		1	−0.168	0.376*	−0.080	0.168	0.187
WPC			1	−0.165	0.024	0.170	0.016
SPM				1	−0.162	0.433**	0.267
NFP					1	−0.177	0.163
EFC						1	0.398*
TVC							1

*Correlation is significant at the 0.05 level (2-tailed).
**Correlation is significant at the 0.01 level (2-tailed).

TABLE 10 Correlational analysis for the Farsi SR scores and the Farsi microstructure scores.

		Complexity		Fluency		Accuracy	
	SR	RS	WPC	SPM	NFP	EFC	TVC
SR	1	0.292	0.287	0.716**	−0.323	0.566**	0.587**
RS		1	−0.171	0.255	0.033	0.081	0.171
WPC			1	0.033	−0.138	0.107	0.069
SPM				1	−0.317	0.549**	0.422**
NFP					1	−0.165	−0.330
EFC						1	0.825*
TVC							1

*Correlation is significant at the 0.05 level (2-tailed).
**Correlation is significant at the 0.01 level (2-tailed).

has been vocabulary tasks (Lambert et al., 1959; Fishman and Cooper, 1969; Cromdal, 1999; Reyes and Azuara, 2008; Treffers-Daller's and Korybski, 2015). Therefore, one of the tasks used to identify language dominance in the present study was the CLT Expressive and Receptive task for English and Farsi. The analyses showed no difference between the children's two languages in receptive vocabulary of verbs but an advantage was observed for English in their expressive vocabulary and receptive vocabulary in nouns. This suggests that children are mostly English dominant in their vocabulary.

Bilingualism is multifaceted, and therefore, language dominance and proficiency cannot simply be concluded based on one simple measure (Bedore et al., 2012; Treffers-Daller's and Korybski, 2015; Caffarra et al., 2016; Meir and Armon-Lotem, 2017; Meir, 2018). Vocabulary skills, although significant in demonstrating proficiency, are often positively and highly correlated with morphosyntax (Chondrogianni and Marinis, 2011; Thordardottir and Brandeker, 2013; Gharibi and Boers, 2016; Montrul, 2016; Hamann and Abed Ibrahim, 2017; Meir and Novogrodsky, 2020), implying a certain degree of overlap between the two. The multifaceted nature of bilingual communication, however, encouraged us to collect more evidence about the linguistic skills of the participants that allow for a comparison of their performance in English and Farsi on the SR and narrative microstructure tasks. The analyses on the SR tasks showed no differences in the children's scores between English and Farsi, indicating balance in their morphosyntactic abilities between the two languages. In contrast, the findings of the microstructure analyses showed that children produced more accurate, complex, and fluent oral narratives in the ML (English) than in the HL (Farsi). Of the three microstructure components, accuracy was the most sensitive dimension, with both accuracy measures being significantly higher in English than in Farsi. The complexity and fluency measures were also effective in highlighting the differences between the children's two languages with WPC demonstrating their ability to use more words per clause in English than Farsi, and NFP indicating they had a larger number of filled pauses in Farsi than in English. Given the ample research evidence supporting the reliability of CAF measures in representing language proficiency (Housen and Kuiken, 2009; Norris and Ortega, 2009), these findings are important as they suggest employing a microstructure analysis would be effective in identifying the differences between the bilingual children's proficiency in their two languages and highlighting the nuanced differences that remain unexplored when using other tasks (e.g., SR) that may be less sensitive than microstructure particularly for children at the end of primary/elementary school.

It should also be noted that dominance is a dynamic construct that changes with time and circumstances. Even though the participants in the current study started out as having Farsi as the dominant language on the basis of their total language use in the parental questionnaire, dominance appeared to shift such that as a group, the participants now seem to be English dominant based on the language richness score in the parental questionnaire as well as expressive vocabulary. The literature looking at language dominance and proficiency with respect to narrative microstructure has shown that the LITMUS—MAIN generally finds in favor of the ML (Hipfner-Boucher et al., 2015; Altman et al., 2016; Bohnacker, 2016; Kapalkova et al., 2016; Roch et al., 2016; Méndez et al., 2018; Hao et al., 2019). The findings of the present study reiterate those found in the literature such that the participants were found to be more proficient and dominant in the ML (English) than the HL (Farsi). The participants produced more complex, fluent and accurate oral narratives in the English than in Farsi. Of the three microstructure components—complexity, fluency and accuracy- the latter was the most telling as both accuracy measures were significantly higher in English than in Farsi.

4.2. Discussion of research question 2: how do the scores on the vocabulary, morphosyntax, and narrative microstructure tasks relate to one another?

A significant amount of research suggests that there is a relationship between vocabulary and grammar (Bates et al., 1988; Caselli et al., 1999; Thal et al., 2000; Devescovi et al., 2005; Hamann and Abed Ibrahim, 2017; Hoff et al., 2018; Kaltsa et al., 2020). However, these studies generally compared vocabulary and grammar in different groups of children. Research looking at the

TABLE 11 English profile analysis.

