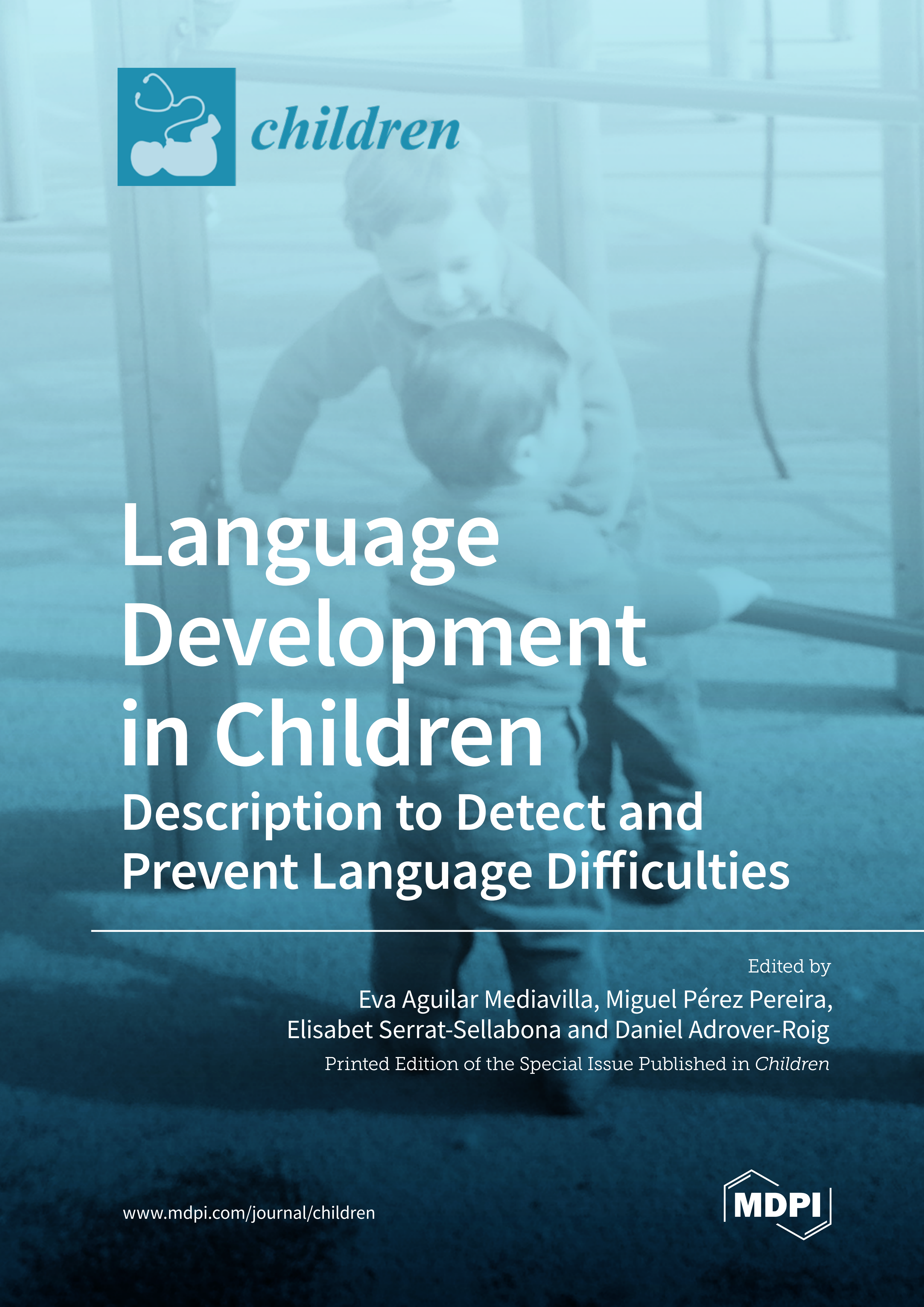


children

Language Development in Children

Description to Detect and Prevent Language Difficulties

Edited by

Eva Aguilar Mediavilla, Miguel Pérez Pereira,
Elisabet Serrat-Sellabona and Daniel Adrover-Roig

Printed Edition of the Special Issue Published in *Children*

www.mdpi.com/journal/children

MDPI

Language Development in Children: Description to Detect and Prevent Language Difficulties

Language Development in Children: Description to Detect and Prevent Language Difficulties

Editors

Eva Aguilar Mediavilla
Miguel Pérez Pereira
Elisabet Serrat-Sellabona
Daniel Adrover-Roig

MDPI • Basel • Beijing • Wuhan • Barcelona • Belgrade • Manchester • Tokyo • Cluj • Tianjin

Editors

Eva Aguilar Mediavilla
Institute for Educational Research and Innovation
University of the Balearic Islands
Palma
Spain

Miguel Pérez Pereira
Departament of Developmental Psychology and Education
University of Santiago de Compostela
Santiago de Compostela
Spain

Elisabet Serrat-Sellabona
Psycology
University of Girona
Girona
Spain

Daniel Adrover-Roig
Institute for Educational Research and Innovation
University of the Balearic Islands
Palma
Spain

Editorial Office

MDPI
St. Alban-Anlage 66
4052 Basel, Switzerland

This is a reprint of articles from the Special Issue published online in the open access journal *Children* (ISSN 2227-9067) (available at: www.mdpi.com/journal/children/special_issues/Language_Child).

For citation purposes, cite each article independently as indicated on the article page online and as indicated below:

LastName, A.A.; LastName, B.B.; LastName, C.C. Article Title. *Journal Name* **Year**, *Volume Number*, Page Range.

ISBN 978-3-0365-6841-6 (Hbk)
ISBN 978-3-0365-6840-9 (PDF)

Cover image courtesy of Eva Aguilar-Mediavilla

Contents

About the Editors

Eva Aguilar Mediavilla

Eva Aguilar Mediavilla holds a Degree and Ph.D. in Psychology and Speech & Language Therapy from the University of Barcelona. She has taught and researched at several universities (UB, UNED, UOC) and currently she is a Professor at the University of the Balearic Islands (UIB).

She currently holds the positions of Director of the Institute for Educational Research and Innovation (IRIE), coordinator of the Education Ph.D. program at this university, as well as presidency of the International Association for the Study of Language Acquisition (AEAL).

She is director of the Laboratory of Research in Development, Education and Language and principal investigator of the I + DEL research group of the IRIE of the UIB. Her research has focused on language acquisition, language difficulties, bilingualism, and phonology. She is the author of 45 articles, 23 of them indexed in JCR, in addition to being the author of 48 books and book chapters, including several evaluation tests. She has also participated in 134 national and international conferences, 12 of them as an invited speaker.

She has also won two awards of research in Language and Speech therapy 'Jordi Perelló' at the Catalan Speech and Language Therapist professional College.

Miguel Pérez Pereira

Miguel Pérez Pereira is an emeritus professor of Developmental and Educational Psychology at the University of Santiago de Compostela. My fundamental field of research has been Evolutionary Psychology, and more specifically the study of Language Development. I have researched on language development in Spanish-speaking or Galician-speaking children.

I have been the main researcher of several projects and contracts related to research on blind children, from which several publications in impact journals have been derived. Since 1998, I have developed several projects to elaborate the Galician version of the MacArthur-Bates scales (CDI). Such research has resulted not only in the construction of the scales, but also in obtaining a significant amount of data on the acquisition of language in Galician children. Since 2008, we have started a longitudinal project on language acquisition in preterm infants. Multiple publications in impact journals have been derived from these investigations, and it is the main topic of my research at present.

I have been the first President of the Official College of Psychologists (COP) in Galicia, and a member of the first Governing Boards of the COP in Spain. I have also been the founder and first President of the Association for the Study of Language Acquisition (AEAL), and I have had a very active participation in the International Association for the Study of Child Language (IASCL) of which I have been a member of the Committee Executive, and President of the Electoral Committee (2005).

I have been Associate Editor of Journal for the Study of Education and Development, and currently (since 2018) I am of the Revista de Logopedia, Foniatría y Audiología, and I belong to the Editorial Board of several journals (First Language...).

Elisabet Serrat-Sellabona

Elisabet Serrat-Sellabona is a senior lecturer in the Psychology Department of the Faculty of Education and Psychology in the Universitat de Girona. She has taught subjects and courses at undergraduate, graduate, postgraduate, master and doctoral levels on language and cognition, language disorders and language acquisition.

She has been director of the Department of Psychology, secretary of this department, coordinator of studies in the Faculty of Education and Psychology, as well as vice-dean of infrastructures. She is head of the 'Language and Cognition' Group, in the Department of Psychology. She has been president of the Association for the Study of Language Acquisition (AEAL). In addition, she has been responsible for national projects on the influence of language on socio-cognition. Her scientific production has been published in national and international scientific journals, as well as in books and book chapters. She has also directed or co-directed several doctoral theses. In terms of the subject matter of her scientific career, we can highlight a first stage in which her publications and participation in research projects focused on the study of language acquisition, both in the case of first and second languages. This line of research continues to the present day. In a second stage, she has also worked on the influence of language and executive functions on the development of socio-cognitive skills and, in particular, on the relationship between language and the development of emotional understanding. In addition, she has chaired the organising committee of an international conference on language acquisition, has participated in scientific committees of international conferences, as a project evaluator and as a reviewer for national and international journals.

Daniel Adrover-Roig

Daniel Adrover-Roig holds Ph.D. in Psychology from the University of the Balearic Islands (UIB) where he currently works as an associate professor (Department of Applied Pedagogy and Psychology of Education).

I have been a lecturer and researcher at several universities (Universitat Oberta de Catalunya). My pre-doctoral stays were carried out at the Universität Würzburg, Germany; Department of Experimental Psychology, Oxford University, United Kingdom. After conducting my Ph.D. on age-related behavioral and neural correlates I did a 2-year post-doctoral research stay at the Centre de Recherche de l'Institut Universitaire de Gériatrie de Montréal (Canada), focusing my studies on the associations of bilingualism with higher-order cognitive functions. Since 2011, my teaching has been centered on Developmental Psychology in School Age and Language Disorders in Primary Education. I have also been Vice-Dean of the Faculty of Education. I am privileged to head the Research Group in Development, Education, and Language (I+DEL). My research interests involve studying how bilingualism influences human cognition, with a particular interest in children with language difficulties, healthy adults, and older adults without cognitive impairment. In recent years, I have also been exploring the phenomenon of bullying and its relationship to educational needs and language difficulties, in particular, Developmental Language Disorder (DLD). I have supervised several doctoral theses on these topics and have been the principal investigator of two consecutive projects funded by the Ministry of Science and Innovation. Thanks to ongoing teamwork, I have been able to publish about 35 articles indexed in the Journal of Citation Reports (JCR) database and over 30 books and book chapters. I have also participated in more than 100 national and international conferences. Our research group was awarded the "Jordi Perelló" prize in research in Speech Therapy from the College of Speech Therapists of Catalonia, and in January 2022 the "La Caixa" Foundation awarded us a Flash Call in Social Sciences, with which we are expanding existing research on cyberbullying.

Preface to ”Language Development in Children: Description to Detect and Prevent Language Difficulties”

The present book is devoted to publish studies on language acquisition in children. We are especially focused on the description of language development and the variables affecting the early detection and prevention of language difficulties. Although language difficulties are very common (14% of children present a primary or secondary language difficulty), these difficulties are misdiagnosed due to the lack of visibility and updated knowledge from professionals of their long-term consequences in education and mental health. To prevent the misdiagnosed identification and assessment of language difficulties, more typical and atypical language studies are needed. In this sense, a good description of language acquisition could help to detect and prevent language difficulties. Nevertheless, most of the research on child language development has been conducted in English. However, studies in other languages and cross-linguistic studies have shown that some results about language development in English may not be directly transferred to other languages. Despite the increase in the number of studies, there is still little research about typical and atypical language acquisition in other languages and in bilingual populations. Therefore, this work aims to fill the current void in these studies, to give them visibility, and show the latest research about language acquisition in children. In this regard, this book addresses works with several perspectives of child language from a psycholinguistic, psychological, linguistic and/or educational point of view, including theoretical and empirical studies on typical and atypical child language acquisition and their association with other variables (either social or genetic) that could affect them.

Eva Aguilar Mediavilla, Miguel Pérez Pereira, Elisabet Serrat-Sellabona, and Daniel Adrover-Roig

Editors

children

Editorial

Introduction to Language Development in Children: Description to Detect and Prevent Language Difficulties

Eva Aguilar-Mediavilla [1,*], Miguel Pérez-Pereira [2], Elisabet Serrat-Sellabona [3] and Daniel Adrover-Roig [1]

1 Institute of Educational Research and Innovation, Universitat de les Illes Balears, 07122 Palma, Spain; daniel.adrover@uib.es
2 Department of Developmental and Educational Psychology, Faculty of Psychology, University of Santiago de Compostela, 15705 Santiago de Compostela, Spain; miguel.perez.pereira@usc.es
3 Department of Psychology, Universitat de Girona, 17004 Girona, Spain; elisabet.serrat@udg.edu
* Correspondence: eva.aguilar@uib.es; Tel.: +34-971172566

Citation: Aguilar-Mediavilla, E.; Pérez-Pereira, M.; Serrat-Sellabona, E.; Adrover-Roig, D. Introduction to Language Development in Children: Description to Detect and Prevent Language Difficulties. *Children* **2022**, *9*, 412. https://doi.org/10.3390/children9030412

Received: 9 February 2022
Accepted: 11 February 2022
Published: 14 March 2022

1. Introduction

The present Special Issue focuses on studies of language acquisition in children. We particularly addressed the description of language development and the variables affecting it for early detection and prevention of language difficulties. Although language difficulties are very common (14% of children present a primary or secondary language difficulty), these difficulties are misdiagnosed [1,2]. This might be due to the lack of visibility and the scarcity of knowledge in professionals in terms of the long-term consequences of language disorders in education and mental health. To prevent misdiagnosed identification and boost assessment of language difficulties, more typical and atypical language studies are needed. In this sense, a good description of language acquisition could help to detect and prevent language difficulties. Nevertheless, most of the research on child language development has been conducted in English. However, studies in other languages and cross-linguistic studies have shown that some results regarding language development in English may not be transferred into other languages [3]. Despite the increase in the number of studies, there is still a dearth of research on typical and atypical language acquisition in other languages and in bilingual populations. Therefore, this Special Issue aims to fill the current void in these studies, give them visibility, and show the latest research in language acquisition in children.

This Special Issue address child language from different perspectives. In this sense, it includes theoretical and empirical studies on typical and atypical child language acquisition. The contributions include studies about markers of language development in typical development, studies about language development in bilingual populations and several studies about language development in atypical populations including Developmental Language Disorder (DLD), reading disorders, Autism Spectrum Disorder (ASD), preterm children, hearing loss and genetic syndromes.

2. Markers of Language Development

Several studies in this Special Issue describe important factors that affect language development at different ages, thus depicting several key aspects to be considered in the prevention of communication and language difficulties throughout childhood. These studies range from the beginnings of word production [4] or gestures [5] to the impact of the use of technological devices in preadolescent children [6]. In addition, they cover different aspects or linguistic components, from phonetics [7], vocabulary [4], or syntax [8], to non-referential gestures related to narrative development [5].

The study by Serrat et al. [4] shows that prelinguistic factors have a greater weight than sociodemographic factors in explaining initial expressive vocabulary. This study shows that children under 18 months of age who imitate more are those who have a greater amount of vocabulary. In a related study, in the Special Issue Rujas et al. [8] shows that imitation or repetition of sentences, as an assessment task, is a useful tool for detecting language difficulties in older children. On the other hand, the study by Liu et al. [7] allows us to observe that the acoustic analysis of the production of certain consonants can provide accurate information on speech development and, therefore, is presented as an aspect to be considered in the evaluation of children's speech. In their review, Vilà-Giménez et al. [5] note another indicator, not much explored previously, according to which non-referential gestures act as predictors of narrative performance. This suggests that these gestures have important pragmatic functions that help to frame discourse. In older children, the study by Acebedo et al. [6] analyzes a variable that has a negative influence on language development: greater access to and use of media devices. The authors show that preadolescent children who use media devices more frequently and for communication purposes (not for school aid, or to learn new things) present lower language scores, without being influenced by sociodemographic factors.

In short, these studies show the importance of various markers of language development, both as indicators that may be related either to adequate development [4,5,7] or may contribute negatively to language development [6]. On the other hand, repetition of words (imitation) or sentences appears as an indicator of adequate development [4], as well as an important assessment tool to identify difficulties in language development [8].

3. Bilingual Development

In terms of bilingual language development, this Special Issue includes two studies. The first study by Kan et al. [9] explores the detection of language impairment in bilingual children by monolingual adults, and the second study by Diaz et al. [10] looks at the mutual longitudinal associations between vocabulary and executive functioning (EF) in monolingual and bilingual children.

As stated, Kan et al. [9] aimed at detecting the risk of language impairment in bilingual children by monolingual adults. The authors focused on how bilingual children's response speed during a narrative task can serve for categorizing language impairment. To do so, monolingual adults listened to several audio clips from an interactive story-retell task in both Cantonese and in English. Children were six sequential Cantonese-English bilinguals of 4 years of age; three of them had a language impairment and three were TD. Results showed that the interrater reliability was high for both languages, logistic regression and ROC curves revealed that adults were able to identify language impairment in bilingual children by judging their response speed, with higher sensitivity and specificity values in L1 conditions (Cantonese) than in L2 (English). These results highlight the potential relevance of looking at response speed to complement language assessment in bilingual children with language impairment.

Focusing on the potential links between EF and receptive vocabulary, Diaz et al. [10] tested monolingual and bilingual children with 4 years of age on average. The authors used a longitudinal approach with two temporal moments spread one year and departed from the theory of dual language processing as one of the potential sources for the frequently reported gains in EF in bilinguals. Their main goal was to determine whether EF exerted a direct influence on language proficiency or vice versa. Several measures of vocabulary and cognitive flexibility were administered to a sample of bilingual children and to a control group of monolingual preschool children. Results revealed that, only in the monolingual group, vocabulary at moment 1 predicted EF at moment 2. However, EF did not predict vocabulary at moment 2. The authors interpret the lack of longitudinal relations between EF and language abilities in the bilingual group together with the absence of differences in EF between both groups as a potential challenge to the purported advantage in EF in bilinguals and claim for the need of future similar studies.

4. Atypical Language Development

The Special Issue included several papers that focused on atypical language development, considering different conditions such as DLD, reading disorders, ASD, preterm, deafness and genetic syndromes.

4.1. Developmental Language Disorders (DLD)

DLD, previously named specific language impairment (SLI), is a persistent language delay affecting everyday social interactions or educational progress, in the absence of other biomedical conditions such as ASD, brain injury, hearing loss, genetic conditions or intellectual disability [11,12]. Four of the papers in this Special Issue focused on DLD, evidencing the increasing interest and the need of further studies of this atypical language condition. The works presented covered syntactic processing, lexical and syntactic errors, the use of non-word repetition task as a marker of DLD, and the relation between structural aspects of language, pragmatics, social cognition, and executive functions.

The work by Roa-Rojas et al. [13] explored a common error in Spanish children with DLD, the gender agreement in clitics, with a real-time processing technique of event-related brain potentials (ERP). Their results evidenced that children with DLD, contrary to their controls, did not show an enhanced anterior negativity between 250 and 500 ms post-target onset when they listened to gender-agreement violations. This result evidences a weaker lexical representation of morphosyntactic gender features in children with DLD.

Additionally, Kornev and Balčiūnienė [14] focused on the grammatical and lexical errors in children with DLD in narrative tasks in Russian. They found that the genre of discourse and age of assessment impacted not only the error distribution in children with DLD, but also in their controls, showing a relation between the cognitive load of the task and the number of errors produced. Their results support the resource deficit model that considers that the DLD is a delay in language performance but not in language competence, with errors being directly influenced by the cognitive demands of utterance and text production.

Following this hypothesis that children with DLD exhibit a limited cognitive load, and thus that language processing can easily overload their cognitive systems, non-word repetition has been proposed as a measure of the phonological working memory capacity and a marker of DLD [15,16]. In this sense, the work of Ahufinger et al. [17] explored the consistency of a non-word repetition task of 3-, 4-, 5- and 6-syllables presented in a random order and with varied wordlikeness ratings. Their results showed that the task discriminated correctly children with and without DLD (from 5 years and 16 years) speaking Catalan–Spanish (bilinguals) and European Portuguese (monolinguals). In this sense, children with DLD were less accurate repeating syllables than typical language developing (TD) children. Interestingly, children with DLD were more accurate repeating non-words with high wordlikeness than low, a pattern that had not been found in TD children. In addition, bilingual children performed worse than monolingual ones. Therefore, this task identified correctly children with DLD and differentiated them from TD children in the three languages (Catalan, Spanish and Portuguese) and in bilingual and monolingual children, making non-word repetition a promising task to detect children with DLD.

The last work in this section, by Andres-Roqueta et al. [18], focused on the association between the results of the parents' reports in the Children's Communication Checklist-2 (CCC-2) and several direct-child measures of structural language (phonetics, receptive and expressive grammar, receptive and expressive vocabulary and a composite score), pragmatics (receptive and expressive pragmatics and a composite score), social cognition (strange stories), and executive functions (sustained attention, inhibitory skills and a composite score). The results showed that children with DLD (between 3; 10 and 9 years old) performed worse than their TD peers in all the direct-child measures. The CCC-2 correlated with all direct child assessments in the group of DLD, but only formal measures of structural language predicted parent's reports in CCC-2. This indicates that CCC-2 answered by parents is a reliable measure to assess formal language, being structural language its best predictor.

4.2. Reading and Writing Disorders

Close to DLD and commonly comorbid with this disorder are reading and writing learning disabilities. Reading and writing disabilities are the most prevalent type of learning disabilities, with a prevalence between 7 and 10% and one of the main factors of school failure [19]. It includes impairments in reading decoding (i.e., letter–phoneme correspondence) resulting from problems in phonological processing skills and/or naming problems [20]. Children with RD also show impaired oral language skills, although not as severe as children with DLD [21].

One paper in the present Special Issue focused on reading and writing learning disabilities. González-Valenzuela et al. [22] explored the relationship between the type of delivery (vaginal or caesarean) and the occurrence of learning disabilities in reading (reading accuracy) and writing (phonetic and visual orthography), controlling for several gestational, obstetric, and neonatal variables (maternal age at delivery, gestational age, foetal presentation, Apgar 1, and new-born weight), in six-year-old children born in twin births. Their results showed a relation between the caesarean delivery and the presence of difficulties in reading accuracy, and phonetic and visual orthography. Although the authors advise that more evidence is needed, these findings could be useful in clinical practice to avoid the use of caesarean section on demand or without specialised indication.

4.3. Autism Spectrum Disorder (ASD)

Children with ASD show a communication deficit that sometimes is accompanied by formal language difficulties. Two papers in this Special Issue aboard the language and communication deficits in ASD.

One paper in this Special Issue looked at the integration of multimodal information within the communicative setting in toddlers at risk of developing ASD by means of eye-tracking measures. The study by Camero et al. [23] investigated visual attention to establish potential early markers of ASD. A group of 10 age-paired TD children and another group composed of 10 children with an increased likelihood of developing ASD looked at a human face when pronouncing pseudowords on a monitor, which were associated with several pseudo-objects. They found that children with higher odds of developing ASD showed a lower number of fixations to the eyes and larger number of gaze fixations to the mouth than the TD children. ASD children also had a slightly larger non-significant pupil dilation to faces, which was constant during the distinct task periods. They also looked more at the pseudo-object and for a longer time than TD children. In contrast, TD children showed a greater pupil dilation when hearing the pseudowords. The authors discuss that objective measures of eye tracking could be considered as potential markers for early detection of ASD and serve as relevant measures of word processing in both ASD and TD toddlers.

In another paper dedicated to ASD, Torrens and Ruiz [24] explored language and communication in preschool children with ASD compared to other developmental disorders using direct measures and parental reports of language. Results revealed that ASD children show a delay in language comprehension in contrast to language production, together with several problems in non-verbal communication, as compared with children with other developmental disorders. A high association was also observed between participant measures and parental reports of language and communication. These results lead the authors to suggest that complementing participants' measurements with parental reports is a valuable tool for language assessment. They also suggest that language comprehension deficits and difficulties in non-verbal communication might help diagnostic purposes between children with ASD and children with other neurodevelopmental disorders.

4.4. Preterm

Preterm children (very and extremely preterm in particular) are generally considered to present atypical language development. In this Special Issue, two papers are related to this topic. The first one by Pérez-Pereira [25] is a longitudinal study on the prevalence and determinants of language delay carried out with low-risk preterm children. The study spans the period between 10 and 60 months of age, with four measuring points. The participants were grouped into four groups of different gestational ages (GA) corresponding to (1) Extremely and Very preterm, (2) Moderately preterm, (3) Late preterm, and (4) Full-term children. Comparisons of the results obtained in the different language tests indicate that there are hardly any differences between the GA groups in the incidence of language delay (scores below the 10th percentile). The results found suggest that healthy preterm children do not seem to have a higher risk of language delay than full-term children. Logistic regression analyses permitted the identification of those factors that better predicted language delay at different ages, highlighting among these factors, previous language, and cognitive delay.

The second study by Joensuu et al. on the topic of preterm children's language development [26] investigates the associations between early language development at 2 years and literacy skills at seven years of age, in a sample of 136 very preterm (VP)/very low birth weight children (VLBW) and 137 full-term controls. Their results indicate that lexical production and MLU (Mean Length of Utterances) of the three longest utterances (measured through the Finnish CDI) and the expressive language score (from the Bayley Scales of Infant Development-II) are quite good predictors of prereading skills, reading, and writing at 7 years of age. In addition, most VP/VLBW children who were below the 10th percentile in language measures at 2 years of age had weak literacy skills at 7 years of age.

4.5. Deafness

Children with hearing impairment without hearing implants use to show a delay or difficulties in language development. Research with hearing children has revealed that the combination of music (rhyme and rhythm) with phonological awareness activities in intervention programs increments language outcomes. The paper by Holcomb and Wolbers [27] attempts to test the benefits of American Sign Language (ASL) rhyme, rhythm, and phonological awareness for deaf children. An intervention program was provided to five deaf children between 3 and 6 years of age to examine the effects of explicit handshape rhyme awareness instruction on increasing engagement behavior and accuracy in recitation. The findings indicate that recitation skills (although not engagement) in young deaf children can be supported through interventions utilizing ASL rhyme and rhythm supplemented with ASL phonological awareness activities.

4.6. Genetic Syndromes

Most genetic syndromes involve cognitive and language developmental impairments. In the present Special Issue, the study by Zanaboni et al. [28] investigates oral motor, speech and language abilities of eight Italian-speaking children (aged 4.6 to 15.4 years) with glucose transporter type 1 deficiency syndrome (GLUT1DS). This syndrome, due to mutations in SLC2A1 gene, implies impaired glucose transport into the brain. Congruently, patients are treated with the ketogenic diet (KD) to meet the energy demands of the developing brain, as it was the case for the participants in this study. The children were assessed with different standardized tests. The results indicated that the patients showed deficits in orofacial praxis, the speech domain, and the language domain (semantic/phonological fluency and receptive grammar, in particular), as well as in the development of several cognitive functions. The authors highlight the importance of a complete speech and language evaluation in GLUT1DS patients to obtain a typical linguistic phenotype, which could guide and improve early diagnosis and intervention.

5. Conclusions

The present Special Issue focuses on the major topics of typical and atypical language development with monolingual and bilingual children, covering new and highly innovative studies that have increased the evidence for detecting and preventing language impairment especially in several languages such as Spanish, Catalan, Portuguese, Italian, Russian, Cantonese, Finnish, and American Sign Language.