English raw scores on all tasks													
ID	CLTNP	CLTNC	CLTVP	CLTVC	CLTExp	CLTRec	SR task	RS	WPC	NFP	SyPM	EFC %	TVC %
1	29	31	27	29	56	60	25	1.7	5.53	0	160	100	100
2	26	30	20	28	46	58	21	2.27	6.12	1.29	79	88	100
3	26	30	19	25	45	55	23	1.67	6.87	3.58	109.25	86.67	95
4	27	31	22	29	49	60	18	1.22	7.18	2.35	97.6	100	100
5	29	31	27	30	56	61	21	1.67	5.36	0	127.9	92	100
6	29	31	26	30	55	61	25	1.88	5.53	0	84.34	95	94
7	28	31	22	29	50	60	19	1.81	6.35	7.88	126.09	100	100
8	20	31	12	25	32	56	1	1.67	6.4	0	69	40	81.81
9	25	31	25	29	50	60	13	2	5.86	0	123.29	81.81	95.8
10	23	30	21	29	44	59	23	2.25	5.22	0	133.47	83.33	85.71
11	26	31	23	30	49	61	29	2.09	6.78	0.67	128.27	95.6	100
12	32	31	23	29	55	60	15	1.2	6.16	2.46	98.63	94.44	100
13	26	31	14	27	40	58	11	1.5	4.86	1.39	114.42	80	86.67
14	30	31	24	30	54	61	25	2.5	6.08	12	121.33	92	100
15	29	31	26	30	55	61	25	2.5	5.92	1	172	100	100
16	32	31	27	31	59	62	27	3	5.88	0	148.23	100	100
17	26	31	19	30	45	61	23	2.5	5.95	0	139	85	95.83
18	26	31	18	28	44	59	20	2	7.21	0.71	80.71	92.8	93.75
19	16	25	10	14	26	39	0	1.75	5.14	7	81.00	35.57	100
20	23	28	17	23	40	51	9	1.875	6.67	1.75	69.9	86.67	94.12
21	24	28	21	30	45	58	8	1.8	6.67	4	132	90	86.9
22	29	31	25	28	54	59	20	2.00	7.14	0	151.91	100	100
23	32	31	21	28	53	59	23	2.28	6.25	1.09	120	87.5	93.3
24	29	31	26	28	55	59	26	2.38	6.16	2	132	78.94	100
25	31	31	28	31	59	62	27	2	5.82	0	106.78	100	100
26	31	31	25	29	56	60	23	1.9	6.71	0.67	115.33	90.47	96
27	31	31	27	31	58	62	28	2.67	5.92	0.76	126.07	83.3	100
28	25	31	26	28	51	59	22	2	5.38	2.3	114.23	100	100
29	31	31	26	29	57	60	21	2.09	6.17	0	169.28	86.96	96.15
30	32	31	25	30	57	61	27	2.44	5.59	1.05	163.16	100	100
31	25	31	21	27	46	58	29	2.75	6.09	0	135.43	100	100
32	25	29	23	28	48	57	17	1.92	5.87	0	143.82	86.95	92.59
33	32	31	28	28	60	59	25	2	5.22	2.45	128.57	100	100
34	30	31	23	31	53	62	28	2.8	5.64	0	143.25	92.85	96.77
35	25	30	19	30	44	60	17	2.33	5.71	0.63	85	90.47	95.65
36	32	31	27	30	59	61	22	1.45	6.13	0	138.75	93.75	100
37	30	31	21	28	51	59	13	2.3	5.65	4.29	133.71	91.3	96
Mean	27.62	30.51	22.54	28.35	50.16	58.86	20.24	2.06	6.02	1.66	121.70	89.23	96.65
−1.5 SD	22.1 (22)	28.68 (29)	15.96 (16)	24.33 (24)	38.75 (39)	52.87 (53)	9.34 (9)	1.37	5.15	5.53*	80.63	61.55	89.52
−2 SD	20.26 (20)	26.85 (27)	13.76 (14)	22.39 (22)	34.94 (35)	50.88 (51)	5.7 (6)	1.15	4.86	6.82*	66.94	68.17	87.15
SD	3.68	1.22	4.39	2.98	7.61	3.99	7.27	0.43	0.58	2.58	27.38	14.04	4.75

*For NFP the SD was added as the more filled pauses per minute the more disfluent the child. Numbers in brackets are the decimal scores rounded to the nearest whole numbers, in order to make cut-off values clearer.

The cells which are not highlighted show scores that are within 1.5 SD of the mean of the group. The scores highlighted in orange represent scores which were 1.5 SD below the group mean, while cells highlighted in red represent scores which were 2SD below the group mean.

TABLE 12 Farsi profile analysis.