Author Contributions: Writing—original draft preparation, E.A.-M., M.P.-P., E.S.-S., D.A.-R.; writing—review and editing, E.A.-M., M.P.-P., E.S.-S., D.A.-R. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Grant EDU2017-85909-P funded by MCIN/AEI/10.13039/501100011033 and by "ERDF A way of making Europe". The funding sources was not involved in the study design, analysis, and interpretation of data, and writing of the report or submission for publication.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: We would like to thank all the authors and reviewers who offered their manuscripts and their constructive comments for this Special Issue.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Tomblin, J.B.; Records, N.L.; Buckwalter, P.; Zhang, X.; Smith, E.; O'Brien, M. Prevalence of Specific Language Impairment in kindergarten children. *J. Speech Lang. Hear. Res.* **1997**, *40*, 1245–1260. [CrossRef] [PubMed]
2. García-Mateos, M.; Mayor Cinca, M.Á.; de Santiago Herrero, J.; Zubiauz de Pedro, B. Prevalencia de las patologías del habla, del lenguaje y de la comunicación. Un estudio a lo largo del ciclo vital. *Rev. De Logop. Foniatría Y Audiol.* **2014**, *34*, 163–170. [CrossRef]
3. Leonard, L.B. Specific language impairment across languages. *Child Dev. Perspect.* **2014**, *8*, 1–5. [CrossRef] [PubMed]
4. Serrat-Sellabona, E.; Aguilar-Mediavilla, E.; Sanz-Torrent, M.; Andreu, L.; Amadó, A.; Serra, M. Sociodemographic and pre-linguistic factors in early vocabulary acquisition. *Children* **2021**, *8*, 206. [CrossRef]
5. Vilà-Giménez, I.; Prieto, P. The Value of Non-Referential Gestures: A Systematic Review of Their Cognitive and Linguistic Effects in Children's Language Development. *Children* **2021**, *8*, 148. [CrossRef]
6. Acebedo, L.; Buil-Legaz, L.; Adrover-Roig, D.; Aguilar-Mediavilla, E. Impact of the Use of Media Devices within the Family Context on the Language of Preteens. *Children* **2020**, *7*, 281. [CrossRef]
7. Liu, C.-T. A First Step toward the Clinical Application of Landmark-Based Acoustic Analysis in Child Mandarin. *Children* **2021**, *8*, 159. [CrossRef]
8. Rujas, I.; Mariscal, S.; Murillo, E.; Lázaro, M. Sentence Repetition Tasks to Detect and Prevent Language Difficulties: A Scoping Review. *Children* **2021**, *8*, 578. [CrossRef]
9. Kan, P.F.; Miller, A.; Still, S. Identifying Bilingual Children at Risk for Language Impairment: The Implication of Children's Response Speed in Narrative Contexts. *Children* **2021**, *8*, 62. [CrossRef]
10. Diaz, V.; Borjas, M.; Farrar, M.J. Is there an association between executive function and receptive vocabulary in bilingual children? A longitudinal examination. *Children* **2021**, *8*, 44. [CrossRef]
11. Bishop, D.V.M.; Snowling, M.J.; Thompson, P.A.; Greenhalgh, T.; CATALISE-2 Consortium. Phase 2 of CATALISE: A multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *J. Child Psychol. Psychiatry* **2017**, *58*, 1068–1080. [CrossRef] [PubMed]
12. Bishop, D.V.M.; Snowling, M.J.; Thompson, P.A.; Greenhalgh, T.; CATALISE Consortium. CATALISE: A multinational and multidisciplinary Delphi consensus study. 1. Identifying language impairments in children. *PLoS ONE* **2016**, *11*, e0168066. [CrossRef] [PubMed]
13. Roa-Rojas, P.; Grinstead, J.; Silva-Pereyra, J.; Fernández, T.; Rodríguez-Camacho, M. Syntactic Gender Agreement Processing on Direct-Object Clitics by Spanish-Speaking Children with Developmental Language Disorder: Evidence from ERP. *Children* **2021**, *8*, 175. [CrossRef]
14. Kornev, A.N.; Balčiūnienė, I. Lexical and Grammatical Errors in Developmentally Language Disordered and Typically Developed Children: The Impact of Age and Discourse Genre. *Children* **2021**, *8*, 1114. [CrossRef]
15. Baddeley, A.D. *Human Memory: Theory and Practice*; Psychology Press: New York, NY, USA, 1997.
16. Conti-Ramsden, G.; Botting, N.; Faragher, B. Psycholinguistic markers for Specific Language Impairment (SLI). *J. Child Psychol. Psychiatry* **2001**, *42*, 741–748. [CrossRef] [PubMed]

17. Ahufinger, N.; Berglund-Barraza, A.; Cruz-Santos, A.; Ferinu, L.; Andreu, L.; Sanz-Torrent, M.; Evans, J.L. Consistency of a Nonword Repetition Task to Discriminate Children with and without Developmental Language Disorder in Catalan–Spanish and European Portuguese Speaking Children. *Children* **2021**, *8*, 85. [CrossRef] [PubMed]
18. Andrés-Roqueta, C.; Garcia-Molina, I.; Flores-Buils, R. Association between ccc-2 and structural language, pragmatics, social cognition, and executive functions in children with developmental language disorder. *Children* **2021**, *8*, 123. [CrossRef]
19. Carrillo, M.S.; Alegría, J.; Miranda, P.; Sánchez Pérez, N. Evaluación de la dislexia en la escuela primaria: Prevalencia en español [Assessment of dyslexia in primary school: Prevalence in Spanish]. *Escritos Psicol. Psychol. Writ.* **2011**, *4*, 35–44. [CrossRef]
20. Ramus, F.; Marshall, C.R.; Rosen, S.; Van Der Lely, H.K.J. Phonological deficits in specific language impairment and developmental dyslexia: Towards a multidimensional model. *Brain* **2013**, *136*, 630–645. [CrossRef]
21. Bishop, D.V.M.; Snowling, M.J. Developmental dyslexia and specific language impairment: Same or different? *Psychol. Bull.* **2004**, *130*, 858. [CrossRef]
22. González-Valenzuela, M.-J.; López-Montiel, D.; Cazorla-Granados, O.; González-Mesa, E.-S. Learning Disabilities in Reading and Writing and Type of Delivery in Twin Births. *Children* **2021**, *8*, 834. [CrossRef] [PubMed]
23. Camero, R.; Martínez, V.; Gallego, C. Gaze Following and Pupil Dilation as Early Diagnostic Markers of Autism in Toddlers. *Children* **2021**, *8*, 113. [CrossRef] [PubMed]
24. Torrens, V.; Ruiz, C. Language and communication in preschool children with autism and other developmental disorders. *Children* **2021**, *8*, 192. [CrossRef] [PubMed]
25. Pérez-Pereira, M. Prevalence of Language Delay among Healthy Preterm Children, Language Outcomes and Predictive Factors. *Children* **2021**, *8*, 282. [CrossRef] [PubMed]
26. Joensuu, E.; Munck, P.; Setänen, S.; Lipsanen, J.; Huhtala, M.; Lapinleimu, H.; Stolt, S.K.J. Associations between Language at 2 Years and Literacy Skills at 7 Years in Preterm Children Born at Very Early Gestational Age and/or with Very Low Birth Weight. *Children* **2021**, *8*, 510. [CrossRef] [PubMed]
27. Holcomb, L.; Wolbers, K. Effects of asl rhyme and rhythm on deaf children's engagement behavior and accuracy in recitation: Evidence from a single case design. *Children* **2020**, *7*, 256. [CrossRef] [PubMed]
28. Zanaboni, M.P.; Pasca, L.; Villa, B.V.; Faggio, A.; Grumi, S.; Provenzi, L.; Varesio, C.; De Giorgis, V. Characterization of Speech and Language Phenotype in GLUT1DS. *Children* **2021**, *8*, 344. [CrossRef] [PubMed]

children

Article

Sociodemographic and Pre-Linguistic Factors in Early Vocabulary Acquisition

Elisabet Serrat-Sellabona [1,*], Eva Aguilar-Mediavilla [2,*], Mònica Sanz-Torrent [3], Llorenç Andreu [4], Anna Amadó [1] and Miquel Serra [3]

1 Psychology Department, Universitat de Girona, 17004 Girona, Spain; anna.amado@udg.edu
2 Applied Pedagogy and Educational Psychology, Institute of Research and Innovation in Education (IRIE), Universitat de les Illes Balears, 07122 Palma, Spain
3 Psychology Faculty, Universitat de Barcelona, 08035 Barcelona, Spain; monicasanz@ub.edu (M.S.-T.); miquel.serra@ub.edu (M.S.)
4 Psychology and Education Science Studies, Universitat Oberta de Catalunya, 08018 Barcelona, Spain; landreub@uoc.edu
* Correspondence: elisabet.serrat@udg.edu (E.S.-S.); eva.aguilar@uib.es (E.A.-M.); Tel.: +34-972-418-754 (E.S.-S.); +34-971-172-566 (E.A.-M.)

Abstract: Here, we studied the beginnings of language development, jointly assessing two groups of precursors, sociodemographic and pre-linguistic, that have previously been studied separately. Thus, the general objective of this study was to explore which factors best explained the acquisition of initial expressive vocabulary. The sample consisted of 504 participants from Catalan-speaking homes with ages ranging between 10 and 18 months. The data were obtained through the MacArthur–Bates Communicative Development Inventories (MCB-CDIs). Vocabulary development shows a lexical spurt at 17 months. Regression analyses show that pre-linguistic factors have more explanatory power of than sociodemographic ones. Within the sociodemographic variables, age, birth order and birth weight explain part of the vocabulary variance. With respect to pre-linguistic variables, imitation, late gestures and phrase comprehension are predictors of the initial vocabulary acquisition. Specifically, imitation and late gestures were the pre-linguistic behaviours that made it possible to distinguish between children with higher and lower levels of vocabulary. We discussed these findings in relation to their relevance for language acquisition and for the early assessment of linguistic competence.

Keywords: MacArthur-Bates CDI; Catalan; lexical spurt; sex; birth order; birth weight; parental education; imitation; gestures; comprehension

Citation: Serrat-Sellabona, E.; Aguilar-Mediavilla, E.; Sanz-Torrent, M.; Andreu, L.; Amadó, A.; Serra, M. Sociodemographic and Pre-Linguistic Factors in Early Vocabulary Acquisition. *Children* **2021**, *8*, 206. https://doi.org/10.3390/children8030206

Academic Editor: Eric Dearing

Received: 29 December 2020
Accepted: 3 March 2021
Published: 9 March 2021

1. Introduction

The aim of this work was to study the influence of factors traditionally considered related to initial language acquisition. Until now, personal, and sociodemographic factors have often been considered separately from pre-linguistic factors, but here we present a joint evaluation of how these two factors affect initial language acquisition.

The cognitive abilities of children and their relationship with the environment are two variables that are in constant interaction and are responsible for determining a child's communicative, cognitive, and affective development. With regard to communicative development, typically developing children have already discovered patterns of meaning in speech by the end of their first year of life [1,2]. It is around this time that children start to discover connections between language and the world around them [3,4]. Any disruption at this initial stage will affect a child's linguistic development as well as other related forms of development and subsequent learning experiences [5].

Thus, it is during the first and second years of life that the foundations for communicative and linguistic development are laid. While this kind of development is important enough in itself, it also has repercussions for other aspects of development. Accordingly, our study recognizes the relevance of exploring and evaluating this initial stage of language development and the factors that influence it. To examine some of these factors more closely, we first review the sociodemographic variables that have been considered to play a role in language development, and then examine some of the pre-linguistic skills also considered relevant during initial language acquisition.

1.1. Sociodemographic Variables Related to Language Acquisition

Previous studies have shown that some demographic, personal, and social variables are related to language development. These include sex, birth weight, history of ear infections, birth order, parents' level of education, and parents' education level, and their socio-cultural and economic status.

Although the male sex has traditionally been linked with lower language abilities [6] and a greater prevalence of language difficulties [7], there is still no evidence of any biological causes that explain this [8]. Besides, more recent and large-scale studies have found a more balanced ratio and fewer differences between boys and girls [9]. With respect to brain differences related to language development, a systematic review by Etchell et al. [8] showed that brain differences between the sexes may be more prominent during certain developmental stages but are negligible in other stages, suggesting that such differences are not as significant as previously thought. However, most studies that evaluate linguistic performance between boys and girls during language acquisition have found differences between the sexes. For example, Huttenlocher et al. [10] found sex differences related to lexical growth in children from 22 to 26 months. Galsworthy et al. [11] found that 2-year-old girls outperformed boys in verbal development. Various large-scale studies conducted using the MacArthur–Bates Communicative Development Inventories (CDIs) [12] found early sex differences that increase with age in children from 8 to 36 months, both from English and non-English-speaking backgrounds, with girls outperforming boys in early communicative gestures, vocabulary, and word combinations [13–16]. However, other studies also conducted using the CDIs, such as Kovačević et al. [17], Jackson-Maldonado et al. [18], and Berglund and Eriksson [19], failed to find differences in language development between boys and girls.

Birth weight has also been related with language development. Children born with a low weight are at higher risk of experiencing language problems [20–22]. The risk of language problems increases the lower birth weight was [23]. The relation between birth weight and language development is also influenced by other factors such as medical complications and born prematurely [22]. Thus, healthy children with weights upper than 1900 gr. at born show language abilities adequate to their gestational age [24]. Contrary, a weight lower than 1500 gr. at born is related with a higher incidence of medical complications such as conductive hearing loss, which can interfere with language acquisition and is considered a risk factor for deafness [25].

Despite the fact that hearing loss is a clear factor in delaying spoken language development, the association between otitis (i.e., inflammation that occurs within the middle-ear cavity and causes mild-to-moderate hearing loss) and language development is not so clear [26]. Although traditionally [27] otitis has been linked with a major risk of suffering language development delays, a review by Roberts et al. [28] found that otitis media with effusion may not be a substantial risk factor for later speech and language development in typically developing children. Despite these results, otitis is still considered a risk factor for language delay in clinical settings [29].

Birth order is another variable that has traditionally been linked with language development, with firstborns exhibiting better language abilities than later-born children [13]. In this regard, Fenson et al. [14] noted significant negative correlations between birth order and gestures, vocabulary production, Mean Length of Utterances (MLU) and word combinations measured using the CDI. Berglund et al. [13] found better vocabulary comprehension and production in firstborns assessed at 18 months using the Swedish Early Communicative Development Inventories. Firstborn advantage in language acquisition has not only been found using parent inventories, but also through direct observations. For example, by observing natural language, Hoff-Ginsberg [30] showed that firstborns had an advantage in lexical and grammar development at 18 to 29 months. Nevertheless, the evidence of quantitative differences between firstborn and later-born infants is inconclusive; for example, Pine [31] found that firstborns reached the 50-word milestone earlier than secondborns, but found no differences in reaching the 100-word milestone. Oshima-Takane et al. [32] reported no differences in the MLU, number of intelligible utterances, total vocabulary (types) and total number of words (tokens) between firstborns and secondborns at 21 and 24 months, but an advanced use of pronoun productions was found in secondborns.

Some social aspects have also been linked with language development. Of particular interest to researchers is the relationship between socioeconomic status (SES) and language outcomes, which has been found to be incredibly convoluted and complex [30,33–37]. For example, socioeconomic status is decisive for other variables that can affect language development such as the family home, neighbourhood, child's school, and the resources to which he/she has access [38]. Besides, the relation between SES and language is mediated by other variables, such as parent's educational level, cultural differences and the linguistic input that the child receives [38–40]. Previous studies have highlighted the importance of quantity of input exposure, but most acknowledge that the quality of language that the child is exposed to is more salient [41,42]. In this sense, some studies [33,38,43] have found that a higher maternal education could be a protective factor against language difficulties and a predictor of better child language development. For example, Hirsh-Pasek et al. [36] show in low incoming families that maternal education is related with sensitive parenting, the quantity of language input the child received, the quality of communicative interactions (e.g., use of routines and rituals) and child expressive language. Accordingly, previous studies have found maternal educational level related to early language development, as measured by the MacArthur Communicative Development Inventories [44,45], although others have failed to find this relationship when linguistically, culturally, and developmentally appropriate instruments are used [46].

1.2. Pre-Linguistic Factors

During the pre-linguistic period, children's language develops as follows: (a) a focus on the sounds of speech; (b) understanding first words; (c) communicating needs through language; (d) random vocalizations; and (e) uttering familiar speech sounds. At this time, they also progress from an initial multimodal perception of their postnatal environment to attaching symbolic representations and references to actions, objects, and significant people. While studying this period, researchers have focused mainly on intentional communication, vocalizations, and gestures as precursors and possibly facilitators of a child's first words [47,48].

1.2.1. First Signs of Understanding

Over recent decades, many studies have focused on how babies know and comprehend aspects of the language in early development and before they say their first words. These studies show how babies recognize the voice of their mothers [49,50] and are able to discriminate their language from others [51]. Saffran et al. [52,53] showed how children are able to extract regularities of speech and recognize parts and patterns in the flow of speech heard. These authors found that after just 2 minutes of exposure, 8-month-old infants could extract words embedded in a continuous stream of spoken artificial language. This type of learning has been called statistical language learning [52]. It has also been shown that crying [54] and the intonation of babbling vary depending on the language of exposure [55], which is a sign that children are attentive and analyzing the speech to which they are exposed and trying to approximate those patterns.

Before they say their first words, babies can also understand the pragmatic intention of adults from highly context-dependent situational clues. At 9 months, children understand some words and expressions of adults. They react to their name or respond in some way to very specific words or expressions spoken with a certain intonation and in repetitive or familiar contexts [48,56]. Around the first year of life, children already react to some words or expressions and understand some very simple instructions or phrases related to routines or very familiar situations such as "A dormir" (to bed), "Ja està" (it is over), "Què vols?" (what do you want?) [48,56]. For example, at 12 months, babies would move their head in response to their own name and can begin to understand simple commands or phrases related to routines. At 18 months, they can understand simple commands (one step) such as "Put it here", "Give me a kiss" or "Say goodbye" [57–59]. Additionally, at around the age of 2 years, children use syntactic clues, such as word order, to understand transitive sentences [60,61].

It is possible that the difficulties in understanding spoken language at early ages is one of the main predictors of experiencing a language development disorder later on [62]. In line with this, several studies have found a strong correlation between sentence comprehension in the first 12–18 months and subsequent language level [45,63]. For example, Watt et al. [64] analysed which pre-linguistic skills and behaviours at 1 and 2 years of age predict language abilities at 3 years of age, using communication and symbolic behaviour scales to measure this [65]. The results showed that early comprehension abilities predict subsequent receptive and expressive language outcomes. Several studies have been conducted to assess these early signs of understanding in children with communication or language difficulties. This is the case in a study of children at risk of autistic spectrum disorder (ASD), which observed low scores in the children's social interest or in responding to theirs names and in understanding initial sentences [66]. Another study conducted with English-speaking infants with and without ASD showed that parents reported fewer sentences understood and fewer gestures produced by 12 months of age in children at risk of ASD measured using the CDI. Luyster et al. [67] reported similar results with "first signs of understanding" and "understanding of phrases" in children with ASD. Finally, Charman et al. [68] also observed, using parental reports, delays in early signs of understanding (e.g., "reacting to mother's/father's name") in children with ASD. In this same line of research, the new conception of Developmental Language Disorders (DLD) includes the presence of comprehension problems between the ages of 2 and 3 years as a factor of early detection, which correlates with the subsequent diagnosis [62].

1.2.2. Imitation

Imitation is another precursor to language development. Verbal imitation, or the repetition of new words or parts of sentences, is a pervasive and innate behavior in early development and is used for diverse functions during language acquisition [69]. One of these functions is to internalize language [70]; in this sense, during our everyday interactions we can see that as children learn language, they spontaneously imitate the speech of those around them. Despite the importance of imitation, few studies have been conducted on its role in language acquisition since the seminal studies by Snow in the 1980s. These first studies focused on verbal imitation and mimicry in the early stages. In accordance with Snow [71], the results of these studies can be placed on a continuum that ranges from the non-contribution of verbal imitation in language development [72,73] to the idea that imitation is at least partially credited for parts of a child's language development such as the acquisition of vocabulary [74–76], grammar [77], morphology and syntax [78,79].

With regard to the development of expressive vocabulary, in a recent study conducted by Masur and Olson [80], children who demonstrated more verbal imitations of the language produced by their mothers were found to have a more advanced vocabulary at 17 and/or 21 months. Research carried out with children with atypical development, such as that of Feeley and Jones [81] on children with Down syndrome, or that of Yoder and Layton [82] and Smith et al. [83] on children with ASD, also describe this relationship between verbal imitation and expressive vocabulary.

Studies that deny the contribution of imitation argue that there is considerable individual variance in imitation among children and that only a subgroup of children learn language, or part of it, through imitation. On the other hand, studies that accept a partial contribution of imitation to language development base their assumption on the idea that children imitate syntactic structures when they cannot produce them spontaneously.

Contemporary research has focused primarily on the role of socio-cognitive abilities in verbal imitation, such as understanding the intentions of others or the context in which the sentence is produced. The results of studies conducted by Over and Gattis [84] and Bannard et al. [85] showed that children use the intention perceived in others [84,85] and the functional context of an utterance [85] to imitate a verbal model.

Finally, a last group of studies provide evidence in favour of verbal over non-verbal imitation in humans and the diversity of purposes of the former. For example, instrumental imitation may have the purpose of (a) transmitting language from one individual to another; (b) engaging in a conversation; and (c) establishing affiliation with others [86–89]. Beyond these purposes we can add an additional purpose from the study conducted by Matthews et al. [90] that is unique to the verbal domain: (d) facilitating communication, because the conversation becomes more efficient when speakers construct referential pacts.

1.2.3. Gestures and Actions

As several authors have recognized, gestures are one of the most important precursors of language acquisition [91,92]. In this sense, gestures have been considered as behaviours that precede and prepare the emergence of expressive language [93].

Some authors have focused on deictic gestures, such as the gesture of pointing, which is considered a precursor of child vocabulary. Nevertheless, more research is needed to establish whether it can be considered a predictor of language development (see [94]), because authors such as McGillion et al. [95] found that the presence of pointing gesture does not predict expressive vocabulary, although the presence of pointing gestures is related to receptive vocabulary.

The communicative gestures that children make during their first years are not limited to the gesture of pointing. Nelson [96] suggested that gestures and actions, including those integrated in symbolic play, contribute to the development of the representational abilities that are fundamental to language acquisition. A longitudinal study by Cadime et al. [97], with the MacArthur Bates CDI-I, found that gestures predicted vocabulary comprehension at 9, 12, and 15 months, although gestures only predicted expressive vocabulary at 12 months.

Regarding more complex gestures and actions, such as actions that children perform with dolls or through games in which they imitate the actions of adults, it is important to note that these actions can be interpreted as part of symbolic play. In this sense, the relationship between language development and symbolic play has been shown to be robust throughout development [98,99]. Furthermore, despite the fact that there are different interpretations of this relationship, symbolic play can be considered a precursor of language [98].

1.3. Present Study

Given the results of the studies reviewed above, several sociodemographic and pre-linguistic factors can influence language acquisition. This study seeks to add to the field by investigating two main questions:

RQ1: Which of the sociodemographic or pre-linguistic variable(s) studied explain early vocabulary acquisition?

RQ2: Which of the sociodemographic or pre-linguistic variable(s) studied discriminate children with a high level of vocabulary from those with a low level of vocabulary?

The influence of sociodemographic and pre-linguistic factors in early language development would be evident in the comprehension and production of first words. In this study, we focus on vocabulary production, as it is a more valid measure of the MacArthur

Bates CD1-I questionnaire in early learners [100]. As we have previously stated, this work aims to study the beginnings of language development, jointly assessing two groups of precursors that have previously been studied separately, that is, sociodemographic and pre-linguistic factors. To ensure that the data are treated collectively rather than separately, data were collected using a single instrument, the MacArthur–Bates CDI, which has been shown to be reliable and valid for child language assessment [14].

Thus, the general objective of this study was to explore which factors best explain the acquisition of initial expressive vocabulary and to what extent they do so. We aimed to describe the course taken in the initial acquisition of expressive vocabulary through the use of parental reports. In this study, this description will allow us to verify that our data conformed to the expected course of vocabulary across the ages studied. Then, we aimed to explore and quantify which factors have the most explanatory power in terms of the acquisition of initial vocabulary, contrasting personal and sociodemographic factors with pre-linguistic factors. Finally, we aimed to measure the contribution of these factors to discriminate children with a high level of vocabulary (>25th percentile) from those who are at risk of suffering delays in their language acquisition. A better understanding of the influence of these aspects on language acquisition is important to design effective assessment tools and interventions.

2. Materials and Methods

2.1. Participants

The total sample consisted of 504 participants (259 girls) from Catalan-speaking homes with ages ranging from 10 and 18 months (M age = 14.23; SD = 2.5). Premature children with a weight below 1900 g were excluded due to the medical complications associated with this condition. Table 1 shows the characteristics of the sample, including sociodemographic factors.

Table 1. Main descriptive data of the participants.

Personal and Sociodemographic Characteristics	*N* = 504
Age in months, *M* (SD)	14.23 (2.5)
Sex as % female	51.4
Birth weight in kg, *M* (SD)	3.26 (0.48)
Number of ear infections per year, *M* (SD)	0.52 (1.29)
Birth order in % of children	
First	56.8
Second	36.8
Third	5.4
Fourth, onwards	1
Mother's educational level, %	
No studies	0.2
Primary	4.4
Secondary	29
University	66.4
Father's educational level, %	
No studies	0.2
Primary	12.3
Secondary	40.9
University	46.6

Note: *M* = mean; SD = standard deviation.

2.2. Materials

The data in this study were obtained using the MacArthur–Bates Communicative Development Inventories (MCB-CDIs) adapted for Catalan [101,102]. Specifically, CDI-I was used for this study as it is appropriate for children between the ages of 8 and 18 months. The inventory has two main sections with different sub-sections.

For the first part, the parents were asked about their child's first words based on his/her first signs of understanding (the child's name, "no", and the names of the parents), how the child understands frequently spoken phrases, the child's capacity to imitate language, and the list of vocabulary understood and produced. In the second part of the instrument, the parents were asked about the child's gestures and actions through the use of first gestures of intentional communications, games with adults, turn-taking routines, actions with objects, and symbolic play (e.g., with dolls, imitating adults, and using objects for a different purpose). In the current study, most of the sections of the MCB-CDI-I were considered as independent variables, except the section "Checklist of total vocabulary" that was considered as dependent variable.

The first part of the MCB-CDI-I was divided in several sections as described below. The section "First signs of understanding" contained three items and recorded whether the children reacted when hearing certain words. Specifically, the items in this section referred to whether the child stopped what he/she is doing upon hearing "no" said to him/her; whether he/she responds when called by their name; and whether the child looks around when hearing his/her mother or father be called by name. In the section "Phrases", which contained 27 phrases or utterances, the number of phrases that the children understood from the section was counted. This section refers to the ability of the children to understand frequently spoken linguistic utterances in speech directed at the children [57–59]. Examples of this type of utterance are: "Què és això?" (What is that?), "A dormir" (Go to sleep), "Fes-me un petó" (Give me a kiss), "Quiet!" (Stop!), "Digues adéu" (Say bye-bye), etc. The section "Starting to talk" contained two items (Imitation and Naming), and in this work we only used "Imitation" because "Naming" was considered to be the same as or similar to the "Vocabulary Checklist". In the "Imitation" section the parents were asked whether the children imitate any words or parts of phrases. The possible answers were "Not yet", "Sometimes", and "Often". The section "Checklist of total vocabulary" contained items from different lexical categories that the child "can understand" or "understand and produce", such as sound effects and animal sounds, animals (real or toys), vehicles (real or toys), toys, food and drink, clothes, parts of the body, furniture, domestic objects, objects from outside the home and places to visit, people, games, routines and social formulas, actions, times, qualities and attributes, pronouns and possessive and demonstrative pronouns, questions, prepositions, quantifiers, and articles.

The second part of the MCB-CDI-I evaluates gestures and actions and was divided into two main sections, as described below. The first section is called "First gestures and actions", and contained a sub-section of communicative gestures (e.g., saying "bye" with your hand) and a sub-section of nursery rhymes, children's songs and routines (e.g., peek-a-boo). The second section is called "Late gestures and actions", and contained three sub-sections: (1) Performing actions with objects such as the child putting the telephone to his/her ear; (2) playing at being an adult, where the participants are asked about symbolic play activities with a doll (e.g., combing a doll's hair); and (3) pretending or trying to do adult activities (e.g., pretending to take photos, pretending to sweep, etc.). This section includes many actions or activities of symbolic play.

Several bibliographic sources on language acquisition show that parental reports are reliable and valid, and represent the linguistic abilities of the children in the short and long term [12,103,104]. It is worth noting that Marchmann and Martínez-Sussman [105] have referred to a high concurrent validity between the development of productive vocabulary measured using the CDI questionnaire and that measured in laboratories. In addition, the MacArthur–Bates questionnaire offers the possibility to analyse the communicative and linguistic development of broad samples of participants, as in our case (see [12]). Specifically, the Catalan version of the MCB-CDI-I has an internal consistency of $\alpha = 0.893$; a test-retest reliability of $\alpha = 0.800$ for word comprehension; and a concurrent validity (MCB-CDI-II inventory with display of spontaneous speech) of $r = 0.577$ [102].

2.3. Procedure

The participants were recruited through professional and personal contacts of the authors of the adapted questionnaires, as well as through the participation of several child education centres. The CDI-I forms were delivered to the families either personally, in which case they were given instructions on paper that were briefly discussed, or through the early childhood education centres (0–3 years) by giving families an information letter and consent form, with the instructions provided later along with the booklet. The instrument itself contained instructions in each of the sections and the families were explicitly informed, either verbally or through the letter and informed consent, that they should only record the words/usage that their child produced in any variant of the Catalan language, even if there was a mispronunciation.

The study used the total vocabulary produced by the children as a dependent variable. To calculate the size of each participant's vocabulary, the total number of words marked by the parents in the "Checklist of total vocabulary section" (one point for each word) was added up. The maximum possible point score was 423 (total number of items in the list).

The information items in the "General information" section of the MCB-CDI-I were used as independent personal and sociodemographic variables—sex, birth order, birth weight, how many ear infections per year, and mother's and father's level of education. As for the pre-linguistic independent variables, all sections were scored as indicated in the scoring manual for the instrument.

A stepwise multivariate regression analysis was performed. Preliminary analyses were conducted to ensure that the assumptions of the multiple regression were met. A logistic regression analysis was also carried out. In this case, the dependent variable was dichotomized based on a child's normative scores [102]. Participants were grouped according to whether they showed a high level of vocabulary ($\geq$25th) or a low level of vocabulary (<25th). The data were analysed using SPSS v. 23.

3. Results

First, curvilinear estimation (a type of regression analysis) was used to determine which model best fit the course of acquisition of expressive vocabulary. Linear, quadratic, and exponential models were chosen for the analysis, as they are appropriate for the field of child development [106].

It was observed that the vocabulary acquisition process began slowly, increasing gradually until 16 months of age. However, from the age of 17 months, a substantial change in trend was noted (see Figure 1). The children's vocabulary at this age increases from two words at 10 months to around 20 words on average at 16 months. From this age onwards, and in just two months, the vocabulary increases to over 60 words at 18 months.

Accordingly, the results of the curvilinear estimation reflect the fact that the linear, quadratic, and exponential models adequately conform to the curve. The ANOVA of each model gave the following statistical values: linear ($F(1.502) = 132.276$; $p < 0.001$; $R^2 = 0.209$), quadratic ($F(2.501) = 81.451$; $p < 0.001$; $R^2 = 0.245$) and exponential ($F(1.502) = 281.978$; $p < 0.001$; $R^2 = 0.360$).The model that best fits the data is the exponential model, which has the highest R^2 value and explains more of the variability in the data than the other two models (36% of variance).