ID	Farsi												
	CLTNP	CLTNC	CLTVP	CLTVC	CLTExp	CLTRec	SR task	RS	WPC	NFP	SyPM	EFC	TVC
1	26	32	17	28	43	60	24	1.5	4.71	2.86	154.3	100	100
2	20	28	19	30	39	58	19	2.25	5.15	6.04	92.62	67	82.14
3	19	32	14	30	33	62	27	1.78	5.25	10.11	99.1	62.50	87.5
4	13	27	8	27	21	54	4	2.13	3.06	5.55	94.44	52.90	61.50
5	26	31	21	27	47	58	24	2.36	4.07	0	156.36	100	100
6	28	31	20	25	48	56	13	2.2	3.95	0.44	43.43	77.30	91.30
7	17	32	18	27	35	59	9	4	4.08	11.68	82.83	83.33	88.46
8	27	32	21	26	48	58	23	1.46	4.95	0.6	105.45	94.73	100
9	29	31	20	30	49	61	25	2.09	5.82	1.84	137.14	69.56	93.75
10	20	31	16	31	36	62	24	2.22	4.35	0	126.23	95.00	100
11	19	32	18	28	37	60	29	2.09	4.82	0	129.33	73.90	86.20
12	19	27	14	25	33	52	9	1.67	3.5	6.96	63.21	75.00	95
13	26	31	19	30	45	61	21	2	3.7	1.5	201	95.00	100
14	27	32	22	30	49	62	19	1.9	5.52	4.39	147.07	90.47	100
15	24	32	25	30	49	62	28	1.8	4.44	0	146	94.44	100
16	20	32	17	31	37	63	23	2.54	4.86	2.9	105.48	71.43	84.38
17	30	32	26	29	56	61	26	2.11	4.32	1	147	100.00	100
18	30	31	18	28	48	59	25	3.57	4.56	4.15	85.38	73.00	96.29
19	29	31	24	29	53	60	19	2.63	4.14	5.52	108.95	61.90	92
20	29	32	27	29	56	61	24	2.1	5.48	3.83	128.94	76.19	85.71
21	28	32	25	31	53	63	21	2.46	3.06	4.28	146.57	81.25	93.75
22	29	32	24	29	53	61	20	2.16	5.07	0.77	146.34	100	100
23	9	31	14	28	23	59	15	1.5	4.08	4.28	72.86	83.30	91.67
24	27	32	25	30	52	62	29	3.43	4.54	5	178	95.83	100
25	7	26	7	29	14	55	13	2.42	3.41	0	120.33	35.29	58.82
26	18	30	21	29	39	59	11	1.36	4.8	6.67	94.17	100.00	100
27	22	31	20	28	42	59	26	2.84	4.46	2.38	147.62	75.60	89.59
28	20	32	17	28	37	60	24	2.11	3.52	12	112	94.70	94.40
29	29	32	25	30	54	62	27	2.38	4.58	0	158	94.70	95.23
30	28	31	26	29	54	60	29	2.8	4.68	1.5	163.5	82.14	97.00
31	23	31	21	26	44	57	24	2.8	3.64	0	143.25	82.14	100
32	29	31	24	30	53	61	23	1.83	4.81	7.5	125.25	81.81	100
33	10	20	6	25	16	45	3	1	4.86	9.13	63.91	28.57	42.85
34	22	32	20	30	42	62	25	2.06	5.19	1	130.5	100.00	100
35	9	25	9	28	18	53	6	2.15	4.93	3.47	71.37	53.50	79.30
36	6	24	7	28	13	52	4	1.22	4.36	1.54	32.82	55.56	90.90
37	17	29	16	28	33	57	18	2.18	3.2	10.59	137.65	85.71	89.47
Mean	21.92	30.27	18.67	28.54	40.59	58.81	19.81	2.19	4.41	3.77	118.61	79.56	91.00
−1.5 SD	11.28 (11)	26.15 (26)	10.12 (10)	26.02 (26)	21.92 (22)	53.14 (53)	8.29 (8)	1.28	3.35	9.13	62.06	52.25	71.94
−2 SD	7.74 (8)	24.77 (25)	7.17 (7)	25.18 (25)	15.70 (16)	51.25 (51)	4.45 (4)	0.97	3.00	10.71	43.21	43.14	65.58
SD	7.09	2.75	5.75	1.68	12.45	3.78	7.68	0.61	0.71	3.57	37.70	18.21	12.71

*For NFP the SD was added as the more filled pauses per minute the more disfluent the child. Numbers in brackets are the decimal scores rounded to the nearest whole numbers, in order to make cut-off values clearer.

The cells which are not highlighted show scores that are within 1.5 SD of the mean of the group. The scores highlighted in orange represent scores which were 1.5 SD below the group mean, while cells highlighted in red represent scores which were 2SD below the group mean.

relationship between vocabulary and grammar in the same group of children is far less common with only a few studies noted to date (Jia et al., 2002; Bohman et al., 2010; Chondrogianni and Marinis, 2011; Thordardottir and Brandeker, 2013; Meir and Novogrodsky, 2020). These studies all showed that vocabulary and morphosyntax were correlated within each language, however, results differed on correlations found cross linguistically. The findings of the current study are in line with Thordardottir and Brandeker (2013) and Meir and Novogrodsky (2020), suggesting that vocabulary and grammar are positively and significantly correlated within each language but not across the two languages. It should be noted that, although there are significant correlations between vocabulary and morphosyntax there is also individual variability in the children's performance as shown in the participant's profiles (see Tables 11, 12). The lack of cross linguistic correlations between the English vocabulary scores and the Farsi SR scores or vice versa suggests that vocabulary and morphosyntax develop in each language separately. Therefore, if we want children to develop both languages adequately, it is important to provide sufficient input and exposure to both languages because vocabulary and morphosyntactic skills will not transfer from one language to the other.