3.1. Which of the Sociodemographic or Pre-Linguistic Variable(s) Studied Explain Early Vocabulary Acquisition?

A summary of the data for the pre-linguistic variables is shown in Table 2. The significant findings from these data are that most children (88.6%) displayed the three behaviours corresponding to first signs of understanding, a large minority of the sample (40.9%) did not imitate words or parts of phrases, and 59.1% imitate them sometimes or often.

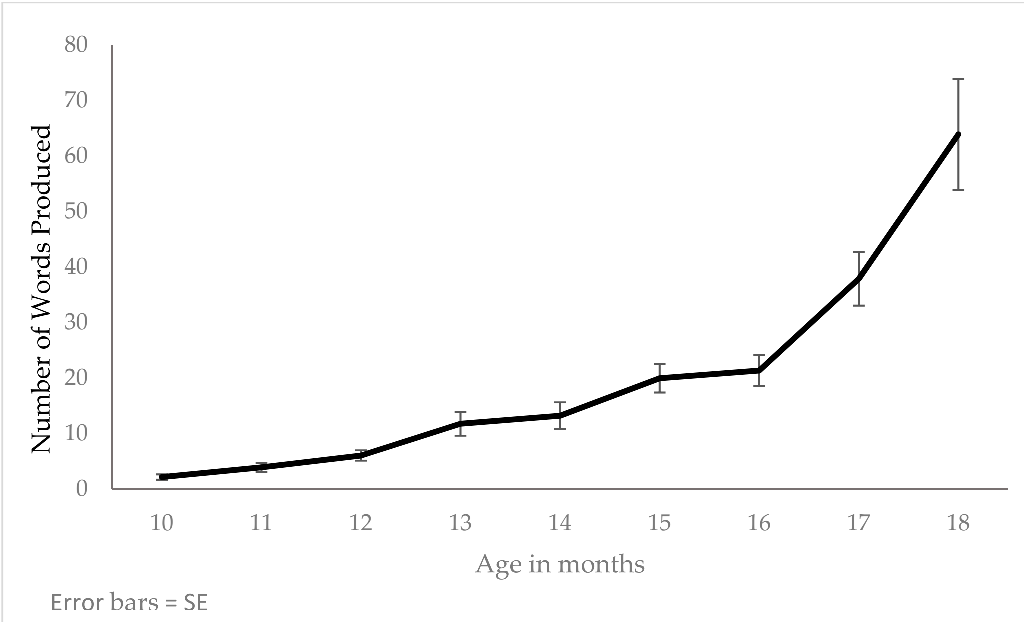


Figure 1. Production of words according to child's age in months.

Table 2. Descriptive data on language precursors.

Language Precursors	*N* = 504
First signs of understanding, %	
0 or 1 behaviour	0.8
2 behaviours	10.6
3 behaviours	88.6
Imitation, %	
Not yet	40.9
Sometimes	41.9
Often	17.2
Phrases, *M* (SD)	17.93 (7.09) [a]
First gestures, *M* (SD)	14.73 (4.12) [b]
Late gestures, *M* (SD)	17.02 (9.52) [c]

[a] Maximum score: 27, [b] Maximum score: 25, [c] Maximum score: 44.

For the second objective of the study, we used a stepwise regression analysis to identify which were the best predictors of vocabulary size. In this regression method, all variables, sociodemographic and pre-linguistic, were considered as predictors, and automatically in each step, the variable accounting for the most proportion of variance was introduced in the model, thus reducing the number of variables in the final model. The significance of the variables in the models, coefficient of determination (R^2) and standardized coefficients (β) were used to interpret the significant predictors, proportion of explained variance and the relative weights of each predictor variable, respectively.

Table 3 shows the data from the sixth model generated by the multivariate regression analysis, based on the data from the stepwise method. The regression coefficient is provided with confidence intervals, the standardized score (β) and the statistical significance. Specifically, in the multiple regression analysis, the six predictors explain 65.1% of the variance in the score for expressive vocabulary (F = 133.695; p < 0.001; adjusted R^2 = 0.651). All the steps that the regression analysis generated can be found in the Appendix A.

Only six of the predictors explained the acquisition of expressive vocabulary in the ages studied, namely, imitation, the understanding of phrases, late gestures and actions, age, birth order, and birth weight. Looking at the standardized scores (β), it can be observed that imitation had the greatest predictive power, followed by late gestures and actions and, to a lesser extent, the understanding of phrases, age, birth order, and birth weight.

Table 3. Multivariate regression analysis of expressive vocabulary with respect to potential pre-linguistic and sociodemographic predictors.

Predictors	Coefficient (95% CI)	β	p
Age	0.1 (0.05 to 0.15)	0.167	<0.001
Birth weight	0.19 (0.02 to 0.36)	0.063	0.031
Birth order	−0.15 (−0.28 to −0.02)	−0.081	0.021
Imitation	0.85 (0.73 to 0.98)	0.445	<0.001
Late gestures and actions	0.04 (0.03 to 0.05)	0.245	<0.001
Phrases	0.03 (0.01 to 0.04)	0.140	0.002

CI indicates confidence interval; R^2 = 0.656; Adj. R^2 = 0.651.

3.2. Which of the Sociodemographic or Pre-Linguistic Variable(s) Studied Discriminate Children with a High Level of Vocabulary from Those with a Low Level of Vocabulary?

The children were classified according to the dependent variable (expressive vocabulary) based on their normative scores [102]. The cut-off point was applied at the 25th percentile score. Children were grouped according to whether they showed a high level of vocabulary (≥25th percentile) or a low level of vocabulary (<25th percentile). Based on this grouping, a binary logistic regression analysis was then carried out (stepwise method).

A total of 352 children (69.8%) obtained vocabulary scores above the 25th percentile, while 152 children were below this percentile. From this sample, the logistic regression analysis included a total of 428 participants (76 missing values). Table 4 shows that out of the ten possible predictors, only two (imitation and late gestures) are associated with vocabulary scores above the 25th percentile at a significance level of 5%.

Table 4. Multivariate logistic regression analysis of large vocabulary.

Predictors	OR (95% CI)	p
First model		
Imitation	6.889 (4.455 to 10.651)	<0.001
Second model		
Imitation	5.348 (3.400 to 8.411)	<0.001
Late gestures and actions	1.067 (1.036 to 1.099)	<0.001

CI, confidence interval; *N* = 428 (76 missing values); R^2 Nagelkerke: 0.374; OR = Odds ratio.

According to Nagelkerke's determination coefficient, the model explains 37.4% of the variance in the dependent variable.

To assess whether the predictor variables enabled the discrimination between a high level of vocabulary and a low level of vocabulary, Table 5 shows the specificity and sensitivity of the model generated by logistic regression analysis. The classification table shows that the model has good specificity (83.3%) but low sensitivity (53.9 %). We can therefore interpret from the results that when the predictor variables are present to a greater degree, there is a high probability that children show a high level of vocabulary, with little concern of vocabulary difficulties occurring. However, an absence or limited presence of imitation and late gestures does not discriminate adequately between children with a low or high level of vocabulary, that is, we cannot distinguish whether or not a vocabulary delay would be present when a child shows low levels of imitation and gesture use.

Table 5. Specificity and sensitivity of the model.

	Observed Percentile		
Predicted percentile	<25	>25	
<25	69	50	
>25	59	250	
	Sensitivity 53.9%	Specificity 83.3%	Accuracy 74.5%

4. Discussion

The results of our study show that both sociodemographic and pre-linguistic variables affect the acquisition of the initial vocabulary, although the latter has more explanatory weight than the former, except for age. In terms of the personal and sociodemographic variables, only the age, birth order and birth weight, in descending order of importance, significantly predicted vocabulary development. With respect to pre-linguistic factors, imitation, late gestures and first sentences comprehension were significant predictors of initial vocabulary production. Thus, demographic factors in conjunction with pre-linguistic ones are useful in explaining the initial vocabulary acquisition with a high amount of variance explained (65.1%). In a detailed analysis, the results of multiple regression and logistic regression indicated that imitation had the greatest explanatory weight. Moreover, the presence of this behaviour can adequately discriminate children with high and low levels of vocabulary.

4.1. Initial Vocabulary

Age is a predictor of vocabulary level, but is not as good as it might seem, as its impact was third in the order of weighting predictors used in this study and was preceded by pre-linguistic variables. It was the best sociodemographic variable in terms of explanatory weight. The effect of age on vocabulary growth reflects a long-established fact regarding the initial course of vocabulary learning: The transition from slow to rapid word-learning in the first half of the child's second year [107,108]. Authors such as Bloom [107] and Nelson [108] observed a sudden increase in new word learning from the age of 17 months, which is the age at which in our data we observed an increase in the rate of vocabulary acquisition. More recently, Fenson et al. [12], using the English MacArthur–Bates CDIs, also observed a considerable gain in the 16–18 month period. Therefore, the results obtained in our study reflect the phenomenon known as the vocabulary or lexical spurt, according to which most children increase their vocabulary notably between these months, as shown by the fact that the function that best explains the rate of acquisition of vocabulary has an exponential nature. This abrupt increase in vocabulary learning can be cognitively interpreted as the acquisition of a new learning procedure. When a "critical mass" of vocabulary is reached (approximately after 50 items, independently of age), words go from being simple gestures or acoustic signals to progressively decontextualized signs. Then, new labels and later words, accepting morphological marks according to their category, are quickly incorporated [48,109,110]. Children are then said to have acquired a new learning strategy which opens the path towards full adult competence [111].

4.2. Sociodemographic and Pre-Linguistic Predictors of Early Vocabulary

Birth order has been associated with language development at earlier ages, with better grammatical and vocabulary skills in firstborns [13,14,30], a finding our results support. In our sample of 10–18-month-old children acquiring Catalan, firstborns show a higher vocabulary production than later-borns. The relation between vocabulary growth and birth order is probably mediated by the direct adult–child speech parents can establish, and is associated with the quantity and quality—more input received from siblings and less directly from parents—of the language received by the child [30,36,37]. However, other studies failed to find this relationship (see for example, [32]), or have found it only temporarily during development [31]. The results of Pine [31] indicated that this relationship could be stronger at the beginning of language acquisition [112] and as the child grows this relationship weakens. Previous studies have also found that second-born children show better communication skills [30] because communication depends largely on socialization, as highlighted by different authors [113]. Thus, as different authors have shown, birth order cannot be considered a risk factor of language delay [30,114]. Differences due to birth order in language development are a reflex of different language contexts where first-borns receive more direct-adult speech that improves grammar and lexical development, and later-borns receive a greater variety of conversations and communication opportunities that improve communication skills [30,114].

Birth weight was also a significant predictor of vocabulary size at 10–18 months in our study. However, its impact compared to the other variables was small, maybe because we only included healthy children with a weight over 1900 g. In fact, a low birth weight has been found to be related with medical complications, such as deafness and cerebral palsy, which could cause subsequent language and developmental difficulties [22], but these children were excluded from our study. Meanwhile, moderately low birth weight (between 1900 and 2500 g) in healthy children has shown divergent results across studies, some indicating that birth weight affects early vocabulary development and others not [24,115].

We failed to find a relationship between vocabulary score and the other sociodemographic factors (sex, otitis episodes, and mother's level of education) previously related with language development in our sample of young Catalan language learners. It may be that the relationship between these sociodemographic factors and language acquisition is mediated by other variables or that they have an influence during later language development.

Sex did not predict a higher vocabulary rate in 10 to 18-month-old Catalan children when pre-linguistic variables were included in the regression model. This result is contrary to other studies of English and non-English languages [13–16]. Our data are in better agreement with a recent large-scale study that found few differences between girls and boys in language development [9]. Some researchers in language development and language disorders have stated that during the last decades, especially in clinical contexts, there has been a diagnostic bias of language difficulties regarding boys, and a misdiagnosis of girls because the latter show less evident symptoms and go unnoticed more often [34,116,117].

With regard to temporal mild hearing loss, our results are consistent with those of a meta-analysis by Roberts et al. [28], which indicated that the number of otitis episodes is not related to the variance of vocabulary production. This is a variable that was considered to be a risk factor for spoken language difficulties in clinical settings [29], but our evidence does not support this.

With respect to the mother's education level, diverse studies have pointed out its influence in language development [33,38,43]. Nevertheless, this influence seems to be mediated by the linguistic input that the child receives [38] and the quality of parental communication (e.g., direct speech, routines . . .) [118]. Our data do not indicate that the mother's education level explains any noticeable variance in early vocabulary development. Although other studies also failed to find this relation in our context [118], it is possible that this variable has a greater impact in later development.

Among prelinguistic factors, our results show that the comprehension of frequent phrases is a significant predictor of vocabulary. This fact is relevant and give rise to some reflections on how the simple fragments and phrases that conform this section ("a dormir"/"go to sleep", "anem a banyar"/"let's take a shower", "fes-me un petó"/"give me a kiss", "obre la boca"/"open your mouth", "molt bé"/"very well", "Què és això?"/"what is that?", "Vols . . . ?"/"do you want...?", "T'has fet caca?"/"did you poop?") are facilitators and precursors of vocabulary learning. This typical language directed at children is redundant, highly contextualized, and with overlapping clues (gestural, visual, and contextual), features that help children to analyse and recognize words and intentions. Thus, children pay attention to language not only by observing the formal composition of parental productions, but also their function, meaning, and referentiality. They profit from gathering the initial understanding of the first orders or demands, the understanding of the pragmatic intentions in the speech acts, and the exchange of questions highly contextualized about actions. In that sense, our findings are aligned with the experimental studies of Cartmill et al. [119], who showed that the quality of parental input, in particular the opportunities they offer for understanding and producing words in a contextualized and informative medium at 14 and 18 months, is a good predictor of the vocabulary level at 3 years of age. Another recent study [120] confirmed that the parent coaching in 8 and 14-month-old infants correlates with vocabulary levels at 18 months.

Parental input is one of the best predictors of a child's later language performance and our study has shown that the understanding of these prototypical examples of child-directed speech is related to the course of learning words. This result agrees with those that have reported a strong correlation between the comprehension of sentences in the initial years and the later language level [45,63,64]. Future studies should analyze more carefully whether the understanding of these first and repetitive adult productions, together with the acquisition of the other pre-linguistic and linguistic factors reported here, is a necessary condition for progress in vocabulary learning.

Verbal imitation was the variable with the strongest predictive power of vocabulary growth in our study. Together with Tennie et al. [121], we consider imitation central to any explanation of our complex culture. Language according to many authors is the most powerful cultural artefact transmitted from one generation to another [122]. Nevertheless, the number of studies on verbal imitation are still scarce compared to the large body of research that exists on the imitation of instrumental actions. In this study, we tried to resume the interest in the study of the role of imitation in language acquisition, specifically in the production of vocabulary during the early years of life. Our study clearly shows that, among pre-linguistic factors, verbal imitation explains the highest percentage of the variance of expressive vocabulary, as was also found in previous studies [74–76,80]. Children who showed more verbal imitations of their mothers' productions at 13 months were those who at 17–21 months had more advanced vocabulary skills.

Similar studies carried out including children with atypical development also corroborate our results. Yoder and Layton [82] found that imitation ability positively predicted the size of the initial spoken vocabulary in children with ASD. In another recent study carried out with children with ASD [83], it was observed that verbal imitation at 20 and 71 months, as evaluated by the CDI inventories, is associated with a subsequent rapid growth of expressive vocabulary. In a study conducted including children with Down syndrome, it was found that poor verbal imitation may negatively influence the extent to which words enter the child's repertoire [81]. Therefore, imitation can be seen as a strategy that infants use for representing and encoding new verbal behaviors, and incorporating them into an existing repertoire involves reproducing and acquiring a new 'word' in its appropriate form and function [123,124]. Imitation also can be seen as a behavior for interacting with others—it can serve to acknowledge interactions with others, maintain the topic, or take turns. Although not all children use imitation, it has the potential to advance vocabulary acquisition by facilitating the processes of mental representation, analysis, and practice of linguistic structures. As a strategy, beyond the first half of the second year it may not be as effective [108].

From our results it can be concluded that verbal imitation can be used to distinguish between children who are going to have good linguistic development and those who are at risk of presenting some difficulties or delay in the development of language. While our results point to a significant effect of verbal imitation in language development, in child development assessments it will be important to explore it as a warning sign to detect atypical children or those who are at-risk in their early development. Furthermore, based on the results of our study, verbal imitation can be understood as an effective pre-linguistic training strategy for professionals working with young children. Thus, in children with early language development difficulties, verbal imitation could be a strategy to favour later language development.

In addition to imitation, our results have shown the importance of late gestures and actions as predictors for initial vocabulary. Beyond deictic gestures or emblems, it was observed that complex actions or gestures constitute the second strongest predictor of vocabulary acquisition. As described in Section 2, "later gestures" include actions of symbolic play or activities related to it. Therefore, the data obtained do not support the idea that communicative gestures or routines ("early gestures") as a whole are adequate predictors of vocabulary acquisition in the age range we have studied. As some authors have argued [93], it is possible that they are only behavioural antecedents that prepare the emergence of expressive language. By contrast, late gestures or actions related to symbolic play are good predictors of the level of vocabulary, in line with the findings of another study [98]. Language and symbolic play reflect the development of underlying mental representative functions. Several studies concerning the relationship between early language and symbolic play have established temporal correlations in functional and structural development [99,125–127]. In the present study, this relationship was also found. Vocabulary acquisition is related to symbolic play, so symbolic play can predict the rate and size of the expressive vocabulary. However, more research is needed to clarify this relationship at other ages and for other language components in order to explain the causal direction of these influences.

Although the present study presents several strengths, including the joint analysis of relevant factors in language development in a large sample, it also has some limitations. One of the most important limitation is the use of a cross-sectional analysis instead of a longitudinal analysis. Future studies would benefit from following children through their development to show how these variables influence the full developmental language process. Another aspect that must be further studied is the time that parents stay at home with their children, whether they work, and whether the child attends preschool centers. These data would allow us to gain a broader overview of the factors related to learning vocabulary.

Our results have several practical implications. First, they highlight that although sociodemographic variables play a role in identifying early language difficulties, they cannot be used alone to detect children at risk and must be combined with other prelinguistic risk factors. Second, our results provide further evidence that the presence of verbal imitations and symbolic play in the first two years of life are indicators of a positive prognostic of language development. In this sense, a greater number of risk factors seem to increase the probability of language delay [128,129]. We recommend that the progress of those children who show low levels of imitation and restricted symbolic play during their first and second years should be followed closely, especially if they were born with a low birth weight and are not the firstborn. It would also be recommended to promote imitation and symbolic play through language interventions delivered in naturalistic contexts (home or kindergarten preschool teachers or parents). In this sense, language interventions based on imitation and symbolic play have shown effective results in children with language difficulties [130–132]. Nevertheless, it is important to note that associated risk factors may differ depending on the age of the child and may change as children develop [128].

5. Conclusions

In this study we have shown that the relationship between pre-linguistic abilities and vocabulary competence is strong and that the former can predict either normal or delayed progress. Although the age, birth order and birth weight of the child are related to vocabulary size, these sociodemographic (non-linguistic) factors have low explanatory power and cannot be used in isolation as an early warning sign for vocabulary delays. The imitation of words or statements, participation in symbolic play activities, and the understanding of highly contextualized phrases are powerful predictors that identify the linguistic functions, the meanings and use of the first words of children and, thus, aid their learning. These results are highly relevant and helpful in early child communication and language development, either for the prevention of difficulties and, if necessary, for early interventions.

Author Contributions: Conceptualization, E.S.-S.; methodology, E.S.-S., E.A.-M., M.S.-T.; L.A., A.A. and M.S.; data collection, E.S.-S., E.A.-M., M.S.-T., L.A., A.A. and M.S.; data analysis, E.S.-S.; writing—original draft preparation, E.S.-S., E.A.-M., M.S.-T., L.A., A.A. and M.S.; writing—review and editing, E.S.-S., E.A.-M., M.S.-T., L.A., A.A. and M.S.; supervision, E.S.-S. and E.A.-M. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partially supported by the Spanish Ministerio de Economía y Competitividaad and the Fondo Europeo de Desarrollo Regional (PSI2015-69419-R; MINECO-FEDER).

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the Ethics Committee of University of Girona (CEBRU0024-2019).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Statistics from the stepwise regression model.

Model	Predictors	Coefficient (95% CI)	β	p	AdjR^2
1	Imitation	1.31 (1.18 to 1.45)	0.677	<0.001	0.456
2	Imitation	0.94 (0.82 to 1.07)	0.486	<0.001	0.622
	Late gestures and actions	0.07 (0.06 to 0.08)	0.450	<0.001	
3	Imitation	0.90 (0.78 to 1.02)	0.464	<0.001	0.637
	Late gestures and actions	0.05 (0.04 to 0.06)	0.327	<0.001	
	Age	0.11 (0.06 to 0.16)	0.183	<0.001	
4	Imitation	0.87 (0.74 to 0.99)	0.477	<0.001	0.645
	Late gestures and actions	0.04 (0.03 to 0,05)	0.266	<0.001	
	Age	0.08 (0.03 to 0.13)	0.142	0.001	
	Phrases	0.03 (0.01 to 0.05)	0.139	0.001	
5	Imitation	0.86 (0.74 to 0.99)	0.445	<0.001	0.648
	Late gestures and actions	0.04 (0.03 to 0.05)	0.270	<0.001	
	Age	0.09 (0.04 to 0.14)	0.154	0.001	
	Phrases	0.03 (0.01 to 0.04)	0.129	0.003	
	Birth order	−0.14 (−0.27 to −0.01)	−0.063	0.032	
6	Imitation	0.85 (0.73 to 0.98)	0.445	<0.001	0.651
	Late gestures and actions	0.04 (0.03 to 0.05)	0.245	<0.001	
	Age	0.1 (0.05 to 0.15)	0.167	<0.001	
	Phrases	0.03 (0.01 to 0.04)	0.140	0.002	
	Birth order	−0.15 (−0.28 to −0.02)	−0.081	0.021	
	Birth weight	0.19 (0.02 to 0.36)	0.063	0.031	

References

1. Bergelson, E.; Swingley, D. At 6-9 months, human infants know the meanings of many common nouns. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 3253–3258. [CrossRef]
2. Pruden, S.M.; Hirsh-Pasek, K.; Golinkoff, R.M.; Hennon, E.A. The Birth of Words: Ten-Month-Olds Learn Words Through Perceptual Salience. *Child Dev.* **2006**, *77*, 266–280. [CrossRef] [PubMed]
3. Hollich, G.; Hirsh-Pasek, K.; Golinkoff, R.M.I. What Does It Take to Learn a Word? *Monogr. Soc. Res. Child Dev.* **2000**, *65*, 1–16. [CrossRef]
4. Levine, D.; Strother-Garcia, K.; Golinkoff, R.M.; Hirsh-Pasek, K. Language Development in the First Year of Life. *Otol. Neurotol.* **2016**, *37*, e56–e62. [CrossRef]
5. Lederberg, A.R.; Schick, B.; Spencer, P.E. Language and literacy development of deaf and hard-of-hearing children: Successes and challenges. *Dev. Psychol.* **2013**, *49*, 15–30. [CrossRef]
6. Anastasi, A. *Differential Psychology*, 3th ed.; MacMiilan: New York, NY, USA, 1958.
7. Zambrana, I.M.; Pons, F.; Eadie, P.A.; Ystrom, E. Trajectories of language delay from age 3 to 5: Persistence, recovery and late onset. *Int. J. Lang. Commun. Disord.* **2014**, *49*, 304–316. [CrossRef]
8. Etchell, A.; Adhikari, A.; Weinberg, L.S.; Choo, A.L.; Garnett, E.O.; Chow, H.M.; Chang, S.E. A systematic literature review of sex differences in childhood language and brain development. *Neuropsychologia* **2018**, *114*, 19–31. [CrossRef] [PubMed]
9. Norbury, C.F.; Gooch, D.C.; Wray, C.; Baird, G.; Charman, T.; Simonoff, E.; Vamvakas, G.; Pickles, A. The impact of nonverbal ability on prevalence and clinical presentation of language disorder: Evidence from a population study. *J. Child Psychol. Psychiatry* **2016**, *57*, 1247–1257. [CrossRef]

10. Huttenlocher, J.; Haight, W.; Bryk, A.; Seltzer, M.; Lyons, T. Early vocabulary growth: Relation to language input and gender. *Dev. Psychol.* **1991**, *27*, 236–248. [CrossRef]
11. Galsworthy, M.J.; Dionne, G.; Dale, P.S.; Plomin, R. Sex differences in early verbal and non-verbal cognitive development. *Dev. Sci.* **2000**, *3*, 206–215. [CrossRef]
12. Fenson, L.; Marchman, V.A.; Thal, D.J.; Dale, P.S.; Reznick, J.S.; Bates, E. *The MacArthur Communicative Development Inventories: User's Guide and Technical Manual*, 2nd ed.; Paul H. Brooks: Baltimore, MD, USA, 2007.
13. Berglund, E.; Eriksson, M.; Westerlund, M. Communicative skills in relation to gender, birth order, childcare and socioeconomic status in 18-month-old children. *Scand. J. Psychol.* **2005**, *46*, 485–491. [CrossRef]
14. Fenson, L.; Dale, P.S.; Reznick, J.S.; Bates, E.; Thal, D.J.; Pethick, S.J.; Tomasello, M.; Mervis, C.B.; Stiles, J. Variability in Early Communicative Development. *Monogr. Soc. Res. Child Dev.* **1994**, *59*, i. [CrossRef]
15. Fenson, L.; Dale, P.S.; Reznick, J.S.; Donna, T.; Elisabeth, B.; Hartung, J.P.; Pethick, S.J.; Reilly, J.; Thal, D.J.; Bates, E.; et al. *The MacArthur Communicative Development Inventories: User's Guide and Technical Manual*; Singular Publishing Group: San Diego, CA, USA, 1993.
16. Eriksson, M.; Marschik, P.B.; Tulviste, T.; Almgren, M.; Pérez-Pereira, M.; Wehberg, S.; Marjanovič-Umek, L.; Gayraud, F.; Kovačević, M.; Gallego, C. Differences between girls and boys in emerging language skills: Evidence from 10 language communities. *Br. J. Dev. Psychol.* **2012**, *30*, 326–343. [CrossRef]
17. Kovačević, M.; Cepanec, M.; Kuvač Kraljević, J. Sex differences in lexical and grammatical development in Croatian. In *Proceedings from the First European Network Meeting on the Communicative Development Inventories*; Marten, E., Ed.; Gävle University Press: Gävle, Švedska, 2007; pp. 5–15.
18. Jackson-Maldonado, D.; Thal, D.J.; Marchman, V.; Newton, T.; Fenson, L.; Conboy, B. *McArthur Inventarios del Desarrollo de Habilidades Comunicativas. User's Guide and Technical Manual*; Brookes: Baltimore, MD, USA, 2003.
19. Berglund, E.; Eriksson, M. Communicative Development in Swedish Children 16-28 months old: The Swedish Early Communicative Development Inventory—Words and Sentences. *Scand. J. Psychol.* **2000**, *41*, 133–144. [CrossRef]
20. Bosch, L.; Agut Quijano, T.; Busquets Ferrer, L. Prematuridad y dificultades del lenguaje. In *Dificultades del Habla, el Lenguaje y la Comunicación en los Trastornos del Neuro-Desarrollo*; Aguilar-Mediavilla, E., Igualada, A., Eds.; UOC: Barcelona, Spain, 2018.
21. McKean, C.; Wraith, D.; Eadie, P.A.; Cook, F.; Mensah, F.; Reilly, S. Subgroups in language trajectories from 4 to 11 years: The nature and predictors of stable, improving and decreasing language trajectory groups. *J. Child Psychol. Psychiatry* **2017**, *10*, 1081–1091. [CrossRef]
22. Resches, M.; Pérez-Pereira, M.; Cruz Guerrero, R.; Fernández Prieto, M. Risk for language delay in healthy preterm and full-term children. In *Atypical Language Development in Romance Languages*; Aguilar-Mediavilla, E.M., Buil-Legaz, L., López-Penadés, R., Sánchez-Azanza, V.A., Adrover-Roig, D., Eds.; John Benjamins Publishing Company: Amsterdam, The Netherlands, 2019; pp. 57–74.
23. Adams-Chapman, I.; Bann, C.; Carter, S.L.; Stoll, B.J. Language outcomes among ELBW infants in early childhood. *Early Hum. Dev.* **2015**, *91*, 373–379. [CrossRef] [PubMed]
24. Pérez-Pereira, M.; Fernández, P.; Gómez-Taibo, M.L.; Resches, M. Language development of low risk preterm infants up to the age of 30months. *Early Hum. Dev.* **2014**, *90*, 649–656. [CrossRef] [PubMed]
25. Robertson, C.M.T.; Watt, M.-J.; Dinu, I.A. Outcomes for the Extremely Premature Infant: What Is New? and Where Are We Going? *Pediatr. Neurol.* **2009**, *40*, 189–196. [CrossRef]
26. Hooper, S.R.; Ashley, T.A.; Roberts, J.E.; Zeisel, S.A.; Poe, M.D. The relationship of otitis media in early childhood to attention dimensions during the early elementary school years. *J. Dev. Behav. Pediatr.* **2006**, *27*, 281–289. [CrossRef]
27. Roberts, J.E.; Wallace, I.; Henderson, D.E. LANGUAGE and otitis media. In *Otitis Media in Young Children: Medical, Developmental, and Educational Considerations*; Roberts, J.E., Wallace, I.F., Henderson, F.W., Eds.; Paul H Brookes Co.: Baltimore, MD, USA, 1997; pp. 133–161.
28. Roberts, J.E.; Rosenfeld, R.M.; Zeisel, S.A. Otitis media and speech and language: A meta-analysis of prospective studies. *Pediatrics* **2004**, *113*, e238–e248. [CrossRef]
29. Rosenfeld, R.M.; Shin, J.J.; Schwartz, S.R.; Coggins, R.; Gagnon, L.; Hackell, J.M.; Hoelting, D.; Hunter, L.; Kummer, A.W.; Payne, S.C.; et al. Clinical Practice Guideline: Otitis Media with Effusion (Update). *Otolaryngol. Neck Surg.* **2016**, *154*, S1–S41. [CrossRef]
30. Hoff-Ginsberg, E. The relation of birth order and socioeconomic status to children's language experience and language development. *Appl. Psycholinguist.* **1998**, *19*, 603–629. [CrossRef]
31. Pine, J.M. Variation in Vocabulary Development as a Function of Birth Order. *Child Dev.* **1995**, *66*, 272. [CrossRef]
32. Oshima-Takane, Y.; Goodz, E.; Derevensky, J.L. Birth Order Effects on Early Language Development: Do Secondborn Children Learn from Overheard Speech? *Child Dev.* **1996**, *67*, 621. [CrossRef]
33. Auza-Benavides, A.; Peñaloza, C.; Murata, C. The influence of maternal education on the linguistic abilities of monolingual Spanish-speaking children with and without Specific Language Impairment. *Atyp. Lang. Dev. Roman. Lang.* **2019**, 93–112. [CrossRef]
34. Conti-Ramsden, G.; Durkin, K. What factors influence language impairment? Considering resilience as well as risk. *Folia Phoniatr. Logop.* **2016**, *67*, 293–299. [CrossRef]