The literature has shown that children are required to reach a threshold level of vocabulary to be able to produce appropriate story narratives (Pearson, 2002; Uccelli and Paez, 2007; Karlsen et al., 2016; Méndez et al., 2018; Hao et al., 2019). The current study reiterates the previous findings that vocabulary correlates with narrative microstructure within each language. However, how vocabulary and microstructure are related in terms of accuracy, fluency, and complexity is unclear from previous research because there are discrepancies between previous studies. For example, Kambanaros et al. (2014), found no correlations between vocabulary and narrative complexity, while Méndez et al. (2018) showed that lexical abilities are highly associated to complexity measures in a bilingual child's languages during narrative retelling tasks. Méndez et al. (2018) also indicated that vocabulary is a strong predictor of narrative skills in the ML but not in the HL. The differences between the results of these studies can be explained in the light of the different narrative tasks and complexity measures they employed. While Méndez et al. (2018) used the *Frog where are you?* narrative retelling task and measured complexity in terms of subordination and length, Kambanaros et al. (2014) used the Bus Story task and other measures of complexity.

Overall, our results suggest that within each language children produce more accurate and fluent narratives when they have higher vocabulary skills. In terms of complexity, although the present study measured complexity using the same method as Méndez et al. (2018), we did not find a correlation between vocabulary and complexity, which is in line with Kambanaros et al. (2014). Methodological differences between the current study and Méndez et al. (2018) can account for the discrepancies found in the results. The methodological differences potentially affecting the results can be summarized in relation to sample size (14 participants in Méndez et al., 2018 vs. 37 in the current study), the narrative task (the *Frog Story* in Méndez et al., 2018 vs. the LITMUS-MAIN in the current study), the average age of the participants (younger in Méndez et al., 2018 than in the current study), and the different language pairs in terms of language distance (Spanish-English in Méndez et al., 2018 vs. Farsi-English in the current study).

In addition to correlations between vocabulary and microstructure, we found that the results of the SR tasks correlated with the microstructure within each language. These findings are in line with the studies by Hesketh (2004), Ellis (2005), Erlam (2006), and Bowles (2011), each of which looked at how language performance on an oral imitation task compared to performance on an oral narrative task. All four studies found that performance on the sentence imitation tasks was highly correlated to the oral narrative tasks. In terms of the measures of accuracy, fluency, and complexity, our participants' scores on the Farsi SR task were significantly correlated to the accuracy and fluency scores on the Farsi LITMUS-MAIN. The correlations were even stronger for the English tasks in that the scores on the English SR task related to the scores on all three aspects of the microstructure (complexity, fluency, and accuracy), although the correlations with the accuracy scores were the strongest. This is similar to Hesketh (2004) who found significant correlations between accuracy scores on the narrative task and scores on the sentence elicitation task.

4.3. Discussion of research question 3: based on the results of each task in both languages, are any of the children at risk for a Developmental Language Disorder?

In order for a bilingual child to be diagnosed as having a DLD, they need to have a score of <2SD below the mean on at least two language measures in both their languages (Kohnert, 2010; Leonard, 2014). Since the children in the present study were tested in both languages across various language domains, we were able to determine if any of the children were at risk for DLD. The prediction was that no child in the sample would be at risk for DLD, as they were reported to be typically developing (TD) and had no previous clinical diagnosis. This hypothesis was confirmed. The analysis showed that no child scored −1.5 or −2 SD below the mean on more than two tasks in both languages. However, interesting patterns emerged indicating that some participants had a greater proficiency in one language vs. their other language due to dominance effects. For example, participants 8, 19, 20, and 21 demonstrated a low proficiency in English, as they had low scores on multiple English tasks. Looking at their demographic data, these children were exposed to English before the age of 4 and had two Farsi speaking Parents. However, participant 19 (age 6;20) arrived in Canada when he was 4;5, although parental reports indicated that he was exposed to English from birth via television and movies and that he attended English classes in Iran before the age of 4. It is possible that his low language scores in English could be the result of limited amount of time residing in an English-speaking society and limited amount of schooling in English. On the other hand, participants 4, 25, 33, 35, and 36 had low proficiency in Farsi, as they were performing −1.5 and/or −2SD below the mean on the majority of the Farsi tasks, while their scores on the English tasks were within 1 SD. These children were all exposed to Farsi from birth and to English before the age of 4. An exception is participant 35 who had some low scores in Farsi. This child had one parent who was not a Farsi speaker suggesting Farsi was not a common home language between the parents. However, the child spoke Farsi fluently and was able to produce a narrative with appropriate complexity, fluency and accuracy. She was exposed to both English and Farsi from birth and used both languages regularly with friends and family. The results suggested that no child was at risk for DLD, as none of the participants had significant difficulties in both languages.

The findings from this study demonstrate the importance of testing bilingual children in both their languages across various language domains. The results indicate that it is possible to see low language scores in one language but not in the other language. To have a valid and reliable assessment of bilingual children's proficiency, it is necessary to have a sufficiently broad and rich sample of their language performance across their two languages. Using single measures of proficiency in one of their language systems may fail to provide an insight into their linguistic abilities. If appropriate testing is not done in both languages, misdiagnosis of a language impairment may occur. From a clinically practical perspective, it is often not possible to test children in both languages, as clinicians may not have access to the appropriate testing material and/or to speak the HL of the child. To make up for such limitations, therefore, conducting parental questionnaires or interviews to obtain information on the child's language history is extremely important. The information obtained by parents can, to some extent, help clinicians identify the child's language dominance and proficiency and potentially aid in the diagnostic process.