35. Gilkerson, J.; Richards, J.A.; Warren, S.F.; Montgomery, J.K.; Greenwood, C.R.; Kimbrough Oller, D.; Hansen, J.H.L.; Paul, T.D. Mapping the Early Language Environment Using All-Day Recordings and Automated Analysis. *Am. J. Speech-Lang. Pathol.* **2017**, *26*, 248–265. [CrossRef]
36. Hirsh-Pasek, K.; Adamson, L.B.; Bakeman, R.; Owen, M.T.; Golinkoff, R.M.; Pace, A.; Yust, P.K.S.; Suma, K. The Contribution of Early Communication Quality to Low-Income Children's Language Success. *Psychol. Sci.* **2015**, *26*, 1071–1083. [CrossRef]
37. Romeo, R.R.; Leonard, J.A.; Robinson, S.T.; West, M.R.; Mackey, A.P.; Rowe, M.L.; Gabrieli, J.D.E. Beyond the 30-Million-Word Gap: Children's Conversational Exposure Is Associated With Language-Related Brain Function. *Psychol. Sci.* **2018**. [CrossRef]
38. Sultana, N.; Wong, L.L.N.; Purdy, S.C. Natural Language Input: Maternal Education, Socioeconomic Deprivation, and Language Outcomes in Typically Developing Children. *Lang. Speech Heart Serv. Sch.* **2020**, *51*, 1049–1070. [CrossRef]
39. Sperry, D.E.; Sperry, L.L.; Miller, P.J. Language Does Matter: But There is More to Language Than Vocabulary and Directed Speech. *Child Dev.* **2019**, *90*, 993–997. [CrossRef]
40. Sperry, D.E.; Sperry, L.L.; Miller, P.J. Reexamining the verbal environments of children from different socioeconomic backgrounds. *Child Dev.* **2019**, *90*, 1303–1318. [CrossRef]
41. Golinkoff, R.M.; Hoff, E.; Rowe, M.L.; Tamis-LeMonda, C.S.; Hirsh-Pasek, K. Language Matters: Denying the Existence of the 30-Million-Word Gap Has Serious Consequences. *Child Dev.* **2019**, *90*, 985–992. [CrossRef] [PubMed]
42. Lowry, L. Closing the 30 million word gap: It's about quality, not just quantity. *Hanen Early Lang. Progr.* **2016**, 1–4.
43. Maria-Mengel, M.R.S.; Linhares, M.B.M. Risk factors for infant developmental problems. *Rev. Lat. Am. Enferm.* **2007**, *15*, 837–842. [CrossRef]
44. Dollaghan, C.; Campbell, T.F.; Paradise, J.L.; Feldman, H.M.; Janosky, J.E.; Pitcairn, D.N.; Kurs-Lasky, M. Maternal education and measures of early speech and language. *J. Speech Lang. Heart Res.* **1999**, *42*, 1432–1443. [CrossRef]
45. Feldman, H.M.; Dollaghan, C.A.; Campbell, T.F.; Kurs-Lasky, M.; Janosky, J.E.; Paradise, J.L. Measurement Properties of the MacArthur Communicative Development Inventories at Ages One and Two Years. *Child Dev.* **2000**, *71*, 310–322. [CrossRef]
46. Jackson-Maldonado, D.; Barcenas Acosta, R.P. Assessing early communicative abilities in Spanish-speaking children from low-SES families. *Zero Three* **2006**, *27*, 22–27.
47. Papoušek, M.A.; Papoušek, H.B. Early verbalizations as precursors of language development. In *Infant Development: Perspectives From German-Speaking Countries*; Taylor and Francis: New York, NY, USA, 2019; pp. 299–328.
48. Serra, M.; Serrat, E.; Solé Planas, R.; Bel Leal, A.; Aparici, M. *La Adquisición del Lenguaje*; Ariel: Barcelona, Spain, 2000.
49. DeCasper, A.; Fifer, W. Of human bonding: Newborns prefer their mothers' voices. *Science* **1980**, *208*, 1174–1176. [CrossRef]
50. Hepper, P.G.; Scott, D.; Shahidullah, S. Newborn and fetal response to maternal voice. *J. Reprod. Infant Psychol.* **1993**, *11*, 147–153. [CrossRef]
51. Mehler, J.; Christophe, A. Maturation and learning of language in the first year of life. In *The Cognitive Neurosciences: A Handbook for the Field*; Gazzaniga, M.S., Ed.; MIT Press: Cambridge, MA, USA, 1995; pp. 943–954.
52. Saffran, J.R.; Aslin, R.N.; Newport, E.L. Statistical Learning by 8-Month-Old Infants. *Science* **1996**, *274*, 1926–1928. [CrossRef]
53. Saffran, J.R.; Newport, E.L.; Aslin, R.N. Word Segmentation: The Role of Distributional Cues. *J. Mem. Lang.* **1996**, *35*, 606–621. [CrossRef]
54. Mampe, B.; Friederici, A.D.; Christophe, A.; Wermke, K. Newborns' Cry Melody Is Shaped by Their Native Language. *Curr. Biol.* **2009**, *19*, 1994–1997. [CrossRef]
55. Whalen, D.H.; Levitt, A.G.; Wang, Q. Intonational differences between the reduplicative babbling of French- and English-learning infants. *J. Child Lang.* **1991**, *18*, 501–516. [CrossRef]
56. Sanz-Torrent, M.; Andreu, L. La adquisición del léxico. In *La Adquisición del Lenguaje*; Pearson Educación: Madrid, Spain, 2018; pp. 90–114.
57. Dosman, C.F.; Andrews, D.; Goulden, K.J. Evidence-based milestone ages as a framework for developmental surveillance. *Paediatr. Child Health* **2012**, *17*, 561–568. [CrossRef]
58. Rossetti, L. *The Rossetti Infant-Toddler Language Scale: A Measure of Communication and Interzaction*; LinguiSystems: East Moline, IL, USA, 2006.
59. Sax, N.; Weston, E. *Language Development Milestones*; University of Alberta: Alberta, Canada, 2007.
60. Gertner, Y.; Fisher, C.; Eisengart, J. Learning Words and Rules. *Psychol. Sci.* **2006**, *17*, 684–691. [CrossRef] [PubMed]
61. Hirsh-Pasek, K.; Golinkoff, R.M. *The Origins of Grammar: Evidence from Early Language Comprehension*; MIT Press: Cambridge, MA, USA, 1996.
62. Bishop, D.V.M.; Snowling, M.J.; Thompson, P.A.; Greenhalgh, T. CATALISE consortium CATALISE: A multinational and multidisciplinary Delphi consensus study. 1. Identifying language impairments in children. *PLoS ONE* **2016**, *11*, 1–26. [CrossRef] [PubMed]
63. Fish, M.; Pinkerman, B. Language skills in low-SES rural Appalachian children: Normative development and individual differences, infancy to preschool. *J. Appl. Dev. Psychol.* **2003**, *23*, 539–565. [CrossRef]
64. Watt, N.; Wetherby, A.M.; Shumway, S. Prelinguistic predictors of language outcome at 3 years of age. *J. Speech Lang. Heart Res.* **2006**, *49*, 1224–1237. [CrossRef]
65. Wetherby, A.M.; Prizant, B.M. *Communication and Symbolic Behaviour Scales–Infant/Toddler Checklist*; Paul H. Brookes Publishing Co.: Baltimore, MD, USA, 2001.

66. Zwaigenbaum, L.; Bryson, S.; Rogers, T.; Roberts, W.; Brian, J.; Szatmari, P. Behavioral manifestations of autism in the first year of life. *Int. J. Dev. Neurosci.* **2005**, *23*, 143–152. [CrossRef] [PubMed]
67. Luyster, R.; Lopez, K.; Lord, C. Characterizing communicative development in children referred for autism spectrum disorders using the MacArthur-Bates Communicative Development Inventory (CDI). *J. Child Lang.* **2007**, *34*, 623–654. [CrossRef] [PubMed]
68. Charman, T.; Drew, A.; Baird, C.; Baird, G. Measuring early language development in preschool children with autism spectrum disorder using the MacArthur Communicative Development Inventory (Infant Form). *J. Child Lang.* **2003**, *30*, 213–236. [CrossRef]
69. Meltzoff, A.N. Imitation and other minds: The "Like Me" hypothesis. In *Perspectives on Imitation: From Neuroscience to Social Science*; Hurley, S., Chater, N., Eds.; MIT Press: Cambridge, MA, USA, 2005; Volume 2, pp. 55–77.
70. Saville-Troike, M. Private speech: Evidence for second language learning strategies during the 'silent' period. *J. Child Lang.* **1988**, *15*, 567–590. [CrossRef] [PubMed]
71. Snow, C.E. The uses of imitation. *J. Child Lang.* **1981**, *8*, 205–212. [CrossRef] [PubMed]
72. Ervin-Tripp, S. Imitation and structural change in children's language. In *New Directions in the Study of Child Language*; Lennenberg, E.H., Ed.; MIT Press: Cambridge, MA, USA, 1964.
73. Rood, L. Braine Children's imitations of syntactic constructions as a measure of linguistic competence. *J. Verbal Learn. Verbal Behav.* **1970**, *10*, 430–441. [CrossRef]
74. Hauser, M.D.; Chomsky, N.; Fitch, T. The Faculty of Language: What Is It, Who Has It, and How Did It Evolve? *Science* **2002**, *298*, 1569–1579. [CrossRef]
75. Ramer, A.L.H. Syntactic styles in emerging language. *J. Child Lang.* **1976**, *3*, 49–62. [CrossRef]
76. Rodgon, M.M.; Kurdek, L.A. Vocal and Gestural Imitation in 8-, 14-, and 20-Month-Old Children. *J. Genet. Psychol.* **1977**, *131*, 115–123. [CrossRef]
77. Tomasello, M. *Constructing a Language: A Usage-Based Theory of Language Acquisition*; Harvard University Press: Cambridge, MA, USA, 2003.
78. Bloom, L.; Hood, L.; Lightbown, P. Imitation in language development: If, when, and why. *Cogn. Psychol.* **1974**, *6*, 380–420. [CrossRef]
79. Moerk, E.L. *Pragmatic and Semantic Aspects of Early Language Development*; University Park Press: Baltimore, MD, USA; London, UK; Tokyo, Japan, 1977.
80. Masur, E.F.; Olson, J. Mothers' and infants' responses to their partners' spontaneous action and vocal/verbal imitation. *Infant Behav. Dev.* **2008**, *31*, 704–715. [CrossRef]
81. Feeley, K.; Jones, E. Strategies to address challenging behaviour in young children with Down syndrome. *Down Syndr. Res. Pract.* **2008**, *12*, 153–163. [CrossRef] [PubMed]
82. Yoder, P.J.; Layton, T.L. Speech following sign language training in autistic children with minimal verbal language. *J. Autism Dev. Disord.* **1988**, *18*, 217–229. [CrossRef]
83. Smith, V.; Mirenda, P.; Zaidman-Zait, A. Predictors of Expressive Vocabulary Growth in Children With Autism. *J. Speech Lang. Heart Res.* **2007**, *50*, 149–160. [CrossRef]
84. Over, H.; Gattis, M. Verbal imitation is based on intention understanding. *Cogn. Dev.* **2010**, *25*, 46–55. [CrossRef]
85. Bannard, C.; Klinger, J.; Tomasello, M. How selective are 3-year-olds in imitating novel linguistic material? *Dev. Psychol.* **2013**, *49*, 2344–2356. [CrossRef]
86. Branigan, H.P.; Pickering, M.J.; Cleland, A.A. Syntactic co-ordination in dialogue. *Cognition* **2000**, *75*, B13–B25. [CrossRef]
87. Garrod, S.; Anderson, A. Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition* **1987**, *27*, 181–218. [CrossRef]
88. Garrod, S.; Clark, A. The development of dialogue co-ordination skills in schoolchildren. *Lang. Cogn. Process.* **1993**, *8*, 101–126. [CrossRef]
89. Garrod, S.; Doherty, G. Conversation, co-ordination and convention: An empirical investigation of how groups establish linguistic conventions. *Cognition* **1994**, *53*, 181–215. [CrossRef]
90. Matthews, D.; Lieven, E.; Tomasello, M. What's in a manner of speaking? Children's sensitivity to partner-specific referential precedents. *Dev. Psychol.* **2010**, *46*, 749–760. [CrossRef] [PubMed]
91. Bates, E. *Language in Context*; Academic Press: New York, NY, USA, 1976.
92. Iverson, J.M.; Goldin-Meadow, S. Gesture paves the way for language development. *Psychol. Sci.* **2005**, *16*, 367–371. [CrossRef]
93. Gullberg, M.; de Bot, K.; Volterra, V. Gestures and some key issues in the study of language development. *Gesture* **2008**, *8*, 149–179. [CrossRef]
94. Esteve Gibert, N.; Igualada, A.; Prieto, P. El gest com a facilitador i precursor del desenvolupament del llenguatge. *Leng. Soc. Commun.* **2019**, *17*, 26–41. [CrossRef]
95. McGillion, M.; Herbert, J.S.; Pine, J.M.; Vihman, M.M.; DePaolis, R.; Keren-Portnoy, T.; Matthews, D. What Paves the Way to Conventional Language? The Predictive Value of Babble, Pointing, and Socioeconomic Status. *Child Dev.* **2017**, *88*, 156–166. [CrossRef]
96. Nelson, K. *Young Minds in Social Worlds: Experience, Meaning, and Memory*; Harvard University Press: Cambridge, MA, USA, 2007.
97. Cadime, I.; Silva, C.; Ribeiro, I.; Viana, F.L. Early lexical development: Do day care attendance and maternal education matter? *First Lang.* **2018**, *38*, 503–519. [CrossRef]

98. Lillard, A.S.; Lerner, M.D.; Hopkins, E.J.; Dore, R.A.; Smith, E.D.; Palmquist, C.M. The impact of pretend play on children's development: A review of the evidence. *Psychol. Bull.* **2013**, *139*, 1–34. [CrossRef] [PubMed]
99. Quinn, S.; Donnelly, S.; Kidd, E. The relationship between symbolic play and language acquisition: A meta-analytic review. *Dev. Rev.* **2018**, *49*, 121–135. [CrossRef]
100. Tomasello, M.; Mervis, C.B. The instrument is great, but measuring comprehension is still a problem. *Monogr. Soc. Res. Child Dev.* **1994**, *59*, 174–179. [CrossRef]
101. Serrat, E.; Sanz-Torrent, M.; Badia, I.; Aguilar-Mediavilla, E.; Olmo, R.; Lara Díaz, M.F.; Andreu, L.; Serra, M. The relationship between lexical learning and grammatical development. *Infanc. Aprendiz.* **2010**, *33*. [CrossRef]
102. Serrat-Sellabona, E.; Aguilar-Mediavilla, E.; Sanz-Torrent, M.; Andreu, L.; Amadó, A.; Badia, I.; Serra, M. *Inventaris del Desenvolupament D'habilitats Comunicatives MacArthur-Bates en Català. Guia d'ús i Manual Tècnic*; UOC: Barcelona, Spain, in press.
103. Can, D.D.; Ginsbrug-Block, M.; Golinkoff, R.M.; Hirsh-Pasek, K. A long-term predictive validity study: Can the CDI Short Form be used to predict language and early literacy skills four years later? *J. Child Lang.* **2013**, *40*, 821–835. [CrossRef]
104. Pérez-Pereira, M.; Resches, M. Concurrent and predictive validity of the Galician CDI. *J. Child Lang.* **2011**, *38*, 121–140. [CrossRef]
105. Marchman, V.A.; Martínez-Sussmann, C. Concurrent Validity of Caregiver/Parent Report Measures of Language for Children Who Are Learning Both English and Spanish. *J. Speech Lang. Heart Res.* **2002**, *45*, 983–997. [CrossRef]
106. Thordardottir, E.T.; Weismer, S.E.; Evans, J.L. Continuity in lexical and morphological development in Icelandic and English-speaking 2-year-olds. *First Lang.* **2002**, *22*, 3–28. [CrossRef]
107. Bloom, L. *One Word at a Time: The Use of Single-Word Utterances before Syntax*; Mouton: The Hague, The Netherlands, 1973.
108. Nelson, K. Structure and strategy in learning to talk. *Monogr. Soc. Res. Child Dev.* **1973**, *21*, 1–135. [CrossRef]
109. Dromi, E. *Esarly Lexical Development*; C.U.P.: Cambridge, MA, USA, 1987.
110. Serra, M.; Sanz-Torrent, M. Las 400 palabras: Datos para una perspectiva constructivista de la interfaz léxico-sintáctica. *Annu. Psicol./UB J. Psychol.* **2004**, *35*, 235–255.
111. Goldfield, B.A.; Reznick, J.S. Early lexical acquisition: Rate, content, and the vocabulary spurt. *J. Child Lang.* **1990**, *17*, 171–183. [CrossRef]
112. Bronfenbrenner, U. *The Ecology of Human Development: Experiments by Nature and Design*; Harvard University Press: Cambridge, MA, USA, 1979; ISBN 0-674-22457-4.
113. Duranti, A.; Ochs, E.; Schieffelin, B.B. *The Handbook of Language Socialization*; Wiley Blackwell: West Sussex, UK, 2014.
114. Tomblin, J.B. The effect of birth order on the occurrence of developmental language impairment. *Int. J. Lang. Commun. Disord.* **1990**, *25*, 77–84. [CrossRef] [PubMed]
115. Martínez-Nadal, S.; Bosch, L. Cognitive and learning outcomes in late preterm infants at school age: A systematic review. *Int. J. Environ. Res. Public Health* **2021**, *18*, 74. [CrossRef]
116. Gould, J.; Ashton-Smith, J. Missed diagnosis or misdiagnosis? Girls and women on the autism spectrum. *Good Autism Pract.* **2011**, *12*, 34–41.
117. Whitehouse, A.J.O. Is There a Sex Ratio Difference in the Familial Aggregation of Specific Language Impairment? A Meta-Analysis. *J. Speech Lang. Heart Res.* **2010**, *53*, 1015–1025. [CrossRef]
118. Valera-Pozo, M.; Adrover-Roig, D.; Pérez-Castelló, J.A.; Sanchez-Azanza, V.A.; Aguilar-Mediavilla, E. Behavioral, emotional and school adjustment in adolescents with and without developmental language disorder (DLD) is related to family involvement. *Int. J. Environ. Res. Public Health* **2020**, *17*, 1949. [CrossRef]
119. Cartmill, E.A.; Armstrong, B.F.; Gleitman, L.R.; Goldin-Meadow, S.; Medina, T.N.; Trueswell, J.C. Quality of early parent input predicts child vocabulary 3 years later. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 11278–11283. [CrossRef]
120. Ferjan Ramírez, N.; Lytle, S.R.; Kuhl, P.K. Parent coaching increases conversational turns and advances infant language development. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 3484–3491. [CrossRef] [PubMed]
121. Tennie, C.; Call, J.; Tomasello, M. Ratcheting up the ratchet: On the evolution of cumulative culture. *Philos. Trans. R. Soc. B Biol. Sci.* **2009**, *364*, 2405–2415. [CrossRef]
122. Klinger, J.; Mayor, J.; Bannard, C. Children's Faithfulness in Imitating Language Use Varies Cross-Culturally, Contingent on Prior Experience. *Child Dev.* **2016**, *87*, 820–833. [CrossRef]
123. Tomasello, M. The social bases of language acquisition. *Soc. Dev.* **1992**, *1*, 67–87. [CrossRef]
124. Masur, E.F. Infants' verbal imitation and their language development: Controversies, techniques and consequences. In *Methods for Studying Language Production*; Menn, L., Bernstein-Ratner, N., Eds.; LEA: Mahwah, NJ, USA, 2000; pp. 27–43.
125. Ogura, T. A longitudinal study of the relationship between early language development and play development. *J. Child Lang.* **1991**, *18*, 273–294. [CrossRef] [PubMed]
126. Lyytinen, P.; Poikkeus, A.-M.; Laakso, M.-L. Language and Symbolic Play in Toddlers. *Int. J. Behav. Dev.* **1997**, *21*, 289–302. [CrossRef]
127. McCune, L. A normative study of representational play in the transition to language. *Dev. Psychol.* **1995**, *31*, 198–206. [CrossRef]
128. Bishop, D.V.M.; Snowling, M.J.; Thompson, P.A.; Greenhalgh, T. CATALISE-2 consortium Phase 2 of CATALISE: A multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *J. Child Psychol. Psychiatry* **2017**, *58*, 1068–1080. [CrossRef]
129. McKean, C.; Reilly, S.; Bavin, E.L.; Bretherton, L.; Cini, E.; Conway, L.; Cook, F.; Eadie, P.A.; Prior, M.; Wake, M.; et al. Language outcomes at 7 years: Early predictors and co-occurring difficulties. *Pediatrics* **2017**, *139*, e20161684. [CrossRef] [PubMed]

130. Bauer, S.M.; Jones, E.A. Requesting and Verbal Imitation Intervention for Infants with Down syndrome: Generalization, Intelligibility, and Problem Solving. *J. Dev. Phys. Disabil.* **2015**, *27*, 37–66. [CrossRef]
131. Ingersoll, B.; Schreibman, L. Teaching reciprocal imitation skills to young children with autism using a naturalistic behavioral approach: Effects on language, pretend play, and joint attention. *J. Autism Dev. Disord.* **2006**, *36*, 487–505. [CrossRef] [PubMed]
132. Sussman, F.; Weitzman, E. *Take out the Toys: Building Early Toy Play for Children with Autism Spectrum Disorder and Other Social Communication Challenges*; The Hanen Centre: Toronto, ON, Canada, 2014.

children

MDPI

Review

Sentence Repetition Tasks to Detect and Prevent Language Difficulties: A Scoping Review

Irene Rujas [1,*], Sonia Mariscal [2], Eva Murillo [3] and Miguel Lázaro [4]

1 Departament of Psychology, Faculty of Psychology, CES Cardenal Cisneros (UCM), 28006 Madrid, Spain
2 Department of Developmental and Educational Psychology, Faculty of Psychology, Universidad Nacional de Educación a Distancia, 28040 Madrid, Spain; smariscal@psi.uned.es
3 Department of Basic Psychology, Faculty of Psychology, Universidad Autónoma de Madrid, 28049 Madrid, Spain; eva.murillo@uam.es
4 Department of Experimental Psychology, Cognitive Processes and Logopedics, Faculty of Psychology, Universidad Complutense de Madrid, 28223 Madrid, Spain; mlazar01@ucm.es
* Correspondence: irenerujas@universidadcisneros.es

Abstract: Sentence repetition tasks (SRTs) have been widely used in language development research for decades. In recent years, there has been increasing interest in studying performance in SRTs as a clinical marker for language impairment. What are the characteristics of SRTs? For what purposes have SRTs been used? To what extent have they been used with young children, in different languages, and with different clinical populations? In order to answer these and other questions, we conducted a scoping review. Peer reviewed studies published in indexed scientific journals (2010–2021) were analyzed. A search in different databases yielded 258 studies. Research published in languages other than English or Spanish, adult samples, dissertations, case studies, artificial models, and theoretical publications were excluded. After this exclusion, 203 studies were analyzed. Our results show that most research using SRT were conducted with English monolingual speakers older than 5 years of age; studies with bilingual participants have mostly been published since 2016; and SRTs have been used with several non-typical populations. Research suggests that they are a reliable tool for identifying language difficulties and are specifically suitable for detecting developmental language disorder.

Keywords: early detection; sentence repetition task; sentence imitation task; early language assessment; specific language impairment; developmental language disorder

Citation: Rujas, I.; Mariscal, S.; Murillo, E.; Lázaro, M. Sentence Repetition Tasks to Detect and Prevent Language Difficulties: A Scoping Review. *Children* **2021**, *8*, 578. https://doi.org/10.3390/children8070578

Academic Editors: Eva Aguilar Mediavilla, Miguel Pérez Pereira, Elisabet Serrat-Sellabona and Daniel Adrover-Roig

Received: 31 May 2021
Accepted: 29 June 2021
Published: 5 July 2021

1. Introduction

Sentence repetition tasks (SRTs) have been widely used in language development research for decades [1,2], with sentence repetition being part of the language batteries commonly used by clinicians to evaluate children's language skills [3]. Despite its apparent simplicity, repeating a sentence is more than a simple memorization task. To accurately recall a sentence, participants must parse the sentence, analyze the thematic relations (i.e., order of events), interpret the underlying syntactic representation, elaborate an articulation plan, and finally, produce it [4]. It has also been argued that sentence repetition converges with comprehension data and with data from spontaneous production. In this vein, it was found that both quantitative (mean length of utterances) and qualitative measures of children's spontaneous productions correlate with measures obtained from a repetition task in Italian [5]. Given these properties, in recent years, there has been increasing interest in studying SRT to detect and prevent language difficulties [6]. Carrying out these tasks with very young children increases the possibilities of early detection and prevention. Even though language disorders are usually diagnosed after 4 years of age, it has already been stated that there are early signs in communicative and language development that predict further difficulties [7].

In a seminal paper [8], the authors conducted a study with 11-year-old children with and without a history of developmental language impairment. The results showed that the sentence repetition task in English yielded high levels of sensitivity and specificity for diagnosing specific language impairment (SLI) in monolingual children. We are aware that the new terminology, developmental language disorder (DLD), has become increasingly accepted since 2017. However, as the term SLI has been more widely used in the studies reviewed from 2010, we have maintained it through this article. In fact, the authors observed that sentence repetition performance in recalling sentences from the Clinical Evaluation of Language Fundamentals-Revised (CELF-R) [9] was the most accurate of four clinical marker candidates of SLI; the other three markers were non-word repetition, past tense, and third person singular use. However, in a meta-analysis [10], it was concluded that although the evidence for sentence repetition as an identifier of SLI was positive, it was still inconclusive. For that research, 13 studies, undertaken with English speaking participants, were compared regarding the use of three markers for language impairment: verb tense use, non-word repetition, and sentence repetition. Sentence repetition either outperformed all other tasks or was equivalent to all other tasks in terms of identification accuracy in each of the studies that included more than one identification task. The author emphasized the need to design and carry out more studies to confirm the effects and refine the stimuli. More recently, in a community-based study [11], it was concluded that, together with an index of past tense marking in English, a sentence repetition task was sufficiently reliable for use as a language screener.

SRTs can be used not only to explore their relationship to language development in monolingual children and contribute to derive possible diagnoses, but also to explore the abilities of bilingual and multilingual children. It is important to bear in mind that, despite the fact that the majority of the world's population is bi/multilingual, most of the phenomena related to language development have initially been carried out with monolingual samples. In this regard, it has also been suggested that SRTs have a potential advantage in second language (L2) assessment, as it has been shown that performance on this kind of task is less influenced than any other tests (e.g., standardized tests) by length of exposure to L2 and experience, which are known to be limited in L2 [12].

Regarding the construction of SRT in different languages, it is important to remember that languages can drastically differ from each other in a number of linguistic features, and this can have an impact on performance in these tasks. For this reason, cross-linguistic assessment using SRTs seems particularly relevant both in monolingual, bilingual, and multilingual participants. These kinds of tasks should be developed in different languages, according to their linguistic characteristics and particularities, and can be used afterwards with different populations. A number of studies with monolingual children who speak languages other than English, and with bi/multilingual participants with typical and non-typical language development have been carried out in recent years (see [13] for Hebrew-Russian; [14] for (European) Spanish; [15] for Welsh-English; [16] for Catalan; [17] for Arabic-German; [18] for Hungarian; [19] for Vietnamese; [20] for Czech; [21] for Cantonese; [22] for (Latino) Spanish-English; and [23] for Danish). However, drawing on the review in [13], until relatively recently, little work had focused on diagnostic accuracy of repetition tasks (SRTs and non-word repetition) in bilinguals with SLI that speak languages other than English. This situation has been compensated for in the last few years as SRTs have been developed for a European project on bilingual children with SLI within the context of a multilingual project (COST Action IS0804 "Language Impairment in a Multilingual Setting: Linguistic Pattern and the Road to Assessment" (LITMUS), http://www.bi-sli.org, accessed on 1 July 2021) [24]. Within this project, linguistically motivated sentence repetition tasks were developed for identifying bilingual children with SLI aged 5 to 8. These studies have revealed clear differences between children with SLI and typically developing (TD) children in several languages [25,26].