5. Conclusions

5.1. Limitations and future directions

This study provides new insights into the importance of measuring language dominance and proficiency via different language tasks that measure different language domains and demonstrates that the different language domains interact with one another in bilingual children. Future research can build on the current study by exploring other language groups as well as looking at how internal and external factors come into play. Similar studies using the same methodology can be used comparing TD bilingual children to bilingual children with DLD. While there is a breath of studies which compares the language performance of TD bilinguals to either monolinguals or bilinguals with DLD, very few of these studies look at the interactions between all three language domains in addition to the influences of internal and external factors on those domains within the same group of children. Replication studies using different language combinations will allow for greater generalization of the results and would help in terms of diagnostic purposes. However, any future studies need to take into account the limitations set forth by the current study. The first limitation was the small sample size. This study focused on TD children and was attempting to identify information on the language skills of TD Farsi-English bilingual children which could potentially be used for diagnostic purposes. Consequently, having a much larger sample size would have been more favorable. In addition, a larger sample size would be more representative of the population. A second limitation of the study was that in many ways the sample was rather homogenous in that the participants were all from families from middle to upper socio-economic status (SES) in Toronto, both parents of all children attended at least some kind of higher education programs (college or university), and the majority of the children had been exposed to Farsi from birth. A sample with children from both high and low SES as well as children with different ages of onset would be more representative of the variability of the population of Farsi speaking children growing up in Canada.

5.2. Conclusion

Two main conclusions can be drawn from this study. The first is that while the results indicated that overall the participants in this study were English dominant, an in-depth analysis of performance via profile effects demonstrated that some children had greater proficiency in Farsi while others had greater proficiency in English. These findings show that dominance and proficiency are two distinct constructs; it is important to measure language dominance and proficiency via multiple language domains because differences between these domains may emerge. This points to the significance of identifying individual profile effects across the domains. Bilingual children can have low proficiency in both languages but still have one language as more dominant (i.e., be their stronger or preferred language). For HS we know that dominance shifts through the lifespan often with the HL becoming less dominant and the SL becoming more dominant but there is a lot of individual variability between children. Therefore, both dominance and proficiency need to be measured and considered in research and practice across multiple language domains. Secondly, vocabulary is highly correlated with morphosyntax and is also related to the complexity, accuracy and fluency of oral narratives within languages.

Using multiple language measures and profile effects in this study and conducting rigorous analyses of the data have provided a more in-depth understanding of bilingual linguistic abilities in this sample. The present study is novel in that no known study to date had identified language dominance and proficiency in such a broad and in-depth manner. The results also demonstrate the importance of testing bilingual children in both their languages across a number of language domains in order to obtain a more accurate picture of their language skills in the HL and ML. Theoretically, such information enriches the literature by providing further information on the language skills of these types of dual language learners. Detailed and accurate information on the language abilities of a child in both their HL and ML helps reduce the potential for misdiagnoses to occur and ultimately leads to better treatment outcomes.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

This study has been reviewed by The School of Psychology and Clinical Language Sciences Research Ethics Committee, University of Reading and has been given favorable ethical opinion for conduct (2018-102-TM). Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Author contributions

MK collected and analyzed the data. All authors contributed to the conceptualization, design of this study, and write up of the manuscript. All authors contributed to the article and approved the submitted version.

Funding

During the write up of the manuscript, TM received funding from the project: The Multilingual Mind, which is part of the European Union's Horizon 2020 Research and Innovation Program under the Marie Skłodowska Curie Grant Agreement No. 765556.

Acknowledgments

We would like to thank the participants of the study for their willingness to take part.

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.

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

Altman, C., Armon-Lotem, S., Fichman, S., and Walters, J. (2016). Macrostructure, microstructure, and mental state terms in the narratives of English–Hebrew bilingual preschool children with and without specific language impairment. *Appl. Psycholinguist.* 37, 165–193. doi: 10.1017/S0142716415000466

Bates, E., Bretherton, I., and Snyder, L. (1988). *From First Words to Grammar: Individual Differences and Dissociable Mechanisms*. New York, NY: Cambridge University Press.

Bates, E., and Goodman, J. (1999). "The emergence of grammar from the lexicon," in *The Emergence of Language*, ed B. MacWhinney (Mahwah, NJ: Lawrence Erlbaum), 29–79.

Bedore, L. M., Peña, E. D., Summers, C. L., Boerger, K. M., Resendiz, M. D., Greene, K., et al. (2012). The measure matters: language dominance profiles across measures in Spanish-English bilingual children. *Bilingualism* 15, 616–629. doi: 10.1017/S1366728912000090

Benmamoun, E., Montrul, S., and Polinsky, M. (2013). Heritage languages and their speakers: opportunities and challenges for linguistics. *Theoret. Linguist.* 39, 129–181. doi: 10.1515/tl-2013-0009

Bialystok, E., Craik, F. I. M., and Luk, G. (2008). Lexical access in bilinguals: effects of vocabulary size and executive control. *J. Neurolinguist.* 21, 522–538. doi: 10.1016/j.jneuroling.2007.07.001

Bohman, T. M., Bedore, L. M., Peña, E. D., Mendez-Perez, A., and Gillam, R. B. (2010). What you hear and what you say: language performance in Spanish-English bilinguals. *Int. J. Bilingual Educ. Bilingualism* 13, 325–344. doi: 10.1080/13670050903342019