Criteria to Construct, Present, and Score Sentence Repetition Tasks

In relation to the previously mentioned issue, i.e., the significant differences between languages (for instance, English is a relatively simple language concerning inflection, while

Finnish is very complex), SRTs vary in the way they are constructed and may, consequently, differ in the linguistic and cognitive abilities they measure. The construction of the SRTs will differ according to the participants to be tested. Logically, depending on the age of the children to be assessed, the type and length of the sentences must be different, leading to another difficulty when comparing results across studies. For example, for school age children, [24] recommended that all sentence repetition tasks include structures that are difficult for children with SLI across languages, including wh-questions and relative clauses. Another important factor to be considered when developing a sentence repetition task is the way it is presented. In this regard, different possibilities exist, that is, the task can be presented orally or be pre-recorded and using different presentation formats (with or without visual images). Some authors suggest that although recording the items adds homogeneity, it disrupts communication between the child and the person conducting the test, while a live voice helps engage children in the task [27]. Together, all these factors add a wide range of variability in the ways the list of sentences to be repeated can be constructed and presented.

Finally, different scoring systems can be used, from the simplest system, where the whole sentence must be repeated correctly in order to receive credit (binary scoring or 0/1), to more detailed approaches that index the number (scaled-score system) or even the types of errors per sentence. When using sentence repetition for clinical purposes, however, trade-offs may arise in using more detailed scoring systems, as simple scoring ones are faster and possibly more reliable to implement.

Summarizing, the evidence concerning SRT is complex to address because not only were typically developing children and children with SLI tested, but also a number of other populations (e.g., attention deficit hyperactivity disorder in [28]; consistent speech/phonological disorder in [29]; resolved late talkers in [30]; and children with autism spectrum disorder and SLI in [31]). Additionally, SRTs have been used to explore monolingual, bilingual, and multilingual children with different languages as their L1 and L2. Furthermore, the tasks constructed followed different criteria depending on the age of the participants, their experimental vs. clinical use, the specific interest in particular characteristics of a given language, etc. Considering these variations, different aspects of language and cognitive processing can be at play depending on the SRT developed or used. Moreover, given this broad heterogeneity in research, drawing clear conclusions regarding the use of sentence repetition tasks is not as simple as expected. For this reason, we conducted a scoping review with the aim of shedding light on the following research questions (RQ).

Regarding languages:

RQ1: To what extent have SRTs been used in different languages?

Regarding the populations studied:

RQ2: Are these populations monolingual, bilingual, or multilingual?

RQ3: What populations have been studied using SRT?

RQ4: Can SRTs be administered to very young children (e.g., under four years of age)?

Regarding the task:

RQ4: What kinds of SRTs have been used?

Regarding the aim:

RQ5: For what purposes have SRTs been used?

2. Materials and Methods

We followed the guidelines of the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) statement [32] for conducting this scoping review.

2.1. Identification of Studies and Inclusion Criteria

The process of identifying studies for this scoping review is summarized in the PRISMA flow diagram in Figure 1.

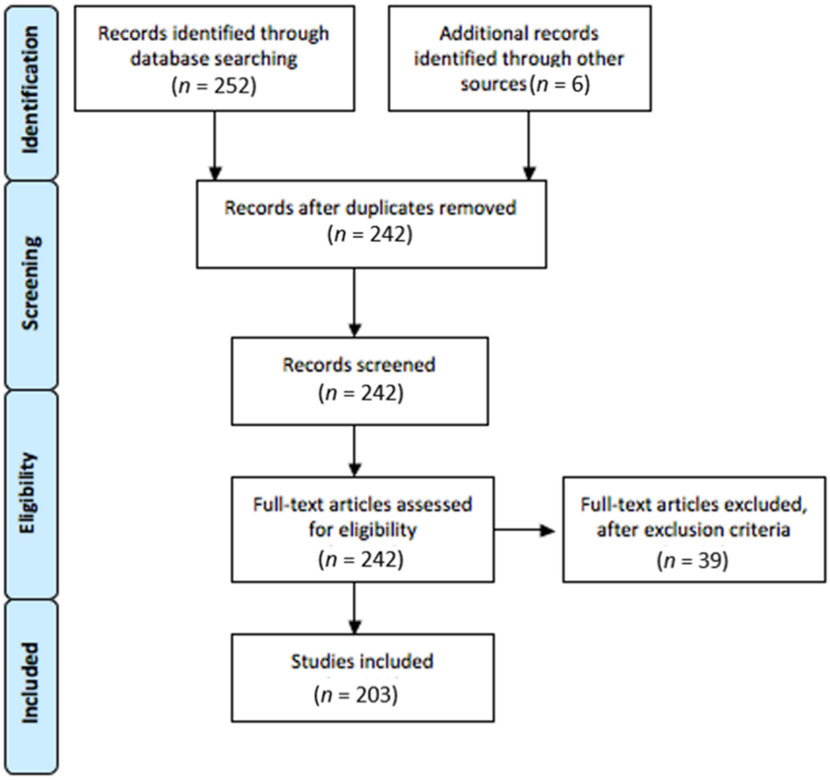


Figure 1. PRISMA Flow Diagram.

The search was conducted in March 2021 using 11 electronic databases: Academic Search Premier; APA PsycArticles; APA PsycBooks; APA PsycInfo CINAHL Complete; EBESCO eClassics Collection (EBESCOhost); Education Source ERIC; Medline; PSICODOC; and Psychology and Behavioral Sciences.

In addition, Google Scholar was used to complete the search for studies published in 2021.

The search was limited to peer reviewed studies published in indexed scientific journals between 2010 and 2021, in English or Spanish. Search terms included "sentence repetition task", OR "sentence imitation task", OR "sentence recall". Age was limited to participants under 18.

2.2. Exclusion Criteria

Studies that did not meet the inclusion criteria were excluded, that is:

Theoretical studies, meta-analysis, computational modeling, case studies, dissertations and conference proceedings.

Studies published in languages other than English or Spanish.

Studies that only included adult samples (ones that included both participants under 18 and adult samples were considered).

2.3. Data Analysis

The initial search led to 258 studies. After removing the 16 duplicates, 242 were screened. Thirty-nine were excluded because they did not meet the inclusion criteria. Finally, 203 full text articles were considered for the analysis. Appendix A Table A1 lists the empirical studies included in the scoping review, and the full database can be found in the (Supplementary Material Table S1). For each study, we obtained the following information:

Authors and year of publication.

Journal.

Number of languages spoken by the participants: monolingual, bilingual, both, other.

Language studied.

Populations studied: typically developing children vs. non-typically developing children.

Sample size.

Age range of the total sample included in the study.

Type of repetition task: belonging to an assessment battery, not original (taken from a previously published task), adapted (modified from a previously published task), original task (specifically developed for the particular study).

Number of sentences included in the task.

Aim of the task: we analyzed whether the SRT was used as a tool for language assessment, as a tool for cognitive assessment, as a clinical marker for language or development difficulties, or for other purposes.

3. Results

3.1. RQ1: To What Extent Have SRTs Been Used in Different Languages?

We found 33 different languages in the studies analyzed. Half of the studies included English speaking samples (103/203) and 11% (23/203) included Spanish speaking participants. The rest of the languages appeared in less than 10% of the studies. Table 1 shows the frequency of the languages included in the studies (the total is higher than 203 because several studies include more than one language).

Table 1. Frequency of languages included in the studies published between 2010 and 2021.

Language	*N*
English	104
Spanish	23
French	15
Italian	11
German	10
Hebrew	9
Hungarian	8
Russian	7
Arabic	6
Catalan	5
Finnish	4
Greek	4
Norwegian	4
Swedish	4
Czech	3
Danish	3
Dutch	3
Cypriot Greek	2
Kannada	2
Mandarin	2
Persian	2
Polish	2
Portuguese	2
Turkish	2
Welsh	2
Albanian-Greek	1
Cherokee	1
Farsi	1
Indian	1
Romanian	1
Vietnamese	1
British Sign Language	1
Other	1
Language Not Specified	2

The percentage of studies including English remains around 50% across years. Considering the tendency towards an increase in bilingual studies, this suggests they include mostly English speaking samples as a monolingual comparison group.

3.2. RQ2: Are These Populations Monolingual, Bilingual, or Multilingual?

Most of the studies carried out between 2010 and 2021 included only monolingual samples (74%; 149/203). A total of 22% (45/203) included bilingual samples (13% only bilingual participants and 9% bilingual and monolingual groups). Of the studies, 1% (3/203) included other populations, mainly L2 learners, while another 2% (4/203) provided no information on the number of languages spoken by the participants.

As can be seen in Figure 2, despite the predominance of monolingual studies, the last decade has witnessed an increase in the inclusion of bilingual samples in the studies.

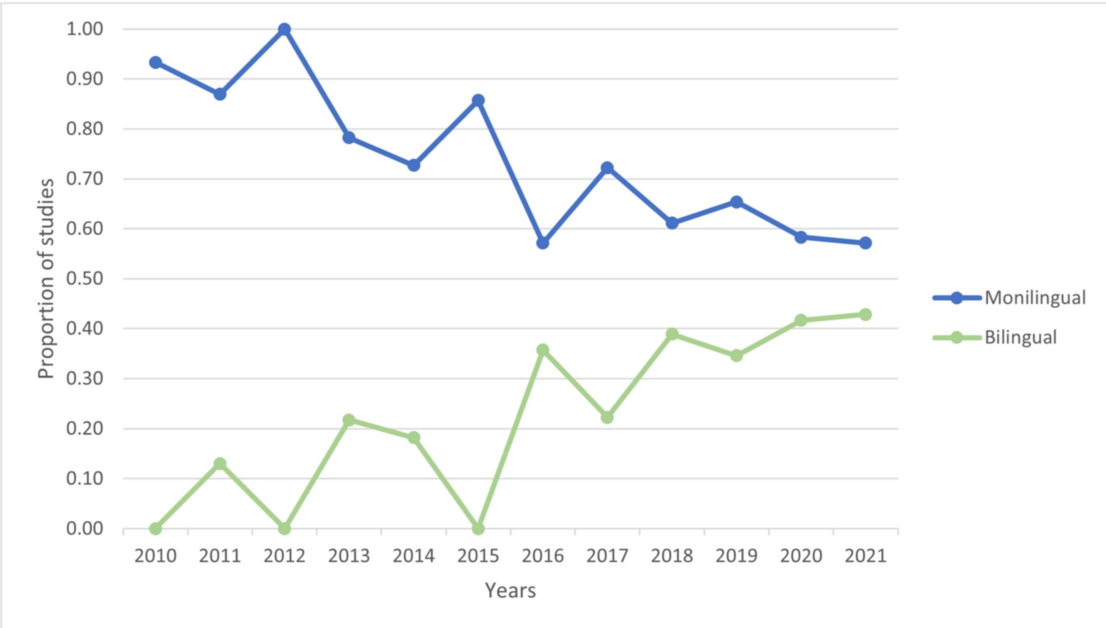


Figure 2. Proportion of monolingual and bilingual studies published between 2010 and 2021.

3.3. RQ3: What Populations Have Been Studied Using SRT?

Most of the studies (68%; 139/203) were carried out with children with non-typical development (NTD) with or without a TD control group, compared to 32% (64/203) of studies including only typically developing samples. For this study, we have used the term "non-typical" to cover both children with developmental disorders and children at risk of developmental difficulties (due to biological or social variables).

Of the studies, 139 included children with special characteristics. Table 2 shows the most frequent ones. As can be seen, more than half of these focused on language impairment (understanding this as a broad term covering language delay, specific language disorder, or developmental language disorder).

Table 2. Sample characteristics studied between 2010 and 2021.

Sample Characteristics	*N*
LI or SLI or DLD or language delay	76
Deafness, hearing difficulties, or hearing loss and/or cochlear implant	14
Autism spectrum disorder	14
Children at risk for language or learning difficulties	7
Reading difficulties or dyslexia	6
Cleft palate	4
Genetic syndrome	5
Cerebral palsy or brain damage	4
ADHD	5
Speech sound disorder	5
Stuttering	2
Learning disabilities	2
Anorexia	1
Auditory processing disorders (APD)	1
Developmental coordination disorder	1
HIV-infected and HIV-exposed	1
Infantile Thiamine deficiency	1
Oncological patients	1
Pediatric bipolar disorder	1
Phonological processing deficit (PPD)	1
Preterm (very low birth weight)	1
Adopted kids	1

Note: Some studies included groups with different conditions in their samples, and thus, the total from the table is higher than 139.

Sample size varies from 5 participants to 2212. However, almost 80% of the studies included samples of fewer than 150 participants, and 64% included 100 participants or less. Figure 3 shows the number of studies according to sample size.

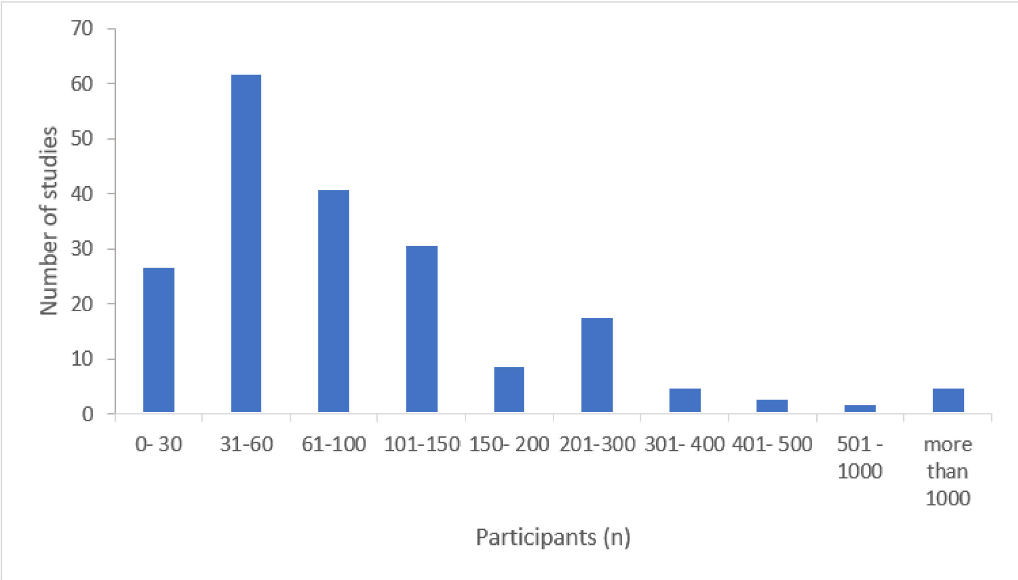


Figure 3. Number of studies according to the sample size.

3.4. RQ4: Can SRTs Be Administered to Very Young Children (e.g., under Four Years of Age)?

The age of the participants varied from 1, 10, to 25 years of age. Half of the studies (51.72%) had initial ages of between 5 and 8 years, and only 10% had initial ages of 9 years or above. Considering participants younger than 5 years, we found that 37.9% of the studies had participants aged 4 years or below. Only 15% of the studies included participants of below 4 years of age.

Regarding the final age, that is, the age of the oldest participant or at the final point of longitudinal studies, we also found that half of the studies had final ages between 5 and 8 years. Only 25% had final ages between 9 and 12 years.

As can be seen in Figure 4, we found the highest density of studies corresponded to age ranges from 5 years to around 6 years.

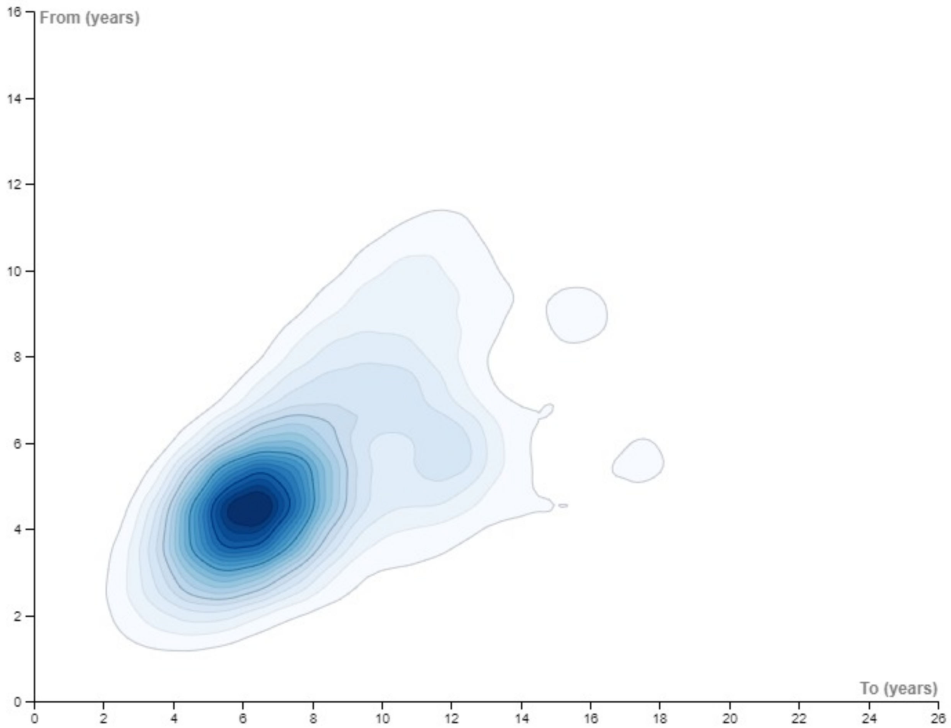


Figure 4. Distribution of studies according to the age range of the samples included.

3.5. RQ4: What Kinds of SRTs Have Been Used?

Most of the studies (41%; 83/203) opted to use SRTs that had been developed and published before, that is, "not original" tests, or they partially modified previously used tests, thus being "adapted" tests. A total of 33% (68/203) of the studies used SRTs belonging to a language/cognitive assessment battery (for example, the CELF, the NEPSY, or the TOLD). Only 25% (50/203) of the studies developed original sentence repetition tests that were specifically created for the research. Two papers (1%) did not specify the SRT used in the study.

Regarding the number of sentences included in the different SRTs, the shortest one was comprised of 10, while the longest one was comprised of 180. Not all studies that used standardized tests or "not original" tasks provided information on the number of sentences included. If we only considered the SRTs that were specifically created (original) for the research, most of the studies designed 20 sentence tasks (Mode = 20), while the mean number of sentences included was 38.

3.6. RQ5: For What Purposes Have SRTs Been Used?

Most of the research (62%, 125/203) based on SRTs used them as a tool to assess different language abilities; 14% (28/203) used the SRT to measure cognitive abilities (for example, short-term verbal memory); and 12% (25/203) employed the SRT for other purposes (for example, to study the psychometric properties of a particular SRT or to study specific linguistic units). It is worth mentioning that 18% (36/203) of the research leveraged SRTs as a tool to identify language difficulties (clinical marker).

3.7. Sentence Repetition Tasks as a Clinical Marker for Language Impairment

As shown before, 36 studies were specifically undertaken to analyze the potential of sentence repetition performance to identify children with language impairment. Most of these studies (69.44%; 25/36) were designed to assess the value of SRTs used as a clinical marker for SLI; 22.22% (8/36) aimed to evaluate the potential of the SRT to identify children with language impairment, language delay, or low language abilities; 5.55% (2/36) used SRT as a clinical marker for language impairment in children with reading difficulties or dyslexia; and only two studies used SRT as a clinical marker for ASD.

Regarding the ages of the participants included in this group of studies, mean range was 5; 4 to 8; 6 (years; months) years of age. Most of the studies were conducted with participants over 4 years of age and only two studies included children below this age. Compared to the complete set of studies reviewed, a slightly lower percentage of papers using SRT as a clinical marker included monolingual participants (66.66% compared to 74% of the whole set). Studies including bilingual populations and children with language impairment (SLI above all) have increased since 2016.

Regarding the type of task, a third of the studies (12/36) used an original format, followed by an adapted task (10/36; 7 being adaptations from the original LITMUS task to other languages), standardized batteries (9/36), and a non-original task (5/36).

Finally, in relation to the number of sentences included in the tasks, the observed range varied from 19 to 70 items (mean = 37, 26), with around 20 items being the most frequent length of SRT (mode = 20).

4. Discussion

The scoping review carried out with the terms “sentence repetition task”, “sentence imitation task”, and “sentence recall” revealed more than two hundred studies in the last ten years. This first result is indeed significant as it showed that SRTs have been a topic of great interest in the last decade. This is not surprising given that it is a simple task to administer, with several advantages over other language assessment tasks. For instance, as [29] highlights, SRTs enable a good number of carefully selected targets to be elicited in a more systematic way than is possible with spontaneous production. Moreover, the sentences to be repeated can include different lexical or morphosyntactic targets that are difficult to elicit with other materials or through spontaneous production. However, beyond these practical aspects, language evaluation through SRTs must be supported by experimental and empirical evidence confirming its appropriateness, for instance, for the clinical diagnosis of language impairment. In this review, we analyze evidence from the last ten years in order to offer a clear picture of the state of the field in relation to the use of different kinds of sentence repetition tasks in developmental research.

A first step towards this goal is to consider the languages under study. For this purpose, it is crucial to bear in mind that the search was limited to studies published in English (199/203) or Spanish (4/203). Considering this set of papers, data seem to be clear, with English being the most widely explored language. Spanish, being the second language in our results, was studied in more than half of the cases with Latino bilingual children (e.g., [33–36]), and research with monolingual Spanish speaking children is scarce (e.g., [14,37–39]). The representation of other languages is low, with most of them having only one or two published studies (e.g., in Arabic [40,41]; in Czech [20,42]; and in Kannada [43]). This set of results clearly shows that the evidence regarding the use of

SRTs is biased towards English. The fact that most of the empirical evidence is related to English is important and must be considered a significant issue because differences between languages can be enormous and, therefore, in a task such as an SRT, the results for a given language do not necessarily apply to other ones. Additionally, English is a particularly simple language in terms of morphosyntax, so if we consider morphosyntactic complexity as a continuum, English can be situated at a great distance from other languages such as Finnish or Polish, which have very complex morphosyntactic structures. This is a critical trait when considering a task devoted to exploring language development, such as SRTs. Therefore, a key conclusion of this paper is that results biased towards a single language cannot represent the outcomes for other languages. This statement also holds for the studies with bilingual and multilingual participants. In these cases, English tends to be one of the participants' languages. In fact, only 20% of the studies that included bilingual participants compared languages other than English (e.g., Arabic-German [17]; Russian-German [44]; and Spanish-Catalan [45]). In our view, it is critical that new studies with bilingual and multilingual children incorporate participants with different languages. It is true, however, that the number of studies conducted both with monolingual and bilingual children in languages other than English are slowly increasing. This, to our understanding, is of great importance, as it allows researchers to have a clearer and deeper view of the task and its value in assessing language development.

Regarding the participants involved in the research, we observed that most of the studies included children with different developmental conditions; in fact, only 32% of the studies were conducted with only TD children. Of the remaining studies (68%), most included children with SLI, but this was not the only group of non-typically developing children considered. As shown in Table 2, a considerable number of studies assess cases of children with other conditions, from cerebral palsy [46] to ASD (e.g., [47–50]). These data reflect that SRTs are not only suitable for assessing the development of children with SLI, but can also provide important information for researchers and clinicians interested in the language development of children with a number of other difficulties. Thus, these kinds of tasks have been used, for example, to explore the severity of a case of stuttering [51] or as a marker of language skills in children with dyslexia [52].

This diversity of conditions of the participants in the reviewed studies might explain, but only partially, the variability in the sample size. It is difficult to find large samples of children with rare genetic conditions, for example, but this does not hold for TD children or even for children with language disorders. The range of participants varies from 5 [53] to 2212 [54], but even if we remove these two studies, the differences are still immense (see Figure 4). Nevertheless, the important issue now is not merely the differences between studies, but the fact that the majority of them included fewer than 100 participants, with studies with groups of between 30 and 60 participants being the most numerous. It should be noted that the computation of the sample sizes reported includes all participants per study as a whole, i.e., a study on children with SLI with 60 participants will probably include only 30 children with this impairment and 30 TD children as control. This implies that despite the total number of children taking part in all these studies being large, the research is usually underpowered, and only a relatively small proportion of the papers report results from large samples. This is even more important, considering the small number of studies concerning languages other than English or specific conditions such as cerebral palsy, etc. In summary, although this review shows that SRTs have been used in research with a wide range of languages and developmental conditions, a deeper analysis indicates there is still plenty of room for more studies to be conducted.

Regarding the age of participants, most of the studies focused on children between five and eight years old. This is for two reasons. The first is that clinicians and researchers are able to engage these children in the task and obtain valid data more easily than with younger children. The second is that with a careful construction of the set of sentences to be repeated, it is possible to focus on morphosyntactic or lexical aspects of language development that are unreasonable when children are younger or older. However, our results also clearly show that the task has scarcely been used with children under 4 years of age, and less frequently with children under 3. This outcome might seem surprising, given that SRTs are devoted to assessing language development, and lower performance on the task can be used as a clinical marker for language impairment. It is a fact, however, that the administration of an SRT is complex with very young children as it is difficult for them to keep their attention focused on the task, and the results are heterogeneous and difficult to score. In any event, research has shown that is possible to conduct SRT with children under 4 years of age (see, for example, [14] for Spanish language, or [55] for English language). Nevertheless, it is an issue of major importance for clinicians and also for researchers interested in more theoretical aspects of SRTs to have data referring to children of this age. The clear gap identified in the use of sentence repetition tasks with children under 4 years old was an unexpected, but significant outcome.

Regarding the type of sentence repetition task used in the reviewed studies, 65% administered a task included in a wider battery assessment, with the CELF (in any of its versions, 4, 5, or Pre-school) being clearly the most frequently used standardized test. There are, however, a good number of other studies that developed their own set of sentences. As many as 50 created these corpora, meaning that at least, there are 50 different original tasks). Nevertheless, this number does not represent all the languages, but only part of them. Interestingly, some researchers making use of languages with little representation developed their own sets, simply because there were no previous sets to be administered (e.g., in Kannada language [43]). Therefore, although there are many languages with few studies published, most of them have their own set of sentences. In fact, the 50 studies reporting an original corpus cover 17 different languages. This number is larger than expected if we consider all the previous outcomes regarding the languages explored.

As we mentioned in the results section, not all studies that use standardized tests or “not original” tasks provide information on the number of sentences included. Considering the ones that do inform about it (see Supplementary Material), we found out that the number of sentences in the tasks widely varies from one to another. Examining the data, most of the studies use around 20. This number, being the mode, broadly represents the number of sentences typically used for the evaluations. Therefore, such a low number of sentences seems to be sufficiently representative of the linguistic structures needed for appropriate proper material for language assessment in every language.

Focusing on the aim of the studies, it has been stated that most of them use SRTs as a tool to assess different language abilities. In addition, 18% of the papers reviewed analyze the potential of these tasks to identify children with different language impairments, mainly children with SLI. This outcome is critical both for theoretical and clinical purposes. If a child with language problems is not identified early and does not receive the necessary intervention, behavioral and academic problems may appear.

Even though we have not analyzed the evidence that supports the effectiveness of SRTs as a clinical marker for SLI, it seems to be well stated in the literature that the performance on SRTs contributes to the detection of children with SLI [23,56]. This is important because, as it is well known, there is currently no gold standard for the diagnosis of SLI [57] that requires the assessment of different language skills [24].

In the case of bilingual children, the number of studies using SRTs to identify children with SLI is still small and, therefore, the conclusions must be taken with caution. However, it is worth mentioning that we have observed an increase of research that include bilingual/ multilingual children in the last years, showing that SRTs are also suitable for this population. Moreover, the evidence suggests this is especially true when using the task in both the languages known by the children, as diagnostic accuracy increases in comparison to when it is administered in just one language.

For future research concerning the use of SRTs, there are still many key issues to be explored. For instance, more qualitative analyses of the results (i.e., error profiles) can be helpful to better understand the difficulties underlying SLI and also to better frame linguistic interventions. To date, few studies have addressed qualitative aspects of children´s responses, but this is a promising path for a deeper understanding of the linguistic development of children with and without typical development.

As highlighted above, future research should focus more on children under 4 years of age. More evidence is needed to ascertain whether these tasks provide useful information to detect and prevent language difficulties in young children. Should the task also prove useful for this purpose, then clinicians would have an efficient tool for assessing and guiding early intervention.

From a theoretical point of view, it remains unclear if these tasks only measure language abilities [58]. Several aspects of verbal memory, lexical knowledge, and morphosyntactic skills appear to be involved in performing the task, but more experimental designs should be carried out to obtain new data to answer the question of what the task measures.

In spite of the different needs that have been detected in this scoping review, the results in this study highlight the utility of SRTs as useful tools for assessing language abilities, and detecting and preventing language difficulties in children.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/children8070578/s1, Dataset: Full database included in the scoping review.

Author Contributions: Conceptualization, E.M., I.R., S.M., and M.L.; formal analysis, E.M. and I.R.; writing—original draft preparation, S.M. and M.L.; writing—review and editing, I.R., E.M., S.M., and M.L.; All authors have read and agreed to the published version of the manuscript.

Funding: Study funded by FEDER, Ministerio de Innovación y Universidades, Agencia Estatal de Investigación PGC-2018-095275-A-100 and CM S12/PGB/2020-00016.

Institutional Review Board Statement: This study is part of a project approved by the Ethics Committee of the Universidad Autónoma de Madrid (protocol code CEI-101-1896).

Data Availability Statement: The full database used in the scoping review is available as Supplementary Material.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Appendix A

Table A1. Studies included in the scoping review (2010–2021).

YEAR	AUTHORS (Year)	MONO/BILINGUAL	LANGUAGE	SAMPLE	N (TOTAL)	AGE RANGE (y;m)	TYPE OF TASK
2021	Andreou, M., et al. (2021)	Both	Greek; Albanian-Greek	TD	70	8 to 12	Adapted
	Friesen, D. C., Edwards, K. & Lamoureux, C. (2021)	Both	English; other languages (14)	TD	94	9 to 12	Not original
	Maher, Z. K., et al. (2021)	Monolingual	African American English (dialects)	TD	475	5;8 to 7;2	Battery
	Pagliarini, E., et al. (2021)	Monolingual	Catalan	TD	56	4;6 to 6;1	Not original
	Polisenká, K., et al. (2021)	Monolingual	English	TD	50	4;0 to 12;0	Original
	Quirk, E. (2021)	Bilingual	French; English	TD	30	5 to 8	Adapted
	Smolík, F. & Matiasovitsová, K. (2021)	Monolingual	Czech	Both	34	3;8 to 7;6	Not original
2020	Antonijevic-Elliott, S., et al. (2020)	Both	English; Polish; Russian	TD	88	5;6 to 7;9	Adapted
	Bedore, L.M. et al. (2020)	Bilingual	English; Spanish	NTD	21	6;4 to 7;7	Battery
	Bogliotti, C., Aksen, H., & Isel, F. (2020)	Monolingual	French	NTD	62	6;8 to 12;8	Original
	Bravo, N., Lázaro, M. & Mariscal, S. (2020)	Monolingual	Spanish	TD	130	2 to 4	Original
	Ebert, K.D., Ochoa-Lubinoff, C., & Holmes, M.P. (2020)	Monolingual	English	TD	66	6;0 to 8;11	Not original
	Gale, R., et al. (2020)	Monolingual	English	Both	104	6	Original
	Hamann, C., et al. (2020)	Bilingual	Arabic L1; German L2	TD	54	6;0 to 12;9	Adapted
	Magimairaj, B.M., et al. (2020)	Monolingual	English	Both	52	7 to 11	Battery
	Matov, J., et al. (2020)	Monolingual	English	Both	182	5;1 to 6;07	Original
	Pham, G. & Ebert, K.D. (2020)	Monolingual	Vietnamese	Both	104	5;2 to 6;2	Original
	Scheidnes, M. (2020)	Both	English; French L2	TD	33	6;04 to 7;05	Not original
	Silleresi, S., et al. (2020)	Both	French	NTD	51	6 to 12	Not original
2019	Abed Ibrahim, L. & Fekete, I. (2019)	Both	German; Arabic, Portuguese, Turkish	Both	77	5;6 to 9;1	Adapted
	Aguilar-Mediavilla, E., et al. (2019)	Bilingual	Spanish; Catalan	Both	28	6 to 12	Battery
	Blything, L.P. & Cain, K. (2019)	Monolingual	English	TD	67	3 to 6	Original
	Chondrogianni, V. & Kwon, H. (2019)	Bilingual	Welsh; English	NTD	52	4 to 9	Adapted
	Cockcroft, K. & Milligan, R. (2019)	Bilingual	Mother Tongue; English (South Africa)	Both	273	6 to 8	Not original
	Fitton, L., et al. (2019)	Bilingual	Spanish; English (American)	TD	291	4;0 to 6;11	Battery
	Hong, T., et al. (2019)	Monolingual	Mandarin	Both	113	4;7 to 21;7	Original
	Jacobson, P.F. & Thompson Miller, S. (2019)	Monolingual	English	NTD	64	6;0 to 7;0	Battery
	Kambanaros, M., Christou, N. & Grohmann, K. K. (2019)	Monolingual	Greek	Both	8	5;11 to 8;9	Adapted
	Keilmann, A., Friese, B. & Hoffmann, V. (2019)	Both	German; other L1	NTD	125	3;0 to 10;11	Battery
	Korat, O., Graister, T. & Altman, C. (2019)	Monolingual	Hebrew	Both	40	4;0 to 5;11	Adapted
	Ladany, E. & Lukács, L. (2019)	Monolingual	Hungarian	Both	54	8	Not original
	Lam, E., et al. (2019)	Monolingual	English (Australian)	Both	16	7;5 to 10;2	Not original
	Leonard, L.B., et al. (2019)	Monolingual	English (American)	Both	27	4;4 to 5;9	Original
	Meir, N. & Novogrodsky, R. (2019)	Both	Hebrew; Russian	Both	85	4 to 9	Adapted
	Méndez, L, I. & Simon-Cereijido, G. (2019)	Bilingual	Spanish; English (American)	NTD	74	3;9 to 4;10	Original
	O'Neill, H., Murphy, C.A. & Chiat, S. (2019)	Monolingual	English	NTD	22	2 to 5	Battery

Table A1. *Cont.*

YEAR	AUTHORS (Year)	MONO/BILINGUAL	LANGUAGE	SAMPLE	N (TOTAL)	AGE RANGE (y;m)	TYPE OF TASK
	Patrucco-Nanchen, T., et al. (2019)	Monolingual	French	TD	49	3	Adapted
	Redmond, S.M, et al. (2019)	Monolingual	English (American)	NTD	251	5;1 to 9;9	Not original
	Særvold, T.K., et al. (2019)	Monolingual	Swedish	NTD	123	10	Battery
	Smith, J., et al. (2019)	Monolingual	English (Australian)	TD	1516	10 to 12	Battery
	Snowling, M.J., et al. (2019)	Monolingual	English	Both	234	3;3 to 8;0	Battery
	Svirko, E., et al. (2019)	Monolingual	English	TD	98	8 to 9	Original
	Vang Christensen, R. (2019)	Monolingual	Danish	Both	114	5;3 to 14;1	Adapted
	Wang, J., et al. (2019)	Monolingual	English (Australian)	Both	1483	11 to 12	Battery
	Wood, C. & Hoge, R. (2019)	Bilingual	Spanish; English (American)	TD	25	5;8 to 6;4	Battery
2018	Aguado, G., et al. (2018) *	Monolingual	Spanish	Both	85	5;0 to 7;0	Battery
	Balilah, A.M. (2018)	Both	English; English language learners	Both	1253	6;0 to 9;11	Not original
	Cai, T., McPherson, B., Li, C., & Yang, F. (2018)	Monolingual	Mandarin	Both	82	8;3 to 8;9	Battery
	Chondrogianni, V., & John, N. (2018)	Bilingual	Welsh; English	Both	28	4 to 7	Original
	Fleckstein, A., et al. (2018)	Both	French; Arabic; English	Both	50	5;0 to 8;0	Adapted
	Gósy, M., et al. (2018)	Monolingual	Hungarian	Both	180	8;4, 9;5, & 10;6	Original
	Magimairaj, B.M., Nagaraj, N.K., & Benafield, N.J. (2018)	Monolingual	English	TD	83	7;0 to 11;0	Battery
	Matov, J., et al. (2018)	Monolingual	English	TD	2212	5;0 & 7;0	Battery
	McIlraith, A. L., & LRR Consortium. (2018)	Monolingual	English	TD	293	3;0 to 5;0	Adapted
	Meir, N. (2018)	Bilingual	Russian; Hebrew	Both	119	5;5 to 6;5	Adapted
	Nag, S., Snowling, M.J., & Mirković, J. (2018)	Monolingual	Kannada	TD	135	5;0 to 8;0	Original
	Novogrodsky, R., Meir, N., & Michael, R. (2018)	Monolingual	Hebrew	Both	54	1;10 to 3;3	Original
	Simon-Cereijido, G., & Méndez, L.I. (2018)	Bilingual	Spanish; English	TD	61	3;11 to 4;10	Original
	Sjögreen, L., Mårtensson, Å., & Ekström, A.B. (2018)	Monolingual	Swedish	Both	50	7;0 to 29;0	Battery
	Sterling, A. (2018)	Monolingual	English	NTD	37	12;0 to 13;0	Adapted
	Sukenik, N., & Friedmann, N. (2018)	Monolingual	Hebrew	Both	277	5;0 to 18;0	Adapted
	Taliancich-Klinger, C.L., Bedore, L.M., & Peña, E.D. (2018)	Bilingual	Spanish; English	TD	148	7;0 to 9;11	Adapted
	Tuller, L., et al. (2018)	Bilingual	Arabic; Portuguese; Turkish; French; German	Both	151	5;0 to 9;0	Adapted
2017	Brynskov, C., et al. (2017)	Monolingual	Danish	Both	42	4 to 6	Adapted
	Courtney, L., et al. (2017)	Other	French L2	TD	254	9 to 10	Adapted
	Feragen, K.B., et al. (2017)	Monolingual	Norwegian	NTD	170	10	Battery
	Frizelle, P., O'Neil, C, & Bishop, D.V. (2017)	Monolingual	English	TD	33	5 to 6;06	Original
	Gavarró, A. (2017)	Bilingual	Catalan; Spanish	Both	35	6 to 17;4	Adapted
	Gooch, D., et al. (2017)	Monolingual	English	Both	200	6;0	Not original
	Graham, S., et al. (2017)	Monolingual	French learners in England	TD	252	9;10 to 10;10	Adapted
	Haman, E., et al. (2017)	Both	English; Polish	TD	233	4 to 7;5	Adapted
	Kirjavainen, M., Kidd, E., & Lieven, E. (2017)	Monolingual	Finnish	TD	37	3;7 to 4;6	Original
	Ladanyi, E., Kas, B., & Lukács, Á. (2017)	Monolingual	Hungarian	Both	60	7 to 10	Not original
	Lohmander, A., Lundeborg, I., & Persson, C. (2017)	Both	Swedish; other languages	TD	443	3 to 19	Battery
	Meir, N., & Armon-Lotem, S. (2017)	Both	Hebrew; Russian-Hebrew	TD	120	5;7 to 6;7	Not original

Table A1. *Cont.*

YEAR	AUTHORS (Year)	MONO/BILINGUAL	LANGUAGE	SAMPLE	N (TOTAL)	AGE RANGE (y;m)	TYPE OF TASK
	Norbury, C. F., et al. (2017)	Monolingual	English	Both	529	5;11 to 7;11	Not original
	Oryadi-Zanjani, M.M., et al. (2017)	Monolingual	Persian	Both	92	5 to 7	Not original
	Redmond, S.M., & Ash, A.C. (2017)	Monolingual	English	Both	105	5;6 to 11	Not original
	Riches, N.G. (2017)	Monolingual	English	Both	55	4;0 to 7;3	Original
	Theodorou, E., Kambanaros, M., & Grohmann, K.K. (2017)	Monolingual	Cypriot Greek	Both	38	5 to 9	Original
	Zamani, P., et al. (2017)	Monolingual	Persian	Both	112	4;9 to 5;1	NP
2016	Alonzo, C.N.,et al. (2016)	Monolingual	English	TD	318	5;0	Battery
	Armon-Lotem, S., & Meir, N. (2016)	Both	Russian; Hebrew	Both	230	5;5 to 6;8	Adapted
	Balladares, J., Marshall, C., & Griffiths, Y. (2016)	Monolingual	Spanish (Chile)	TD	126	3;10 to 6;3	Adapted
	Buil-Legaz, L., Aguilar-Mediavilla, E., & Adrover-Roig, D. (2016)	Bilingual	Spanish; Catalan	Both	37	6 to 12	Battery
	Gagarina, N. (2016)	Bilingual	Russian; German	TD	58	2 to 10	Adapted
	Gooch, D.,et al. (2016).	Monolingual	English	Both	243	4;8 to 6;6	Battery
	Kapantzoglou, M., et al. (2016)	Monolingual	Spanish	Both	307	5 to 7	Original
	Klem, M.,et al. (2016)	Monolingual	Norwegian	TD	206	4;5 & 6	Adapted
	Ladányi, E., & Lukács, Á. (2016)	Both	Hungarian	Both	26	7 to 9	Not original
	Mari, G., et al. (2016)	Monolingual	Italian	Both	53	3 to 6	Not original
	Moreno-Torres, I., et al. (2016) *	Monolingual	Spanish	Both	14	4 to 4;8	Not original
	Oetting, J.B., et al. (2016)	NP	English	Both	106	4 to 6	Original
	Oudgenoeg-Paz, O., Volman, M.C., & Leseman, P.P. (2016)	Both	Dutch	TD	31	3;5	Not original
	Van Dijk, C.N., et al. (2016)	Monolingual	Dutch	TD	55	10 to 13	Battery
2015	Annett, R.D., et al. (2015)	NP	NP	NTD	188	4 to 21	Not original
	Babayiğit, S. (2015)	Monolingual	English	TD	183	9	Battery
	Foltz, A., et al. (2015)	Monolingual	German	Both	24	4;0 to 5;7	NP
	Francis, N.M.P., & Thomas, I. (2015)	NP	Indian	Both	80	9 to 15	Battery
	Frizelle, P., & Fletcher, P. (2015)	Monolingual	English	Both	67	4;7 to 7;11	Original
	Garraffa, M., Coco, M. I., & Branigan, H.P. (2015)	Monolingual	Italian	Both	38	4 to 6	Original
	Hämäläinen, J.A., et al. (2015)	Monolingual	Finnish	Both	37	5 to 6	Battery
	Klem, M., et al. (2015)	Monolingual	Norwegian	TD	216	4 to 6	Adapted
	Marshall, C., et al. (2015)	Monolingual	British Sign language	NTD	22	6 to 13	Original
	Moll, K., et al. (2015)	Monolingual	English	Both	97	6 to 12	Original
	Polišenská, K., Chiat, S., & Roy, P. (2015)	Monolingual	English; Czech	TD	100	4;0 to 5;11	Original
	Redmond, S.M., Ash, A.C., & Hogan, T.P. (2015)	Monolingual	English (American)	Both	57	7;0 to 9;9	Not original
	Tambyraja, S.R., et al. (2015)	Monolingual	English	NTD	118	6;3	Battery
	Vernice, M., & Guasti, M.T. (2015)	Monolingual	Italian	TD	48	4;0 to 5;11	Original

Table A1. *Cont.*

YEAR	AUTHORS (Year)	MONO/BILINGUAL	LANGUAGE	SAMPLE	N (TOTAL)	AGE RANGE (y;m)	TYPE OF TASK
2014	Aguilar-Mediavilla, E., et al. (2014)	Bilingual	Spanish; Catalan	Both	34	6;0 to 8;0	Battery
	Anthony, J.L., et al. (2014)	Monolingual	English	TD	529	4;0	Battery
	Armon-Lotem, S. (2014)	Bilingual	Hebrew; English; Russian	Both	43	5;0 to 7;0	Original
	Chilosi, A.M., et al. (2014)	Monolingual	Italian	Both	40	4;0 to 14;0	Adapted
	Ebert, K.D. (2014)	Bilingual	English; Spanish	NTD	47	5;0 to 11;0	Battery
	Erdos, C., et al. (2014)	Other	English (French as L2)	TD	86	5;0 to 6;0	Battery
	Frizelle, P., & Fletcher, P. (2014)	Monolingual	English	Both	84	4;9 to 6;11	Original
	Frizelle, P., & Fletcher, P. (2014)	Monolingual	English	Both	84	4;9 to 6;11	Original
	Heath, S.M., et al. (2014)	Monolingual	English	TD	102	4;3 to 6;8	Battery
	Horton, R., & Apel, K. (2014)	Monolingual	English	TD	113	5;11 to 7;11	Not original
	Leclercq, A.L., et al. (2014)	Monolingual	French	Both	68	7;0 to 12;0	Battery
	Lewis, D.E., & Wannagot, S. (2014)	Monolingual	English	TD	50	8;0 to 11;0	Adapted
	Loukusa, S., et al. (2014)	Monolingual	Finnish	Both	57	5;0 to 7;0	Battery
	Lüke, C., & Ritterfeld, U. (2014)	Both	German	NTD	20	3;0 to 5;0	Battery
	Määttä, S., et al. (2014)	Monolingual	Finnish	Both	91	5;3	Battery
	Mitri, S.M., & Terry, N.P. (2014)	Monolingual	English	TD	119	7;3	Battery
	Polišenská, K., et al. (2014)	Monolingual	English	TD	24	6;0	Original
	Smolík, F., & Vávrů, P. (2014)	Monolingual	Czech	Both	57	4;10 to 7;8	Original
	Taylor, L.J., et al. (2014)	Monolingual	English	Both	117	5;0 to 12;8	Battery
	Terry, N.P. (2014)	Monolingual	English	TD	105	7;2	Battery
	Terzi, A., et al. (2014)	Monolingual	Greek	Both	40	5;0 to 8;0	Original
	Venkatesh, S.K., et al. (2014)	NP	NP	Both	466	6;0 to 18;0	Battery
2013	Abdalla, F., Aljenaie, K., & Mahfoudhi, A. (2013)	Monolingual	Arabic	Both	24	3;7 to 6;2	Battery
	Archibald, L. (2013)	Monolingual	English	TD	374	5 to 9:11	Not original
	Boyle, W., Lindell, A.K., & Kidd, E. (2013)	Monolingual	English	TD	50	4 to 5;10	Battery
	Calderoni, S., et al. (2013)	Monolingual	Italian	Both	69	9 to 16	Battery
	Everitt, A., Hannaford, P., & Conti-Ramsden, G. (2013)	Monolingual	English	Both	94	3 to 5	Battery
	Grunewaldt, K., et al. (2013)	Monolingual	Norwegian	NTD	20	5 to 6	Battery
	Harper-Hill, K., Copland, D., & Arnott, W. (2013)	Monolingual	English	Both	35	9 to 16	Battery
	Hesketh, A., & Conti-Ramsden, G. (2013)	Monolingual	English	Both	272	11	Adapted
	Howard, S. (2013)	Monolingual	English	Both	5	9:5 to 11:0	Not original
	Komeili, M., & Marshall, C.R. (2013)	Bilingual	English; farsi	TD	36	5;7 to 12;5	Not original
	Leonard, L. B., & Dispaldro, M. (2013)	Monolingual	Italian	Both	45	2;9 to 5;8	Battery
	Lidzba, K., et al. (2013)	Monolingual	German	Both	53	8 to 25	Original
	Lukács, Á., Kas, B & Leonard, L. B. (2013)	Monolingual	Hungarian	Both	92	4;10 to 11;4	Original
	Michalczyk, K., et al. (2013)	Both	German; other languages	TD	1343	5 to 6	Not original
	Moreno-Torres, I., Madrid, S., & Moruno, E. (2013)	Monolingual	Spanish	Both	28	2;2 to 3;6	Original
	Narzisi, A., et al. (2013)	Monolingual	Italian	Both	66	5 to 16	Battery
	Nash, H. M.,et al. (2013)	Monolingual	English	Both	112	3;06 to 4;06	Battery
	Petersen, D. & Gillam, R. (2013)	Bilingual	Spanish; English	NTD	63	5;5	Original
	Poll, G.H., et al. (2013)	Monolingual	English	Both	70	6 to 13	Battery

Table A1. *Cont.*

YEAR	AUTHORS (Year)	MONO/BILINGUAL	LANGUAGE	SAMPLE	N (TOTAL)	AGE RANGE (y;m)	TYPE OF TASK
	Thordardottir, E. & Brandeker, M. (2013)	Both	English; French	Both	140	4;06 to 7;02	Battery
	Walsh, B., & Smith, A. (2013)	Monolingual	English	Both	104	3;5 to 5;11	Original
	Windsor, J., et al. (2013)	Monolingual	Romanian	NTD	105	8	Original
	Ziethe, A., Eysholdt, U., & Doellinger, M. (2013)	Bilingual	German; others	Both	73	4;9 to 8;9	Battery
2012	Caselli, M. C., et al. (2012)	Monolingual	Italian	Both	50	3 to 5	Not original
	Christensen, R. V., & Hansson, K. (2012)	Monolingual	Danish	Both	33	5 to 7	Original
	Delcenserie, A., Genesee, F., Gauthier, K. (2013)	Monolingual	French	Both	54	6;8 to 8;10	Battery
	Kenworthy, L., et al. (2012)	Monolingual	English	NTD	76	9	Adapted
	Leonard, L.B., Lukács, Á., & Kas, B. (2012)	Monolingual	Hungarian	Both	63	3;3 to 7;3	Not original
	Narbona, J., Yglesias-Pereira, A., & García-López, C. (2012) *	Monolingual	Spanish (Hispanic)	Both	52	4 to 16	Not original
	Peter, B. (2012)	Monolingual	English	Both	22	4;6 to 7;0	Battery
	Petruccelli, N., Bavin, E.L., & Bretherton, L. (2012)	Monolingual	English	NTD	102	5	Battery
	Puranik, C.S., & AlOtaiba, S. (2012)	Monolingual	English	TD	242	5;0 to 7;8	Battery
	Riches, N.G. (2012)	Monolingual	English	Both	63	4;0 to 7;3	Original
	Yglesias-Pereira, A., Garcia-Lopez, C., & Narbona, J. (2012)	Monolingual	Spanish	Both	52	3;5 to 7;5	Not original
2011	Anthony, J. L., et al. (2011)	Bilingual	English; Spanish	TD	129	4	Not original
	Anthony, J. L.,et al. (2011)	Monolingual	English	Both	204	3;5 to 5;6	Not original
	Elbeheri, G., et al. (2011)	Monolingual	Arabic	Both	387	7 to 10	Original
	Fattal, I., Friedmann, N. & Fattal-Valevski, A. (2011)	Monolingual	Hebrew	Both	59	5 to 7	Original
	Geers, A.E .& Sedey, A.L. (2011)	Monolingual	English	NTD	112	15;0 to 18;6	Not original
	Gou, Z., Choudhury, N., & Benasich, A.A. (2011)	Monolingual	English	NTD	95	5	Battery
	Hirata-Edds, T. (2011)	Monolingual	Cherokee (L2) & English (l1)	TD	23	4;5 to 6;1	Original
	Hooper, S.R., et al. (2011)	Monolingual	English	TD	205	6 to 7;3	Not original
	Klintö, K., et al. (2011)	Monolingual	Swedish	Both	40	5	Not original
	Kronenberger, W.G., et al. (2011)	Monolingual	American English	NTD	9	7 to 15	Battery
	Lewis, B. A., et al. (2011)	Monolingual	English	Both	237	4 to 7	Battery
	Lewis, B. A., et al. (2011)	Monolingual	English	NTD	152	4 to 6	Battery
	Mattis, S., et al. (2011)	Monolingual	American English	Both	53	7 to 15	Not original
	Nag, S. & Snowling, M.J. (2011)	Bilingual	Kannada; English (L2)	Both	103	8 to 10	Original
	Nash, H., Leavett, R., & Childs, H. (2011).	Monolingual	English	TD	106	3 to 6	Not original
	Nelson, K.E., et al. (2011)	Monolingual	English	TD	336	4	Battery
	Nevo, E. & Breznitz, Z. (2011)	Monolingual	Hebrew	TD	97	5;5 to 6;11	Battery
	Nittrouer, S., Shune, S. & Lowenstein, J.H. (2011)	Monolingual	American English	Both	28	8 & adults	Battery
	Pecini, C., et al. (2011)	Monolingual	Italian	Both	13	12 to 50	Not original
	Pérez-Leroux, A.T., Cuza, A. & Thomas, D. (2011)	Bilingual	Spanish; English	TD	23	3 to 8	Adapted
	Redmond, S. M., Thompson, H.L. & Goldstein, S. (2011)	Monolingual	American English	Both	60	7 to 8	Not original
	Spanoudis, G.C. & Natsopoulos, D. (2011)	Monolingual	Greek; Cypriot	Both	100	8 to 12	Adapted
	Thordardottir, E., et al. (2011)	Monolingual	French	Both	92	4 to 5	Adapted

Table A1. *Cont.*

YEAR	AUTHORS (Year)	MONO/BILINGUAL	LANGUAGE	SAMPLE	N (TOTAL)	AGE RANGE (y;m)	TYPE OF TASK
2010	Anthony, J.L., et al. (2010)	Monolingual	English	TD	175	3 to 5	Original
	Coady, J., Evans, J.L., & Kluender, K.R. (2010)	Monolingual	English	Both	36	7;3 to 10;6	Adapted
	Dockrell, J.E., Stuart, M., & King, D. (2010)	Other	English (L2)	TD	142	4	Not original
	Geurts, H., & Embrechts, M. (2010)	Monolingual	Dutch	Both	56	5	Not original
	Ho, A.K., & Wilmut, K. (2010)	Monolingual	English	Both	10	9 to 13	Original
	Kemény, F., & Lukács, Á. (2010)	Monolingual	Hungarian	Both	32	11 & adults	Not original
	Lukács, A.G., Leonard, L.B., & Kas, B. (2010)	Monolingual	Hungarian	Both	60	4;10 to 9;10	Not original
	Marini, A., et al. (2010)	Monolingual	Italian	Both	38	6 to 25	Not original
	Martínez, M. et al. (2010) *	Monolingual	Spanish (Colombia)	Both	16	7 to 9	Battery
	Riches, N.G., et al. (2010)	Monolingual	English	Both	47	14 to 15	Original
	Seeff-Gabriel, B., Chiat, S., & Dodd, B. (2010)	Monolingual	English	Both	72	4 to 6	Original
	Tavano, A., et al. (2010)	Monolingual	Italian	Both	52	3;0 to 29;11	Not original
	Terry, J. M., Jackson, S. C., Evangelou, E., & Smith, R. L. (2010)	Monolingual	African American English	TD	48	7	Battery
	Thordardottir, E., et al. (2010)	Monolingual	Quebec French	TD	78	4;06 to 5;06	Battery
	Wolter, J. A. & Apel, K. (2010)	Monolingual	English	Both	52	5;07 to 6;04	Battery

* Studies published in Spanish. TD = Typical development; NTD = Non-Typical Development. Multilingual studies have been included in "Bilingual" category.