Bohnacker, U. (2016). Tell me a story in English or Swedish: narrative production and comprehension in bilingual preschoolers and first graders. *Appl. Psycholinguist.* 37, 19–48. doi: 10.1017/S0142716415000405

Bowles, M. A. (2011). Measuring implicit and explicit linguistic knowledge: what can heritage language learners contribute? *Stud. Sec. Lang. Acquisit.* 33, 247–271. doi: 10.1017/S0272263110000756

Cabo, D. P. Y., and Rothman, J. (2012). The (il)logical problem of heritage speaker bilingualism and incomplete acquisition. *Appl. Linguist.* 33, 450–455. doi: 10.1093/applin/ams037

Caffarra, S., Zimnukhova, S., and Mancini, S. (2016). What usage can do: The effect of language dominance on simultaneous bilinguals' morphosyntactic processing. *Linguist. Vanguard* 1, 20. doi: 10.1515/lingvan-2016-0020

Carroll, S. E. (2017). Exposure and input in bilingual development. *Bilingualism* 20, 3–16. doi: 10.1017/S1366728915000863

Caselli, M. C., Casadio, P., and Bates, E. (1999). A comparison of the transition from first words to grammar in English and Italian. *J. Child Lang.* 26, 69–112. doi: 10.1017/S0305000998003687

Chondrogianni, V., and Marinis, T. (2011). Differential effects of internal and external factors on the development of vocabulary, tense morphology and morpho-syntax in successive bilingual children. *Linguist. Approach. Bilingualism* 1, 318–345. doi: 10.1075/lab.1.3.05cho

Cromdal, J. (1999). Childhood bilingualism and metalinguistic skills: analysis and control in young Swedish–English bilinguals. *Appl. Psycholinguist.* 20, 1–20. doi: 10.1017/S0142716499001010

Devescovi, A., Caselli, M. C., Marchione, D., Pasqualetti, P., Reilly, J., and Bates, E. (2005). A crosslinguistic study of the relationship between grammar and lexical development. *J. Child Lang.* 32, 759. doi: 10.1017/S0305000905007105

Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: a psychometric study. *Stud. Sec. Lang. Acquisit.* 27, 141–172. doi: 10.1017/S0272263105050096

Erlam, R. (2006). Elicited imitation as a measure of L2 implicit knowledge: an empirical validation study. *Appl. Linguist.* 27, 464–491. doi: 10.1093/applin/aml001

Fishman, J. A., and Cooper, R. L. (1969). Alternative measures of bilingualism. *J. Verb. Learn. Verb. Behav.* 8, 276–282. doi: 10.1016/S0022-5371(69)80076-9

Foster, P., Tonkyn, A., and Wigglesworth, G. (2000). Measuring spoken language: a unit for all reasons. *Appl. Linguist.* 21, 354–375. doi: 10.1093/applin/21.3.354

Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balèiûnienë, I., et al. (2012). *MAIN—Multilingual Assessment Instrument for Narratives. ZAS Papers in Linguistics 56*. Berlin: ZAS.

Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Bohnacker, U., and Walters, J. (2019). *MAIN: Multilingual Assessment Instrument for Narratives - Revised version. ZAS Papers in Linguistics 63*. Berlin: ZAS.

Gharibi, K., and Boers, F. (2016). Influential factors in incomplete acquisition and attrition of young heritage speakers' vocabulary knowledge. *Lang. Acquisit.* 24, 52–69. doi: 10.1080/10489223.2016.1187613

Haman, E., Lunieska, M., and Pomiechowska, B. (2015). "Designing cross linguistic lexical tasks (CLT's) for bilingual preschool children," in *Assessing Multilingual Children: Disentangling Bilingualism from Language Impairment*, eds S. Armon-Lotem, J. De Jong, and N. Meir (Clevedon: Multilingual Matters), 196–240. doi: 10.21832/9781783093137-010

Hamann, C., and Abed Ibrahim, L. (2017). Methods for identifying specific language impairment in bilingual populations in Germany. *Front. Commun.* 2, 16. doi: 10.3389/fcomm.2017.00016

Hao, Y., Bedore, L. M., Sheng, L., and Peña, E. D. (2019). Narrative skills in two languages of Mandarin–English bilingual children. *Int. J. Speech Lang. Pathol.* 21, 325–335. doi: 10.1080/17549507.2018.1444092

Hesketh, A. (2004). Grammatical performance of children with language disorder on structured elicitation and narrative tasks. *Clin. Linguist. Phonet*. 18, 161–182. doi: 10.1080/02699200310001659061

Hipfner-Boucher, K., Milburn, T., Weitzman, E., Greenberg, J., Pelletier, J., and Girolametto, L. (2015). Narrative abilities in subgroups of English language learners and monolingual peers. *Int. J. Bilingualism* 19, 677–692. doi: 10.1177/1367006914534330

Hoff, E., and Core, C. (2013). Input and language development in bilingually developing children. *Semin. Speech Lang.* 17, 215–226. doi: 10.1055/s-0033-1353448

Hoff, E., Quinn, J. M., and Giguere, D. (2018). What explains the correlation between growth in vocabulary and grammar? New evidence from latent change score analyses of simultaneous bilingual development. *Dev. Sci.* 21, 536. doi: 10.1111/desc.12536

Housen, A., and Kuiken, F. (2009). Complexity, accuracy, and fluency in second language acquisition. *Appl. Linguist*. 30, 461–473. doi: 10.1093/applin/amp048