References

1. Brown, R.; Fraser, C. The Acquisition of Syntax. *Monogr. Soc. Res. Child Dev.* **1964**, *29*, 43. [CrossRef]
2. Slobin, D.I.; Welsh, C.A. Elicited imitation as a research tool in Developmental Psycholinguistics. In *Studies in Child Language Development*; Ferguson, C.A., Slobin, D.I., Eds.; Holt, Rinehart & Wilson: New York, NY, USA, 1973.
3. Lee, L. *The Northwestern Syntax Screening Test*; Northwestern University Press: Evanston, IL, USA, 1971.
4. Levelt, W.J.M. Spoken word production: A theory of lexical access. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 13464–13471. [CrossRef] [PubMed]
5. Devescovi, A.; Caselli, M.C. Sentence repetition as a measure of early grammatical development in Italian. *Int. J. Lang. Commun. Disord.* **2007**, *42*, 187–208. [CrossRef] [PubMed]
6. Bishop, D.V. What Causes Specific Language Impairment in Children? *Curr. Dir. Psychol. Sci.* **2006**, *15*, 217–221. [CrossRef]
7. Rescorla, L. Late Talkers: Do Good Predictors of Outcome Exist? *Dev. Disabil. Res. Rev.* **2011**, *17*, 141–150. [CrossRef]
8. Conti-Ramsden, G.; Botting, N.; Faragher, B. Psycholinguistic Markers for Specific Language Impairment (SLI). *J. Child Psychol. Psychiatry* **2001**, *42*, 741–748. [CrossRef] [PubMed]
9. Semel, E.; Wiig, E.; Secord, W. *Clinical Evaluation of Language Fundamentals-Revised (CELF-4)*; The Psychological Corporation: London, UK, 1994.
10. Pawłowska, M. Evaluation of Three Proposed Markers for Language Impairment in English: A Meta-Analysis of Diagnostic Accuracy Studies. *J. Speech Lang. Hear. Res.* **2014**, *57*, 2261–2273. [CrossRef]
11. Redmond, S.M.; Ash, A.C.; Christopulos, T.T.; Pfaff, T. Diagnostic Accuracy of Sentence Recall and Past Tense Measures for Identifying Children's Language Impairments. *J. Speech Lang. Hear. Res.* **2019**, *62*, 2438–2454. [CrossRef]
12. Chiat, S.; Armon-Lotem, S.; Marinis, T.; Polišenská, K.; Roy, P.; Seeff-Gabriel, B. 3. Assessment of Language Abilities in Sequential Bilingual Children: The Potential of Sentence Imitation Tasks. In *Issues in the Assessment of Bilinguals*; Multilingual Matters: Bristol, UK, 2013; pp. 56–89.
13. Armon-Lotem, S.; Meir, N. Diagnostic accuracy of repetition tasks for the identification of specific language impairment (SLI) in bilingual children: Evidence from Russian and Hebrew. *Int. J. Lang. Commun. Disord.* **2016**, *51*, 715–731. [CrossRef]
14. Bravo, N.; Lázaro, M.; Mariscal, S. A Sentence Repetition Task for Early Language Assessment in Spanish. *Span. J. Psychol.* **2020**, *23*, e39. [CrossRef]
15. Chondrogianni, V.; Kwon, H. The development of English tense and agreement morphology in Welsh–English bilingual children with and without specific language impairment. *Appl. Psycholinguist.* **2019**, *40*, 821–852. [CrossRef]
16. Gavarró, A. A Sentence Repetition Task for Catalan-Speaking Typically-Developing Children and Children with Specific Language Impairment. *Front. Psychol.* **2017**, *8*, 1865. [CrossRef]
17. Hamann, C.; Chilla, S.; Ibrahim, L.A.; Fekete, I. Language assessment tools for Arabic-speaking heritage and refugee children in Germany. *Appl. Psycholinguist.* **2020**, *41*, 1375–1414. [CrossRef]
18. Leonard, L.B.; Lukács, Á.; Kas, B. Tense and aspect in childhood language impairment: Contributions from Hungarian. *Appl. Psycholinguist.* **2011**, *33*, 305–328. [CrossRef]
19. Pham, G.; Ebert, K.D. Diagnostic Accuracy of Sentence Repetition and Nonword Repetition for Developmental Language Disorder in Vietnamese. *J. Speech Lang. Hear. Res.* **2020**, *63*, 1521–1536. [CrossRef]
20. Smolík, F.; Matiasovitsová, K. Sentence Imitation With Masked Morphemes in Czech: Memory, Morpheme Frequency, and Morphological Richness. *J. Speech Lang. Hear. Res.* **2021**, *64*, 105–120. [CrossRef]
21. Stokes, S.F.; Wong, A.M.-Y.; Fletcher, P.; Leonard, L.B. Nonword Repetition and Sentence Repetition as Clinical Markers of Specific Language Impairment: The Case of Cantonese. *J. Speech Lang. Hear. Res.* **2006**, *49*, 219–236. [CrossRef]
22. Taliancich-Klinger, C.L.; Bedore, L.M.; Pena, E. Preposition accuracy on a sentence repetition task in school age Spanish–English bilinguals. *J. Child Lang.* **2017**, *45*, 97–119. [CrossRef] [PubMed]
23. Christensen, R.V. Sentence Repetition: A Clinical Marker for Developmental Language Disorder in Danish. *J. Speech Lang. Hear. Res.* **2019**, *62*, 4450–4463. [CrossRef] [PubMed]
24. Armon-Lotem, S.; Marinis, T. 5. Sentence Repetition. In *Assessing Multilingual Children*; Multilingual Matters: Bristol, UK, 2015; pp. 95–122.
25. Fleckstein, A.; Prévost, P.; Tuller, L.; Sizaret, E.; Zebib, R. How to identify SLI in bilingual children: A study on sentence repetition in French. *Lang. Acquis.* **2016**, *25*, 85–101. [CrossRef]
26. Tuller, L.; Hamann, C.; Chilla, S.; Ferré, S.; Morin, E.; Prevost, P.; Dos Santos, C.; Ibrahim, L.A.; Zebib, R. Identifying language impairment in bilingual children in France and in Germany. *Int. J. Lang. Commun. Disord.* **2018**, *53*, 888–904. [CrossRef]
27. Frizelle, P.; O'Neill, C.; Bishop, D. Assessing understanding of relative clauses: A comparison of multiple-choice comprehension versus sentence repetition. *J. Child Lang.* **2017**, *44*, 1435–1457. [CrossRef] [PubMed]
28. Redmond, S.M.; Thompson, H.L.; Goldstein, S. Psycholinguistic Profiling Differentiates Specific Language Impairment From Typical Development and From Attention-Deficit/Hyperactivity Disorder. *J. Speech Lang. Hear. Res.* **2011**, *54*, 99–117. [CrossRef]
29. Seeff-Gabriel, B.; Chiat, S.; Dodd, B. Sentence imitation as a tool in identifying expressive morphosyntactic difficulties in children with severe speech difficulties. *Int. J. Lang. Commun. Disord.* **2010**, *45*, 691–702. [CrossRef] [PubMed]
30. Petruccelli, N.; Bavin, E.L.; Bretherton, L. Children With Specific Language Impairment and Resolved Late Talkers: Working Memory Profiles at 5 Years. *J. Speech Lang. Hear. Res.* **2012**, *55*, 1690–1703. [CrossRef]

31. Taylor, L.J.; Maybery, M.T.; Grayndler, L.; Whitehouse, A.J.O. Evidence for Distinct Cognitive Profiles in Autism Spectrum Disorders and Specific Language Impairment. *J. Autism Dev. Disord.* **2013**, *44*, 19–30. [CrossRef]
32. Liberati, A.; Altman, D.G.; Tetzlaff, J.; Mulrow, C.; Gøtzsche, P.C.; Ioannidis, J.P.A.; Clarke, M.; Devereaux, P.J.; Kleijnen, J.; Moher, D. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *PLoS Med.* **2009**, *6*, e1000100. [CrossRef]
33. Bedore, L.M.; Pena, E.; Fiestas, C.; Lugo-Neris, M.J. Language and Literacy Together: Supporting Grammatical Development in Dual Language Learners With Risk for Language and Learning Difficulties. *Lang. Speech Hear. Serv. Sch.* **2020**, *51*, 282–297. [CrossRef] [PubMed]
34. Méndez, L.I.; Simon-Cereijido, G. A View of the Lexical–Grammatical Link in Young Latinos with Specific Language Impairment Using Language-Specific and Conceptual Measures. *J. Speech Lang. Hear. Res.* **2019**, *62*, 1775–1786. [CrossRef]
35. Wood, C.; Hoge, R. Average change in sentence repetition by Spanish-English speaking children: Kindergarten to first grade. *Int. J. Biling. Educ. Biling.* **2017**, *22*, 725–740. [CrossRef]
36. Fitton, L.; Hoge, R.; Petscher, Y.; Wood, C. Psychometric Evaluation of the Bilingual English–Spanish Assessment Sentence Repetition Task for Clinical Decision Making. *J. Speech Lang. Hear. Res.* **2019**, *62*, 1906–1922. [CrossRef]
37. Aguado, G.; Ripoll, J.C.; Tapia, M.M.; Gibson, M. Marcadores del trastorno específico del lenguaje en español: Comparación entre la repetición de oraciones y la repetición de pseudopalabras. *Rev. Logop. Foniatría Audiol.* **2018**, *38*, 105–112. [CrossRef]
38. Kapantzoglou, M.; Thompson, M.S.; Gray, S.; Restrepo, M.A. Assessing Measurement Invariance for Spanish Sentence Repetition and Morphology Elicitation Tasks. *J. Speech Lang. Hear. Res.* **2016**, *59*, 254–266. [CrossRef]
39. Sánchez, I.M.-T.; Cánovas, S.M.; López, E.M. Prueba repetición de oraciones para niños de 24 a 48 meses (PRO-24). Estudio piloto con niños típicos y niños sordos con implante coclear. *Rev. Logop. Foniatría Audiol.* **2013**, *33*, 25–35. [CrossRef]
40. Abdalla, F.; Aljenaie, K.; Mahfoudhi, A. Plural noun inflection in Kuwaiti Arabic-speaking children with and without Specific Language Impairment. *J. Child Lang.* **2013**, *40*, 139–168. [CrossRef]
41. Elbeheri, G.; Everatt, J.; Mahfoudhi, A.; Abu Al-Diyar, M.; Taibah, N. Orthographic processing and reading comprehension among Arabic speaking mainstream and LD children. *Dyslexia* **2011**, *17*, 123–142. [CrossRef] [PubMed]
42. Smolík, F.; Vávrů, P. Sentence Imitation as a Marker of SLI in Czech: Disproportionate Impairment of Verbs and Clitics. *J. Speech Lang. Hear. Res.* **2014**, *57*, 837–849. [CrossRef] [PubMed]
43. Nag, S.; Snowling, M.J.; Mirković, J. The role of language production mechanisms in children's sentence repetition: Evidence from an inflectionally rich language. *Appl. Psycholinguist.* **2017**, *39*, 303–325. [CrossRef]
44. Gagarina, N. Narratives of Russian–German preschool and primary school bilinguals: Rasskaz and Erzaehlung. *Appl. Psycholinguist.* **2015**, *37*, 91–122. [CrossRef]
45. Aguilar-Mediavilla, E.; Buil-Legaz, L.; López-Penadés, R.; Sanchez-Azanza, V.A.; Adrover-Roig, D. Academic Outcomes in Bilingual Children With Developmental Language Disorder: A Longitudinal Study. *Front. Psychol.* **2019**, *10*, 531. [CrossRef]
46. Pereira, A.Y.; López, C.G.; García, J.N. Test de escucha dicótica para niños hispanohablantes: Validación de unas nuevas series de pares de palabras y de sílabas en castellano. *Rev. Neurol.* **2012**, *55*, 137. [CrossRef]
47. Silleresi, S.; Prévost, P.; Zebib, R.; Bonnet-Brilhault, F.; Conte, D.; Tuller, L. Identifying Language and Cognitive Profiles in Children With ASD via a Cluster Analysis Exploration: Implications for the New ICD-11. *Autism Res.* **2020**, *13*, 1155–1167. [CrossRef] [PubMed]
48. Meir, N.; Novogrodsky, R. Prerequisites of Third-Person Pronoun Use in Monolingual and Bilingual Children With Autism and Typical Language Development. *Front. Psychol.* **2019**, *10*, 2289. [CrossRef]
49. Sukenik, N.; Friedmann, N. ASD Is Not DLI: Individuals With Autism and Individuals With Syntactic DLI Show Similar Performance Level in Syntactic Tasks, but Different Error Patterns. *Front. Psychol.* **2018**, *9*, 279. [CrossRef] [PubMed]
50. Brynskov, C.; Eigsti, I.-M.; Jørgensen, M.; Lemcke, S.; Bohn, O.-S.; Krøjgaard, P. Syntax and Morphology in Danish-Speaking Children with Autism Spectrum Disorder. *J. Autism Dev. Disord.* **2016**, *47*, 373–383. [CrossRef] [PubMed]
51. Zamani, P.; Ravanbakhsh, M.; Weisi, F.; Rashedi, V.; Naderi, S.; Hosseinzadeh, A.; Rezaei, M. Effect(s) of Language Tasks on Severity of Disfluencies in Preschool Children with Stuttering. *J. Psycholinguist. Res.* **2016**, *46*, 261–269. [CrossRef]
52. Moll, K.; Hulme, C.; Nag, S.; Snowling, M.J. Sentence repetition as a marker of language skills in children with dyslexia. *Appl. Psycholinguist.* **2015**, *36*, 203–221. [CrossRef]
53. Howard, S. A Phonetic Investigation of Single Word versus Connected Speech Production in Children with Persisting Speech Difficulties Relating to Cleft Palate. *Cleft Palate-Craniofacial J.* **2013**, *50*, 207–223. [CrossRef] [PubMed]
54. Matov, J.; Mensah, F.; Cook, F.; Reilly, S. Investigation of the language tasks to include in a short-language measure for children in the early school years. *Int. J. Lang. Commun. Disord.* **2018**, *53*, 735–747. [CrossRef]
55. O'Neill, H.; Murphy, C.-A.; Chiat, S. What Our Hands Tell Us: A Two-Year Follow-Up Investigating Outcomes in Subgroups of Children With Language Delay. *J. Speech Lang. Hear. Res.* **2019**, *62*, 356–366. [CrossRef]
56. Leclercq, A.-L.; Quémart, P.; Magis, D.; Maillart, C. The sentence repetition task: A powerful diagnostic tool for French children with specific language impairment. *Res. Dev. Disabil.* **2014**, *35*, 3423–3430. [CrossRef] [PubMed]
57. Spaulding, T.J.; Plante, E.; Farinella, K.A. Eligibility Criteria for Language Impairment. *Lang. Speech Hear. Serv. Sch.* **2006**, *37*, 61–72. [CrossRef]
58. Klem, M.; Melby-Lervåg, M.; Hagtvet, B.E.; Lyster, S.H.; Gustafsson, J.; Hulme, C. Sentence repetition is a measure of children's language skills rather than working memory limitations. *Dev. Sci.* **2014**, *18*, 146–154. [CrossRef] [PubMed]

children

MDPI

Article

A First Step toward the Clinical Application of Landmark-Based Acoustic Analysis in Child Mandarin

Chin-Ting Liu

Department of Applied English, National Chin-Yi University of Technology, Taichung 411030, Taiwan; ctjimboliu@gmail.com

Abstract: As an initial step for the clinical application of landmark-based acoustic analysis in child Mandarin, the study quantified the developmental trajectories of consonants produced by four-to-seven-year-old children who acquired Taiwanese Mandarin as their first language. The results from a total of 80 children (20 in each age group, with gender balanced) indicated that younger age groups produced more +*b* landmark features than seven-year-olds did, showing that the development of obstruents was not completed by the age of six. A multiple regression showed that the participants' speech intelligibility scores could be predicted by landmark features. Additionally, the +*b* landmark feature demonstrated the strongest net effect on speech intelligibility scores. The findings indicated that: (a) the landmark feature +*b* was an essential indicator of speech development in child Mandarin and; (b) the consonantal development in child Mandarin could be predicted by the physiological complexity of the articulatory gestures. Future studies focusing on a wider range of population (e.g., typically developing adults, aging and other clinical groups) with different language backgrounds are encouraged to apply landmark-based acoustic analysis to trace the linguistic development of a particular group.

Keywords: acoustics; landmark analysis; Mandarin Chinese; language acquisition; consonant

Citation: Liu, C.-T. A First Step toward the Clinical Application of Landmark-Based Acoustic Analysis in Child Mandarin. *Children* **2021**, *8*, 159. https://doi.org/10.3390/children8020159

Academic Editor: Eva Aguilar Mediavilla

Received: 17 January 2021
Accepted: 17 February 2021
Published: 20 February 2021

1. Introduction

A massive body of literature has pointed out that the traditional manual segmentation and acoustical analyses of speech are too laborious and time-consuming (c.f. [1–6], among many others). As the transcription and coding are labor-intensive, the number of instances included in analyses is usually limited (c.f. [1,7]). This issue is particularly critical for pediatricians and language therapists because young children with high risks of speech disorders usually have limited energy and attention span. Therefore the speech evaluation sessions are less likely to be long enough to include a larger corpus of speech data produced by the children. In view of this limitation, several newly created devices and software have emerged with the aim of enabling researchers to analyze a larger body of samples with high validity and reliability without consuming too much time. SpeechMark© (Boston, MA, USA) [8] is one of those products and is built upon previous works by Liu [9] and Howitt [10].

SpeechMark© has been developed based on the landmark-based theory proposed by Stevens [11–15]. Unlike the traditionally proposed articulator-bound features (c.f. [16,17]), landmark-based analysis is an articulator-free analysis that focuses on the rapid change in spectrum or amplitude. These abrupt changes in spectrum or amplitude are said to correlate with speech intelligibility [18–22]. That is, listeners rely on those changes to judge what the perceived speech sounds are. At first, there were three types of landmarks, including $\pm g$ (lottis), $\pm b$ (urst) and $\pm s$ (onorant) (c.f. [9]), where the symbols '+' (positive) and '−' (negative) refer to the onset and the offset of the feature, respectively. Additional features are added when researchers develop SpeechMark© based on their observations of speech recordings. The specifications and the articulatory interpretations of the six abrupt-consonantal landmarks based on DiCicco and Patel [23], MacAuslan [24], Ishikawa, MacAuslan and Boyce [19], Atkins, Boyce, MacAuslan and Silbert [21], Huang, Epps and Joachim [25] and Ishikawa, Rao, MacAuslan and Boyce [22] are summarized in Table 1. Landmark-based acoustic analysis has been used to study the linguistic behaviors of several populations, including typically developing (TD) adults [18,19], individuals with dysarthria [23], children with cleft lip and palate [20], simultaneous bilingual children [26] and individuals with dysphonic speech [22].

Table 1. Acoustic rules and articulatory interpretations of the six abrupt-consonantal landmarks.

Symbol	Mnemonic	Acoustic Rule [1]	Articulatory Interpretation
$\pm g$	Glottal	Beginning/end of sustained laryngeal vibration/motion	Onset/offset of vocal folds' free vibration
$\pm p$	Periodicity	Beginning/end of sustained periodicity (syllabicity) lasting for at least 32 milliseconds	The presence of $\pm p$ reflects the speaker's ability to properly control the subglottal pressure and cricothyroid muscle.
$\pm b$	Burst	At least three of five frequency bands show simultaneous power increases/decreases of at least 6 dB in both the finely smoothed and the coarsely smoothed contours in an unvoiced segment (not between $+g$ and the next $-g$)	Presence of a fricative, affricate or aspirated stop burst consonant (i.e., $+b$) or cessation of frication or aspiration noise (i.e., $-b$)
$\pm s$	Syllabic	At least three of five frequency bands show simultaneous power increases/decreases of at least 6 dB in both the finely smoothed and the coarsely smoothed contours in a voiced segment (between $+g$ and the next $-g$)	Closure or release of a nasal or /l/
$\pm f$	Unvoiced frication	At least three of five frequency bands show simultaneous 6 dB power increases/decreases at high frequencies and decreases/increases at low frequencies (unvoiced segment)	Onset/offset of an unvoiced fricative
$\pm v$	Voiced frication	At least three of five frequency bands show simultaneous 6 dB power increases/decreases at high frequencies and decreases/increases at low frequencies (voiced segment)	Onset/offset of a voiced fricative

[1] The descriptions of the rules are from [24].

For the following reasons, landmark-based acoustic analysis, by using SpeechMark©, would be particularly informative for researchers, pediatricians and language therapists to quantify the developmental trajectory of consonants produced by Mandarin-acquiring children. First, infants and young children's productions are quasiphonetic [27] or are protophones [28,29]. That is, their productions might or might not have clear vowels and consonants and could not be sensibly transcribed with the symbols found in the International Phonetic Alphabet. Additionally, human listeners perceive sounds in the categorical fashion [30]. Therefore, categorizing and studying children's productions with the articulator-bounded method might risk incorporating children's protophones into adults' existing sound inventories and failing to reliably represent those children's productions. For instance, Zhu and Dood [31] and Zhu [32] study the consonantal acquisition of Mandarin-acquiring children by inviting human judges to transcribe those children's word productions. The results indicate that the voiceless alveolo-palatal fricative/ɕ/and the voiceless alveolar fricative/s/are acquired sometime before those children are three years old and four and half years old, respectively. However, by studying the same segments with acoustical analyses, Li and Munson [33] show that the spectral energy distribution and the values of the second formant onset frequency of the following vowels generated by five-year-olds (the oldest age group in the study) are still different from those produced by adults. This shows that the adoption of acoustical methods in child language research is essential and could enable researchers to analyze and compare children's speech without associating their quasiphonetic productions with a transcriber's mental phonetic inventory.

Second, an objective and reliable reference of the consonant developmental trajectory in Mandarin Chinese provides a significant contribution in clinical settings. Although many studies have investigated how Mandarin-acquiring children acquired consonants, those studies set different correction rates, including 70, 75, and 90%, as the criteria for acquisition (c.f. [31,32,34–45]; and also [46] for a relevant literature review). That is, once the correction rate of a certain segment produced by a child is higher than the predetermined percentage (i.e., 70, 75, or 90% depending on independent studies), the acquisition of the segment is said to be completed. The differences in the criteria of acquisition results in the inconsistent order of consonantal acquisition reported in the literature. For instance, while some studies claim that the voiceless labiodental fricative /f/ is acquired later than the voiceless alveolar fricative /s/ [35,36], other studies claim that the segment /f/ is acquired earlier than the segment /s/ [31,32]. With the inconsistency of the order of the segmental acquisition, pediatricians and language therapists do not have a reliable and valid reference from the typical population when they assess the speech development of a potentially high-risk individual or when they wish to evaluate the progress in speech development of a particular atypical population. By using landmark-based acoustic analysis and the program SpeechMark©, the developmental trajectories of Mandarin consonants could be efficiently and reliably quantified, which in turn serve as essential references in clinical settings.

As the first step toward the clinical application of landmark-based acoustic analysis in child Mandarin, the purpose of the current study is to quantify the consonantal productions from four-to-seven-year-old Mandarin-acquiring TD children in Taiwan by using landmark-based acoustic analysis. Furthermore, the relationship between landmark features and speech intelligibility is explored. Children ranging from four to seven are selected because the literature shows that most consonants in Mandarin Chinese are acquired before five years old, with some fricatives and affricates being acquired after the age of six (c.f. [39,42–45]). The study is significant in the following aspects. First, based on the review in this section, it is clearly shown that an objective, efficient and reliable reference of the consonantal development of child Mandarin is in great need. The program SpeechMark©, based on the landmark-based theory proposed by Stevens [11–15], might be particularly informative in this respect. Second, the results could be used to test the prediction from the Biological Model of phonetic/phonological development proposed by Kent [27] and Locke [47]. Specifically, the Biological Model claims that the order of the segmental acquisition could be predicted based on the complexity of the speech motor control ability required in the articulation. According to the model, the articulatory gestures for producing fricatives and affricates require higher physiological demand in speech motor control ability. Therefore, those segments are said to be acquired sometime after children are six years old. Based on the model, it is predicted that the $\pm b$ and $\pm f$ features might be the informative indices, and the oldest age groups would demonstrate differences in these two features. Finally, the results of the study could shed light on the relationship between the landmark features and speech intelligibility.

2. Methods

2.1. Participants

Speech samples from 80 participants were included in the analysis. Table 2 summarizes the demographics of the participants. The author actively contacted and visited several kindergartens and elementary schools. After knowing the purpose, the methodology and the inclusion criteria for the participants, the chairpersons or class advisors of the institutions helped first screen the potential participants in the institutions. Specifically, the inclusion criteria required that all the participants acquired Taiwanese Mandarin as their first language and, according to the teachers and those children's caretakers, did not have language-, learning- or hearing-related disorders. After that, the teachers at the institutions contacted the parent(s) or the caregiver of the potential participants.

One of the parents of each participant was required to sign the consent form so that the experimenters could invite the child participants to join the recitation task individually at the kindergarten/elementary school they attended.

Table 2. Demographics of the participants in the current study.

Participants	Total Participants (Number of Girls)	Mean Age in Month (*SD*) [1]
4-year-olds	20 (10)	52.25 (2.552)
5-year-olds	20 (10)	64.20 (2.821)
6-year-olds	20 (10)	76.20 (2.353)
7-year-olds	20 (10)	88.20 (2.238)

[1] *SD* stands for standard deviation.

2.2. Equipment, Procedures and Materials

The unidirectional microphone RODE (NTG3B) was linked to the interface Babyface Pro, which was linked to the DELL Inspiron 15-5570 laptop. The same laptop was used to display the pictures used to elicit the participants' productions. Praat [48] was the software used to record the speech productions from the participants. The sampling rate was set at 44.1 kHz. All the devices had been settled in a quiet room before the experiment formally started. As the unidirectional microphone was used, the ambient noise, if any, could be minimized or eliminated while recording.

A trained experimenter conducted all the recordings. When a child participant entered the quiet classroom in the kindergarten or the elementary school he/she attended, the experimenter invited the child to sit in front of the laptop. After that, the experimenter first verbally interacted with the child with the unidirectional microphone so that the child could be familiar with speaking to the microphone. When the experiment formally started, the experimenter invited the child participant to name the picture they saw. When the participants failed to produce the target word, the experimenter would recite the correct word and invited the child to repeat it. The microphone was held by the experimenter and he would constantly pay attention to the distance between the microphone and the participant's lips. When children's productions overlapped with noise (e.g., the bell ring at the elementary school), the experimenter would invite the child participant to reproduce those words again. After each participant completed the recitation task, he/she could choose three cartoon stickers as rewards.

Ten disyllabic words were included in the analysis and are listed in the Appendix A. The data were collected based on the two projects conducted by the author. As the contents and the length of the word lists used to elicit productions differ among different age groups, the 10 words that were shared among these age groups were included in the analysis.

2.3. Data Analysis

2.3.1. Landmark-Based Acoustic Analysis

One trained assistant first screened the collected sound files and edited them so that the irrelevant sounds (e.g., the sounds from the experimenter and the disyllabic words that were not included in this study) could be deleted. The author double-checked the resulting edited sound files to make sure that all and only the 10 critical disyllabic words were included. After that, the same trained assistant ran the program SpeechMark© (WaveSurfer Plug-in, Windows Edition, Version 1.0.39) to generate the acoustic landmarks for each participant. The "infants" option was selected so that the range of fundamental frequency in the analysis was adjusted to the range from 1200 to 8000 Hz [24]. A custom-written program was used to automatically sum up the number of instances of each landmark symbol.

2.3.2. Intelligibility Scores

A full-time licensed language therapist with more than 17 years of experience in practice was invited to provide the intelligibility score for each participant. The language therapist did not know the purpose of the study and the data presented to her were randomized. A 5-point Likert scale was adopted where a score from 1 to 5 represents that the speech productions were completely unintelligible (1), mostly unintelligible (2), somewhat intelligible (3), mostly intelligible (4), and completely intelligible (5), respectively. The language therapist gave a score to each disyllabic word production, and the final intelligibility score of each participant was the average score from his/her 10 productions.

2.3.3. Statistical Analysis

Kruskal–Wallis H Test, the non-parametric equivalency of one-way ANOVA, was used to explore if there were differences in the total number of landmark features and within each landmark type among the four age groups. The total number of each landmark was the dependent variable, and the age was the independent variable. IBM SPSS Statistics Version 26.0 was the software used to run the statistical tests. Two notes are appropriate here. First, the landmark features whose total instances were fewer than 80 were not included for statistical analyses. As 80 participants were included in the current study, a landmark feature with a total number of instances less than 80 implies that on average each of the participants generated the feature less than once in the speech sample. In this case, the specific landmark feature was not sensitive enough to detect the speech signals produced by the participants and would not be able to inform us much about the developmental trends of the specific aspects of those children's consonantal productions. Second, as this study is the first study to analyze Mandarin-acquiring children's consonantal development by using the acoustic landmark analysis, increased risk of Type 1 errors was considered less of a concern than Type 2 errors. Therefore, the significant alpha value was set at 0.05. However, when there was a main effect for a specific landmark feature, six specific post hoc comparisons (age 7 vs. age 6, age 7 vs. age 5, age 7 vs. age 4, age 6 vs. age 5, age 6 vs. age 4, and age 5 vs. age 4) were computed to investigate if there were any differences in each landmark feature across different age groups by using Mann–Whitney U Test. In this case, the Bonferroni correction method was adopted, and the p value was set at 0.008 (i.e., 0.05/6).

A multiple regression was run using IBM SPSS Statistics Version 26.0 software to investigate how much of the variation in speech intelligibility scores could be explained by the landmark features. The dependent and independent variables were the individual participants' speech intelligibility scores and the numbers of each landmark feature, respectively. As there was only one test for the regression analysis, the p value was set at 0.05.

3. Results

3.1. Descriptive Results

The results of the landmark-based acoustic analysis and the intelligibility scores were summarized in Tables 3 and 4, respectively. According to the standard described in Section 2.3.3, four landmark features, including $+f$, $-f$, $+v$, and $-v$, were excluded from the later statistical analyses. In terms of the speech intelligibility scores, the four age groups demonstrated highly intelligible speech productions.

3.2. Inferential Results

Nine Kruskal–Wallis H Tests were performed to explore whether there were any differences in the number of landmark features (total landmarks without $\pm f$ & $\pm v$, $+g$, $-g$, $+p$, $-p$, $+b$, $-b$, $+s$, and $-s$) among different age groups. The results showed that there was a statistically significant difference in the number of $+b$ landmarks among different age groups, $\chi^2(3) = 14.07$, $p = 0.003$. No other comparisons were statistically significant. Six post hoc tests, using Mann–Whitney U Test, were performed to compare the number of $+b$ landmarks produced by age 7 vs. age 6, age 7 vs. age 5, age 7 vs. age 4, age 6 vs. age 5, age 6 vs. age 4, and age 5 vs. age 4. The results indicated that the differences among three comparisons were statistically significant (age 7 vs. age 6: $U = 73.5$, $z = -3.46$, $p = 0.001$; age 7 vs. age 5: $U = 103$, $z = -2.64$, $p = 0.008$; age 7 vs. age 4: $U = 92.5$, $z = -2.92$, $p = 0.003$). In short, the results from the statistical analyses revealed that, except for $+b$, the differences in the numbers of the landmark features produced among 7-year-olds, 6-year-olds, 5-year-olds and 4-year-olds were not statistically significant. Seven-year-olds produced fewer +*b* acoustic landmarks than did other age groups.

Table 3. Mean and standard deviation (in parentheses) of the landmark features produced by each age group.

Landmark Features	Age 4 (n = 20)	Age 5 (n = 20)	Age 6 (n = 20)	Age 7 (n = 20)
$+g$	19.25 (2.79)	18.55 (2.54)	18.4 (1.7)	18.9 (3.49)
$-g$	19.25 (2.79)	18.5 (2.48)	18.45 (1.73)	18.85 (3.5)
$+p$	26.9 (7.82)	23.95 (4.71)	24.8 (5.03)	25.55 (6.33)
$-p$	24.6 (5.753)	22.35 (3.56)	23.3 (4.14)	24 (5.54)
$+b$	10.05 (3.33)	9.75 (3.73)	10.5 (3.17)	7 (3.1)
$-b$	3.45 (2.31)	2.75 (1.68)	2.8 (1.58)	2.8 (2.09)
$+s$	5.55 (3.09)	5.2 (3.3)	4.7 (3.23)	4.95 (3.4)
$-s$	5.3 (2.96)	5.55 (3.1)	5.4 (2.8)	4.6 (4.31)
$+f$	0.15 (0.49)	0 (0)	0 (0)	0 (0)
$-f$	0.1 (0.31)	0.05 (0.224)	0.15 (0.366)	0 (0)
$+v$	0.05 (0.22)	0 (0)	0.1 (0.31)	0.05 (0.22)
$-v$	0.45 (0.83)	0.25 (0.55)	0.05 (0.22)	0.15 (0.366)
Total	115.1 (17.47)	106.9 (11.35)	108.65 (12.87)	106.85 (16.6)
Total without $\pm f$ & $\pm v$	114.35 (17.37)	106.6 (11.39)	108.35 (12.93)	106.65 (16.69)

Table 4. Mean and standard deviation (in parentheses) of the intelligibility scores produced by each age group.