Jia, G., Aaronson, D., and Wu, Y. (2002). Long-term language attainment of bilingual immigrants: predictive variables and language group differences. *Appl. Psycholinguist*. 23, 599. doi: 10.1017/S0142716402004058

Kaltsa, M., Prentza, A., and Tsimpli, I. M. (2020). Input and literacy effects in simultaneous and sequential bilinguals: the performance of Albanian–Greek-speaking children in sentence repetition. *Int. J. Bilingualism* 24, 159–183. doi: 10.1177/1367006918819867

Kambanaros, M., Grohmann, K., Theodorou, E., and Michaelides, M. (2014). "Can vocabulary size predict narrative abilities in children with SLI?," in *Major Trends in Theoretical and Applied Linguistics 2*, eds N. Lavidas, T. Alexiou and A. Sougari (London: Versita Ltd, Great Britain: De Gruyter Open Poland), 423–436.

Kapalkova, S., PoliŠEnska, K., Markova, L., and Fenton, J. (2016). Narrative abilities in early successive bilingual Slovak–English children: a cross-language comparison. *Appl. Psycholinguist*. 37, 145–164. doi: 10.1017/S0142716415000454

Karlsen, J., Geva, E., and Lyster, S.-A. (2016). Cognitive, linguistic, and contextual factors in Norwegian second language learner's narrative production. *Appl. Psycholinguist*. 37, 1117–1145. doi: 10.1017/S014271641500051X

Kohnert, K. (2010). Bilingual children with primary language impairment: issues, evidence and implications for clinical actions. *J. Commun. Disord.* 43, 456–473. doi: 10.1016/j.jcomdis.2010.02.002

Kohnert, K. J., and Bates, E. (2002). Balancing bilinguals II: lexical comprehension and cognitive processing in children learning Spanish and English. *J. Speech Lang. Hear. Res.* 45, 27. doi: 10.1044/1092-4388(2002/027)

Kohnert, K. J., Hernandez, A. E., and Bates, E. (1998). Bilingual performance on the boston naming test: preliminary norms in Spanish and English. *Brain Lang.* 65, 422–440. doi: 10.1006/brln.1998.2001

Komeili, M., Marinis, T., Tavakoli, P., and Kazemi, Y. (2020). Sentence repetition in Farsi-English bilingual children. *J. Eur. Sec. Lang. Assoc.* 4, 1–12. doi: 10.22599/jesla.55

Lambert, W. E., Havelka, J., and Gardner, R. C. (1959). Linguistic manifestations of bilingualism. *Am. J. Psychol.* 72, 77–82. doi: 10.2307/1420213

Leonard, L. B. (2014). *Children With Specific Language Impairment, 2nd Edn*. Cambridge, MA: A Bradford Book.

Luk, G., and Bialystok, E. (2013). Bilingualism is not a categorical variable: interaction between language proficiency and usage. *J. Cogn. Psychol.* 25, 605–621. doi: 10.1080/20445911.2013.795574

Marinis, T., and Armon-Lotem, S. (2015). Sentence repetition. *Assess. Multilingual Child.* 7, 95–124. doi: 10.21832/9781783093137-007

Meir, N. (2018). Morpho-syntactic abilities of unbalanced bilingual children: a closer look at the weaker language. *Front. Psychol.* 9, 1318. doi: 10.3389/fpsyg.2018.01318

Meir, N., and Armon-Lotem, S. (2017). Independent and combined effects of socioeconomic status (SES) and bilingualism on children's vocabulary and verbal short-term memory. *Front. Psychol.* 8, 1442. doi: 10.3389/fpsyg.2017.01442

Meir, N., and Novogrodsky, R. (2020). Syntactic abilities and verbal memory in monolingual and bilingual children with High Functioning Autism (HFA). *First Lang.* 40, 341–366. doi: 10.1177/0142723719849981

Méndez, L. I., Perry, J., Holt, Y., Bian, H., and Fafulas, S. (2018). Same or different: narrative retells in bilingual Latino kindergarten children. *Bilingual Res. J.* 41, 150–166. doi: 10.1080/15235882.2018.1456984

Montrul, S. (2016). *The Acquisition of Heritage Languages*. Cambridge: Cambridge University Press.

Montrul, S. A. (2008). *Incomplete Acquisition in bilingualism: Re-examining the Age Factor (Vol. 39)*. Amsterdam: John Benjamins Publishing. doi: 10.1075/sibil.39

Norris, J. M., and Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: the case of complexity. *Appl. Linguist*. 30, 555–578. doi: 10.1093/applin/amp044

Ortega, L. (2020). The study of heritage language development from a bilingualism and social justice perspective. *Lang. Learn.* 70, 15–53. doi: 10.1111/lang.12347

Pearson, B. Z. (2002). "Narrative competence among monolingual and bilingual school children in Miami," in *Language and Literacy in Bilingual Children*, eds D. K. Oller and R. E. Eilers (Clevedon: Multilingual Matters), 135–174. doi: 10.21832/9781853595721-008

Reyes, I., and Azuara, P. (2008). Emergent biliteracy in young Mexican immigrant children. *Read. Res. Quart*. 43, 374–398. doi: 10.1598/RRQ.43.4.4

Roch, M., Florit, E., and Levorato, C. (2016). Narrative competence of Italian–English bilingual children between 5 and 7 years. *Appl. Psycholinguist*. 37, 49–67. doi: 10.1017/S0142716415000417

Rothman, J., Tsimpli, I. M., and Pascual y Cabo, D. (2016). Formal linguistic approaches to heritage language acquisition. *Adv. Spanish Heritage Lang.* 2, 13–26. doi: 10.1075/sibil.49.02rot

Silva-Corvalán, C., and Treffers-Daller's, J. (2015). "Digging into dominance: a closer look at language dominance in bilinguals," in *Language Dominance in Bilinguals: Issues of Measurement and Operationalization*, eds C. Silva-Corvalán and J. Treffers-Daller's (Cambridge: Cambridge University Press), 1–14. doi: 10.1017/CBO9781107375345.001

Talabi, A. (2018). "Crosslinguistic lexical tasks: persian version (CLT-FA). A part of LITMUS COST IS804 Battery. Unpublished material. Based on: Haman, E., Łuniewska, M., and Pomiechowska, B. (2015). Designing Cross-linguistic lexical tasks (CLTs) for bilingual preschool children," in *Methods for Assessing Multilingual Children: Disentangling Bilingualism From Language Impairment*, eds S. Armon-Lotem, J. D. Jong, and N. Meir (Bristol: Multilingual Matters), 196–240.

Tavakoli, P. (2018). L2 development in an intensive study abroad EAP context. *System* 72, 62–74. doi: 10.1016/j.system.2017.10.009

Tavakoli, P., and Wright, C. (2020). *Second Language Speech Fluency: From Research to Practice*. Cambridge: Cambridge University Press.

Thal, D. J., Jackson-Maldonado, D., and Acosta, D. (2000). Validity of a parent-report measure of vocabulary and grammar for Spanish-speaking toddlers. *J. Speech Lang. Hear. Res.* 43, 1087–1100. doi: 10.1044/jslhr.4305.1087

Thordardottir, E., and Brandeker, M. (2013). The effect of bilingual exposure versus language impairment on nonword repetition and sentence imitation scores. *J. Commun. Disord.* 46, 1–16. doi: 10.1016/j.jcomdis.2012.08.002

Tomasello, M. (2000). Do young children have adult syntactic competence? *Cognition* 74, 209–253. doi: 10.1016/S0010-0277(99)00069-4

Treffers-Daller's, J., and Korybski, T. (2015). "Using lexical diversity measures to operationalise language dominance in bilinguals," in *Language Dominance in Bilinguals: Issues of Measurement and Operationalization*, eds C. SilvaCorvalan and J. Treffers Daller's (Cambridge: Cambridge University Press), 106123. Available online at: http://centaur.reading.ac.uk/39019/ (accessed October 8, 2020).

Tuller, L. (2015). "Clinical use of parental questionnaires in multilingual contexts," in *Assessing Multilingual Children: Disentangling Bilingualism From Language Impairment*, eds S. Armon-Lotem., and J. de Jong (Bristol: Multilingual Matters), 301–330. doi: 10.21832/9781783093137-013

Uccelli, P., and Paez, M. M. (2007). Narrative and vocabulary development of bilingual children from kindergarten to first grade: developmental changes and associations among English and Spanish skills. *Lang. Speech Hear. Serv. Schools* 38, 225–236. doi: 10.1044/0161-1461(2007/024)

Unsworth, S. (2015). "Amount of exposure as a proxy for dominance in bilingual language acquisition," in *Language Dominance in Bilinguals: Issues of Measurement and Operationalization*, eds C. Silva-Corvalán and J. Treffers-Daller's (Cambridge: Cambridge University Press), 156–173. doi: 10.1017/CBO9781107375345.008

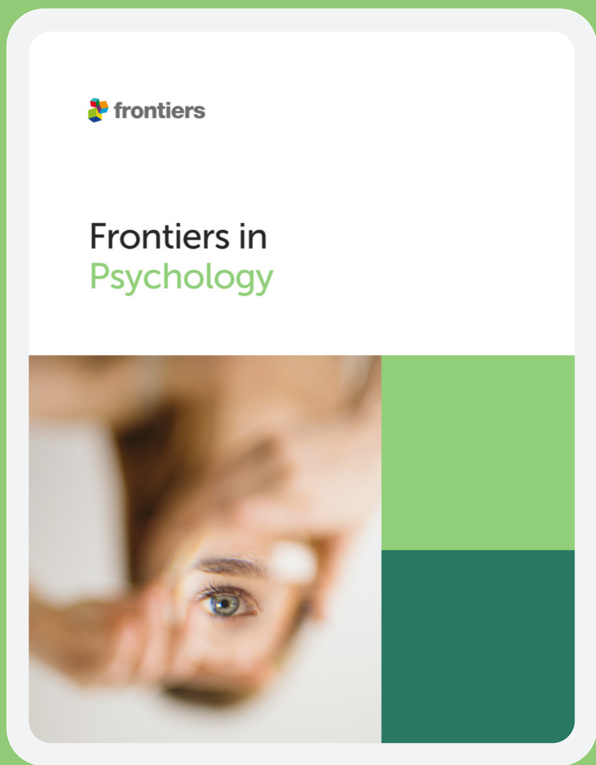
Frontiers in Psychology

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