Age 4 (n = 20)	Age 5 (n = 20)	Age 6 (n = 20)	Age 7 (n = 20)
4.825 (0.259)	4.905 (0.267)	4.855 (0.305)	4.975 (0.079)

A multiple regression analysis was performed in order to investigate how much of the variation in speech intelligibility scores could be explained by the landmark features. The results showed that these landmark features statistically significantly predicted speech intelligibility scores, F (8, 71) = 2.405, p = 0.023, R^2 = 0.213. That is, 21.3% of the total variation in speech intelligibility scores could be accounted for by all the eight landmark features (excluding $+f$, $-f$, $+v$, and $-v$). The landmark feature that added statistically significantly to the prediction was the $+b$ feature (p = 0.0002, B = -0.031). For every one point increase in the number of the $+b$ feature, speech intelligibility scores would be expected to decrease by 0.031 point.

4. Discussion

As the first step toward the clinical application of landmark-based acoustic analysis in child Mandarin, this study was designed to quantify the consonantal productions from Mandarin-acquiring children in Taiwan by using landmark-based acoustic analysis. Furthermore, the relationship between the landmark features and speech intelligibility scores was explored. The disyllabic word productions from 80 children (from four to seven years old) were collected and analyzed by using the program SpeechMark©. The results indicated that seven-year-olds produced statistically significantly fewer $+b$ landmark features than did other age groups. No other statistically significant differences were found among these children's productions. Additionally, all the participants hardly generated $\pm f$ and $\pm v$ landmark features. The results from a multiple regression analysis indicated that the eight landmark features (excluding $+f$, $-f$, $+v$, and $-v$) could statistically significantly account for 21.3% of the total variation in speech intelligibility scores. The net effect of the landmark feature $+b$ was the strongest. For every one point increase in the number of the $+b$ feature, speech intelligibility scores would be expected to decrease by 0.031 point. Based on the obtained results, several issues are discussed.

Three landmark features, *+b*, $\pm f$ and $\pm v$, are first discussed below. First, the presence of the *+b* landmark represented the presence of bursts among obstruents [21,22]. The presence and absence of bursts had been consistently reported to be an essential indicator in speech intelligibility in both English and Chinese [49,50]. Empirical studies also demonstrated that TD young children produced more bursts for affricates than for fricatives [51]. That was an expected phenomenon as a release burst was expected for the first half stop in an affricate (e.g., the/t/in/ts/). Additionally, literature generally agreed that even the consonants with the more complicated articulatory gestures (e.g., the voiceless retroflex fricative/ʂ/) were virtually mastered around six years old [39,42–45]. Therefore, the fact that younger children (four to six years old) produced too many *+b* landmark features indicated that their finer-grained ability to properly control the speech motor was still developing and that progress was observed by the time children reached seven years old. Second, the scarcity of the two speech landmarks, $\pm f$ and $\pm v$, deserves some attention. According to Huang, Epps and Joachim [25] and Ishikawa et al. [22], $\pm f$ is an indicator of the onset/offset of voiceless fricatives while $\pm v$ is an indicator of the onset/offset of voiced fricatives. As all the six fricatives in Taiwanese Mandarin (i.e., /x, ɕ, s, ʂ, zʈ/) are voiceless, the scarcity of the landmark feature $\pm v$ is understandable. At first glance, the scarcity of the landmark feature $\pm f$ might suggest that even the oldest children in the experiment might not be able to properly produce fricatives. However, a closer look at the acoustic rules in Table 1 and the existing literature might reveal a different picture. In fact, the $\pm f$ and $\pm b$ landmark features partially share acoustic rules (i.e., at least three of five frequency bands show simultaneous power increases/decreases of at least 6 dB). Nevertheless, $\pm f$ (and also $\pm v$) further required the lower frequency bands to simultaneously decrease (or increase) when the higher frequency band showed power increases (or decreases). As the acoustic rules for $\pm f$ were more complicated and might be designed to detect a very rare case of fricatives, it was not surprising to learn that the $\pm f$ landmark features were scarce, if not totally unavailable, even among TD adult speakers in Chinese [52] and English [19,22]. These phenomena also indicated that the addition or modification of the landmarks that could be used to distinguish among stops, affricates and fricatives would be especially informative. The redefinition of the acoustic rules for detecting the landmark features $\pm f$ and $\pm v$ might be a solution. At first, the rules for detecting $\pm f$ and $\pm v$ must be less complicated so that fricatives in general could be detected. Second, as fricatives are the only obstruents that are produced without a stop burst, the acoustic rules for *+f* and *+v* should specify the timing of the detection. Specifically, the landmark features *+f* and *+v* could only be detected without a preceding *+b* feature within a certain time domain. In short, the *+b* landmark features, but not the $\pm f$ features, are more sensitive to the quality of the voiceless obstruents produced by Mandarin speakers. The redefinition of the acoustic rules for the landmark features $\pm f$ and $\pm v$ is required in order for the analysis to precisely determine the differences among different obstruents.

The current findings also lent strong support to the Biological Model of phonetic/phonological development proposed by Kent [27] and Locke [47]. According to the model, the phonetic/phonological development of children was substantially affected by their speech motor control ability. Based on Kent [27], consonants involving the fine force control to generate frication (i.e., fricatives and affricates) were acquired the latest, and the completion of the acquisition was sometime after children were six years old. The current findings matched the developmental trajectory predicted by the Biological Model. According to the experimental results, by age seven, Mandarin-acquiring children had made progress in the production of the *+b* landmark features (i.e., producing fewer *+b* landmark features). Please recall that the *+b* landmark features indicated the presence of a burst consonant. The higher number of *+b* landmark features among younger age groups implied that those children were more likely to generate stop bursts for segments even when such stop bursts were not expected (i.e., for fricatives). This phenomenon in turn showed that four-to-six-year-olds were less likely to properly generate the fine force regulation for frication and therefore produced the bursts that were not supposed to be present. In short, the current findings showed that the landmark feature +*b* is particularly sensitive to those children's speech motor control ability.

A note about the relationship between landmark features and speech intelligibility is appropriate here. Boyce et al. [18] showed that speakers with clearer speech produced a higher number of landmark features. Similarly, Ishikawa, MacAuslan and Boyce [19] hypothesized that the greater number of landmark features produced by female speakers in their study might indicate greater intelligibility of their speech. However, according to the current findings from the landmark feature +*b*, the relationship between the number of landmark features and the degree of speech intelligibility might not always be "the more, the better". According to the results from the multiple regression in the current study, the increase of the landmark feature +*b* resulted in the decrease in the speech intelligibility scores. Similar patterns were also reported in Ishikawa et al. [22]. In their study, Ishikawa et al. [22] compared the acoustic landmark features produced by dysphonic speakers and TD speakers. The findings indicated that the speakers from the clinical group produced a statistically significantly higher number of $\pm g$ and $\pm b$ features than did the control group. If the higher number of the landmark features indicated speakers' better speech intelligibility, it was hard to justify why those individuals with dysphonic speech produced higher numbers of the landmark features $\pm g$ and $\pm b$. Therefore, the findings from the literature and the current study suggest that, with regard to landmark features, "the more, the better" is inaccurate. Rather, it is more accurate to say that too many and too few acoustic landmark features would exert equally negative influences on speech intelligibility. As different languages encompass different segmental inventories, the critical landmark features that are strongly related to speech intelligibility might vary from language to language. Future studies focusing on languages other than Taiwanese Mandarin are suggested to directly explore the relationship between each of the acoustic landmark features and speech intelligibility so that researchers could identify the key landmark features that could account for the variation of speech intelligibility in the particular language.

The landmark-based acoustic analysis reported in the current study could be practically applied to several domains. First, as the landmark feature +*b* reflects the Mandarin-acquiring children's speech motor control ability and exerts influences on speech intelligibility, future clinical applications of the analysis should focus on the quantity of the +*b* landmark feature that Mandarin-acquiring children with speech related disorders generate in their word productions. Second, it has been reported that aging people and elderly people with Parkinson's disease generally have decreased speech motor control ability and lower speech intelligibility. In this case, for Mandarin-speaking adults, it is expected that seniors would produce more +*b* landmark features than did the younger generations. In addition, those Mandarin-speaking individuals with Parkinson's disease might also produce more +*b* landmark features than did their TD counterparts. In short, the application of landmark-based acoustic analysis to various TD or disordered populations would inform us about the nature of those individuals' speech motor control ability.

5. Conclusions

By using landmark-based acoustic analysis, the current study quantified the consonantal developments among children ranging in age from four to seven years. The results of the disyllabic word recitation task indicated that the younger children (four, five and six-year-olds) produced a significantly higher number of the +*b* landmark features than did the seven-year-olds. In addition, the number of the +*b* landmark features were negatively correlated with the participant's speech intelligibility scores. The experimental results could be elegantly accounted for by the Biological Model of children's phonetic/phonological development [27,47], which claimed that consonants requiring finer-grained speech motor control ability were acquired sometime after the age of six. Additionally, based on the literature and the current study, it could be concluded that the relationship between the number of landmark features and speech intelligibility is not always "the more, the better".

Instead, too many and too few acoustic landmark features would exert equally negative influences on speech intelligibility. Additionally, the acoustical rules for detecting the landmark features $\pm f$ and $\pm v$ should be refined so that the distinctions among obstruents could be more precisely identified. Pediatricians and language therapists are encouraged to apply landmark-based acoustic analysis in clinical sessions, and the findings from the TD children presented in the current study could serve as essential references for the Mandarin-acquiring population.

Funding: This research was funded by Ministry of Science and Technology in Taiwan, grant numbers MOST 107-2410-H-012-002 and MOST 108-2410-H-167-013. The APC was funded by the Ministry of Science and Technology in Taiwan, grant number MOST 108-2410-H-167-013 and the Higher Education Sprout Project, Ministry of Education, Taiwan.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Central Regional Research Ethics Committee of the China Medical University (CRREC-106-097, 02/21/2018 and CRREC-108-056, 06/19/2019).

Informed Consent Statement: Informed consent was obtained from all participants (or one of each participant's parents) involved in the study.

Data Availability Statement: The data presented in this study are available on request from the author. The data are not publicly available due to original informed consent provisions.

Acknowledgments: The author sincerely appreciates the support from the families of the children and the teachers of the kindergartens and elementary schools in this study. For the kind advice, invaluable suggestions and highly constructive comments about the earlier version of the manuscript, the author thanks the two anonymous reviewers of *Children*. For the help he has received at the various stage of the project, the author extends his gratitude to Eller Chad, Shiao-Wei Chu, Chun-Yi Lin and Hui-Ping Lu.

Conflicts of Interest: The author declares no conflict of interest.

Appendix A

The 10 disyllabic words that were analyzed in the study. Please note that the participants saw pictures of the object and did not read the contents below.

Table A1. Ten disyllabic words included in the present study.

No.	Chinese Characters	Transliteration (Pinyin)	Gloss
1.	鳳梨	fèng lí	Pineapple
2.	飛機	fēi jī	Airplane
3.	火車	huǒ chē	Train
4.	漢堡	hàn bǎo	Hamburger
5.	蝦子	xiā zi	Shrimp
6.	小鳥	xiǎo niǎo	Bird
7.	森林	sēn lín	forest
8.	松鼠	sōng shǔ	squirrel
9.	薯條	shǔ tiáo	french fries
10.	手錶	shǒu biǎo	watch

References

1. Chen, L.-M.; Oller, D.K.; Lee, C.C.; Liu, C.-T. LENA: Computerized Automatic Analysis of Speech Development from Birth to Three. In Proceedings of the 30th Conference on Computational Linguistics and Speech Processing (ROCLING 2018), Hsinchu, Taiwan, 4–5 October 2018.
2. Holmgren, K.; Lindblom, B.; Aurelius, G.; Jailing, B.; Zetterström, R. On the phonetics of infant vocalization. In *Precursors of Early Speech*; Lindblom, B., Zetterström, R., Eds.; Palgrave Macmillan: London, UK, 1986; pp. 51–63.
3. Jones, G.; Nadjibzadeh, N.; Károly, L.; Mohammadpour, M. An integrated dialect analysis tool using phonetics and acoustics. *Lingua* **2019**, *221*, 37–48. [CrossRef]

4. Khan, A.; Steiner, I.; Sugano, Y.; Bulling, A.; Macdonald, R. A multimodal corpus of expert gaze and behavior during phonetic segmentation tasks. In Proceedings of the Language Resources and Evaluation Conference (LREC), Miyazaki, Japan, 7–12 May 2018.
5. Oller, D.K.; Niyogi, P.; Gray, S.; Richards, J.A.; Gilkerson, J.; Xu, D.; Yapanel, U.; Warren, S.F. Automated Vocal Analysis of Naturalistic Recordings from Children with Autism, Language Delay, and Typical Development. *Proc. Nat. Acad. Sci. USA* **2010**, *107*, 13354–13359. [CrossRef] [PubMed]
6. Xu, D.; Richards, J.A.; Gilkerson, J. Automated analysis of child phonetic production using naturalistic recordings. *J. Speech Lang. Hear. Res.* **2014**, *57*, 1638–1650. [CrossRef]
7. Liu, C.-T.; Chen, L.-M.; Lin, Y.-C.; Cheng, C.-Y.; Lin, Y.-C. Fricative productions of Mandarin-speaking children with cerebral palsy: The case of five-year-olds. *Clin. Linguist. Phonet.* **2020**, *34*, 256–270. [CrossRef] [PubMed]
8. Boyce, S.; Fell, H.J.; McAuslan, J. SpeechMark: Landmark detection tool for speech analysis. In Proceedings of the Interspeech 2012, Portland, OR, USA, 9–13 September 2012.
9. Liu, S.A. Landmark detection for distinctive feature-based speech recognition. *J. Acoust. Soc. Am.* **1996**, *100*, 3417–3430. [CrossRef]
10. Howitt, A.W. Automatic Syllable Detection for Vowel Landmarks. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 2000.
11. Stevens, K.N. Evidence or the role of acoustic boundaries in the perception of speech sounds. In *Phonetic Linguistics: Essays in Honor of Peter Ladefoged*; Fromkin, V.A., Ed.; Academic Press: London, UK, 1985; pp. 243–255.
12. Stevens, K.N. On the quantal nature of speech. *J. Phon.* **1989**, *17*, 3–46. [CrossRef]
13. Stevens, K.N. Diverse acoustic cues at consonantal landmarks. *Phonetica* **2000**, *57*, 139–151. [CrossRef] [PubMed]
14. Stevens, K.N. From Acoustic Cues to Segments, Features and Words. In Proceedings of the 6th International Conference on Spoken Language Processing (ICSLP 2000), Beijing, China, 16–20 October 2000.
15. Stevens, K.N. Toward a model for lexical access based on acoustic landmarks and distinctive features. *J. Acoust. Soc. Am.* **2002**, *111*, 1872–1891. [CrossRef]
16. Chomsky, N.; Halle, M. *The Sound Pattern of English*; Harper and Row: New York, NY, USA, 1968.
17. Keyser, S.J.; Stevens, K.N. Feature geometry and the vocal tract. *Phonol* **1994**, *11*, 207–236. [CrossRef]
18. Boyce, S.; Krause, J.; Hamilton, S.; Smilijanic, R.; Bradlow, A.R.; Rivera-Campos, A.; MacAuslan, J. Using Landmark Detection to Measure Effective Clear Speech. In Proceedings of the Meetings on Acoustics, Montreal, QC, Canada, 2–7 June 2013.
19. Ishikawa, K.; MacAuslan, J.; Boyce, S. Toward clinical application of landmark-based speech analysis: Landmark expression in normal adult speech. *J. Acoust. Soc. Am.* **2017**, *142*, EL441–EL447. [CrossRef]
20. Kalita, S.; Mahadeva Prasanna, S.R.; Dandapat, S. Importance of glottis landmarks for the assessment of cleft lip and palate speech intelligibility. *J. Acoust. Soc. Am.* **2018**, *144*, 2656–2661. [CrossRef]
21. Atkins, M.S.; Boyce, S.E.; MacAuslan, J.; Silbert, N. Computer-assisted Syllable Complexity Analysis of Continuous Speech as a Measure of Child Speech Disorders. In Proceedings of the 19th International Congress of Phonetic Sciences, (ICPhS 2019), Melbourne, Australia, 4–10 August 2019.
22. Ishikawa, K.; Rao, M.B.; MacAuslan, J.; Boyce, S. Application of a landmark-based method for acoustic analysis of dysphonic speech. *J. Voice* **2020**, *34*, 645.e11–645.e18. [CrossRef] [PubMed]
23. DiCicco, T.; Patel, R. Automatic landmark analysis of dysarthric speech. *J. Med. Speech-Lang. Path.* **2008**, *16*, 213–219.
24. MacAuslan, J. What are Acoustic Landmarks, and What do They Describe? Available online: https://speechmrk.com/wp-content/uploads/2016/08/Landmark-Descriptions.pdf (accessed on 17 January 2021).
25. Huang, Z.; Epps, J.; Joachim, D. Investigation of Speech Landmark Patterns for Depression Detection. In Proceedings of the 2020 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP 2020), Virtual Conference, Barcelona, Spain, 4–8 May 2020.
26. Atkins, M.S.; Washington, K.; Silbert, N.; MacAuslan, J.; Tuhohy, S.; Blades, R.; Donaldson, M.; Ungruhe, J.; Swanson, K. Towards Automated Detection of Similarities and Differences in Bilingual Speakers. In Proceedings of the 174th Meeting of the Acoustical Society of America, Louisiana, LA, USA, 4–8 December 2017.
27. Kent, R. The biology of phonological development. In *Phonological Development: Models, Research, Implications*; Ferguson, C.A., Menn, L., Stoel-Gammon, C., Eds.; York Press: Maryland, MD, USA, 1992; pp. 65–90.
28. Nathani, S.; Oller, D.K. Beyond ba-ba and gu-gu: Challenges and strategies in coding infant vocalizations. *Behav. Res. Methods Insr. Cmp.* **2001**, *33*, 321–330. [CrossRef]
29. Oller, D.K.; Ramsdell, H.L. A weighted reliability measure for phonetic transcription. *J. Speech Lang. Hear. Res.* **2006**, *49*, 1391–1411. [CrossRef]
30. Liberman, A.M.; Harris, K.S.; Kinney, J.A.; Lane, H. The discrimination of relative onset-time of the components of certain speech and nonspeech patterns. *J. Exp. Psychol.* **1961**, *61*, 379–388. [CrossRef] [PubMed]
31. Zhu, H.; Dodd, B. The phonological acquisition of Putonghua (modern standard Chinese). *J. Child Lang.* **2000**, *27*, 3–42.
32. Zhu, H. *Phonological Development in Specific Contexts: Studies of Chinese-Speaking Children*; Multilingual Matters: Clevedon, UK, 2002.
33. Li, F.; Munson, B. The development of voiceless sibilant fricatives in Putonghua-speaking children. *J. Speech Lang. Hear. Res.* **2016**, *59*, 699–712. [CrossRef]
34. Chao, Y.-R. The Cantian idiolect. *Univ. Calif. Pub. Semitic Philol.* **1951**, *2*, 27–44.

35. Jeng, H. The Acquisition of Chinese Phonology in Relation to Jakobson's Laws of Irreversible Solidarity. In Proceedings of the 9th International Congress of Phonetic Sciences (ICPhS 1979), Copenhagen, Denmark, 6–11 August 1979.
36. Jeng, H. A developmentalist view of child phonology. *Studies Lang. Lit.* **1985**, *1*, 1–29.
37. Wang, N.-M.; Fei, P.; Huang, H.; Chen, C.-W. The phonetic development of Mandarin-acquiring, three-to-six preschool children. *J. Speech-Lang.-Hear. Assoc.* **1984**, *1*, 12–15. (In Chinese)
38. Chang, C.-F.; Chung, Y.-M. Revision and application of preschool children language development scale. *Bull. Spec. Edu.* **1986**, *2*, 37–52. (In Chinese)
39. Hsu, J. A Study of the Various Stages of Development and Acquisition of Mandarin Chinese by Children in Taiwan Milieu. Master's Thesis, Fu Jen Catholic University, New Taipei, Taiwan, 1987.
40. Shiu, H. The Phonological Acquisition by Mandarin-Speaking Children: A Longitudinal Case Study on Children from 9 Months through Three Years Old. Master's Thesis, National Taiwan Normal University, Taipei, Taiwan, 1990.
41. Cheung, H. Three to four-years old children's perception and production of Mandarin consonants. *Lang. Linguist.* **2000**, *1*, 19–38. (In Chinese)
42. Cheung, H.; Hsu, P.-H. Chinese children's production and perception of consonants: A developmental study. *J. Chin. Hear. Speech* **2000**, *15*, 1–10. (In Chinese)
43. Cho, S.-C. The Phonological Development of 3 to 6 Year-Old Preschool Children in Taiwan. Master's Thesis, National Taipei University of Nursing and Health Sciences, Taipei, Taiwan, 2008. (In Chinese)
44. Jeng, J.-Y. Manual of Mandarin Speech Test for Children. Available online: http://giast.nknu.edu.tw/UploadFile/TeaFiles/giast_t_103205415.pdf (accessed on 17 January 2021). (In Chinese)
45. Jeng, J.-Y. The speech acquisition of Mandarin-speaking preschool children. *J. Chin. Lang. Teach.* **2017**, *14*, 109–136. (In Chinese)
46. Li, X.X.; To, C.K. A review of phonological development of Mandarin-speaking children. *Amer. J. Speech-Lang. Pathol.* **2017**, *26*, 1262–1278. [CrossRef] [PubMed]
47. Locke, J. *Phonological Acquisition and Change*; Academic Press: New York, NY, USA, 1983.
48. Boersma, P.; Weenink, D. Praat: Doing Phonetics by Computer [Computer program]. Available online: http://www.praat.org/ (accessed on 17 January 2021).
49. Ansel, B.M.; Kent, R.D. Acoustic-phonetic contrasts and intelligibility in the dysarthria associated with mixed cerebral palsy. *J. Speech Lang. Hear. Res.* **1992**, *35*, 296–308. [CrossRef]
50. Liu, H.-M.; Tseng, C.-H.; Tsao, F.-M. Perceptual and acoustic analysis of speech intelligibility in Mandarin-speaking young adults with cerebral palsy. *Clin. Linguist. Phonet.* **2000**, *14*, 447–464.
51. Liu, C.-T.J.; Chen, L.-M.; Lin, Y.-C.; Cheng, C.-F.A.; Chang, H.-C.J. Speech Intelligibility and the Production of Fricative and Affricate among Mandarin-Speaking Children with Cerebral Palsy. In Proceedings of the 2016 Conference on Computational Linguistics and Speech Processing (ROCLING 2016), Tainan, Taiwan, 6–7 October 2016.
52. Liu, C.-T. Acoustic Landmark Analysis of Adults' Consonants in Mandarin Chinese: The Case of Disyllabic Words. In Proceedings of the Paper Presented at The 56th Linguistics Colloquium, Virtual Conference, 26–28 November 2020.

children

Review

The Value of Non-Referential Gestures: A Systematic Review of Their Cognitive and Linguistic Effects in Children's Language Development

Ingrid Vilà-Giménez [1,2,*] and Pilar Prieto [1,3]

1 Department of Translation and Language Sciences, Universitat Pompeu Fabra, 08018 Barcelona, Spain; pilar.prieto@upf.edu
2 Department of Subject-Specific Education, Universitat de Girona, 17004 Girona, Spain
3 Institució Catalana de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Spain
* Correspondence: ingrid.vila@udg.edu

Citation: Vilà-Giménez, I.; Prieto, P. The Value of Non-Referential Gestures: A Systematic Review of Their Cognitive and Linguistic Effects in Children's Language Development. *Children* **2021**, *8*, 148. https://doi.org/10.3390/children8020148

Academic Editors: Eva Aguilar Mediavilla, Miguel Pérez Pereira, Elisabet Serrat-Sellabona and Daniel Adrover-Roig

Received: 31 December 2020
Accepted: 11 February 2021
Published: 17 February 2021

Abstract: Speakers produce both referential gestures, which depict properties of a referent, and non-referential gestures, which lack semantic content. While a large number of studies have demonstrated the cognitive and linguistic benefits of referential gestures as well as their precursor and predictive role in both typically developing (TD) and non-TD children, less is known about non-referential gestures in cognitive and complex linguistic domains, such as narrative development. This paper is a systematic review and narrative synthesis of the research concerned with assessing the effects of non-referential gestures in such domains. A search of the literature turned up 11 studies, collectively involving 898 2- to 8-year-old TD children. Although they yielded contradictory evidence, pointing to the need for further investigations, the results of the six studies–in which experimental tasks and materials were pragmatically based–revealed that non-referential gestures not only enhance information recall and narrative comprehension but also act as predictors and causal mechanisms for narrative performance. This suggests that their bootstrapping role in language development is due to the fact that they have important discourse–pragmatic functions that help frame discourse. These findings should be of particular interest to teachers and future studies could extend their impact to non-TD children.

Keywords: non-referential gestures; prosody; pragmatics; children; language development; cognitive development; narrative development; information recall; narrative discourse comprehension; oral narrative discourse performance

1. Introduction

Gesture is a powerful embodied form of communication. Apart from their rich communicative value, gestures have been shown to act as cognitive bootstrappers, as they can contribute to changes in children's linguistic knowledge (see [1–3], for reviews). Understanding the relative value of different types of co-speech gestures is crucial to unraveling how gestures pave the way for language development. However, the field of language development has tended to focus on the role of referential gestures, such as deictic or iconic gestures, which imagistically represent the properties of a referent and thus bear a close relationship to the semantic content of the speech. The value of non-referential gestures, such as hand movements which typically associate with prosodically prominent positions in speech but do not encode specific semantic content, has been comparatively neglected. It is the small amount of research on the latter—-non-referential gestures—-that we will systematically review in this article, limiting our focus to typically developing children (henceforth TD), since to our knowledge no studies have been conducted on the role of non-referential gestures in non-TD children. By means of this review, we hope to gain insight into the link between the important discourse–framing properties of non-referential gestures and their potential bootstrapping role in cognitive and language development. An understanding of this relationship will allow us to point to some practical implications for the teaching of TD children, as well as ideas for promoting multimodally-based narrative and pragmatic trainings, and some directions for future research. Future investigations could also extend these findings to non-TD children.

Current research has demonstrated that co-speech gestures are tightly linked to speech production and perception, suggesting that the two modalities are very closely intertwined in creating meaning and make up a well-integrated communicative system (see [4–6], and many others). Previous studies focusing on referential gestures have shown that children's gestures serve as forerunners of future linguistic skills in many populations including not only TD children (e.g., [7–9]; see [10] for a review), but also late talking toddlers (e.g., [11]), and children with an Autism Spectrum Disorder (ASD) diagnosis (e.g., [12–14]; see [15] for a review). It has also been well established by a variety of studies that referential gestures have a positive effect on adults' and children's cognitive and linguistic abilities (see [16] for a meta-analysis review), boosting memory recall, for example, in TD children [17–21].

Typically developing children start producing their first gestures, typically deictic or pointing gestures that identify objects, people, events, or locations, between 9 and 12 months of age and before they produce their first words [22–24]. These gestures help children carry out successful dyadic interactions with their parents and caregivers (see [25], for a review on the development of deictic pointing in infancy). Already in the transition between the babbling stage and single-word period, infants start to semantically and temporally coordinate their pointing-speech combinations, and gesture is used to complement or reinforce speech [26,27]. Deictic declarative gestures have proven to be a reliable predictor of language skills not only in TD [7,28–30], but also in children with speech and language impairments, such as ASD infants [14] (see also [31] for a meta-analysis review). This type of gesture has also been identified as the most markedly impaired in ASD [32–34]. Importantly, the fact that deictic gestures place high social and interactive demands on early interactions is an indication that these gestures constitute a powerful tool for early interventions in ASD programs (see also the positive effects of a pointing gesture intervention in generating larger vocabulary repertoires in TD children by [35]).

At this early stage, TD children also start producing iconic gestures that allow them to represent information about a referent in speech, such as an object, an action, or a space. Early iconic gestures are used to depict actions or attributes associated with objects, such as raising arms to indicate big size or flapping arms to represent a bird flying (see [36]) [22,23,37,38]. At around two years of age, there is a sharp increase in the number of iconic gestures produced (e.g., [39–41]), corresponding with the period in which children also show an increased sensitivity to iconicity in gesture comprehension [42,43]. For instance, a study assessing spontaneous gestures performed by 40 TD children observed from 14 to 34 months of age reported a spurt in iconic gesture production at roughly 26 months, with children past this threshold not only using iconic gestures more frequently but also employing them to convey a more varied set of meanings [40]. Moreover, a longitudinal study by [41] also found an increase in the production of iconic gestures between 22 and 26 months of age, which were usually used to convey action meanings not yet conveyed in the first verbs (e.g., "go like this" + move fisted empty hand in circles as if stirring, p. 9). Other studies have reported that TD children benefit from observing referential iconic gestures in complex linguistic processes, such as narrative comprehension [44,45]. There is also evidence that a specific type of iconic gesture ("character-viewpoint" or CVPT gestures, in which the gesturer takes on the role of a character in a story; see [6]) can serve as the precursor [46] and predictor [47] of more complex narrative abilities undergoing development. Concerning non-TD children, while research shows that young children with ASD produce deictic gestures less often than TD children, empirical evidence about the role of other forms of gesture in ASD language development comes from only one study. The study by [14] explored the gesture-language relation in autistic children by tracking gesture type production (deictic, give, iconic, or conventional) and subsequent language outcomes of 18-month-old TD and 30-month-old ASD children (n = 23). They found that only deictic gestures predicted vocabulary growth in both TD and ASD children, but such gestures occurred at significantly lower prevalence (70% ASD vs. 96% TD) and frequency (45% ASD vs. 60% TD) rates in the ASD group.

Later on, between 2 and 3 years of age, another type of gesture starts to emerge, the non-referential gesture, often called a beat gesture (McNeill [6] describes this type of gesture as being a rhythmically short and quick "simple flick of the hand or fingers up and down, or back and forth" (p. 15) that lacks referentiality and associates with prosodic prominent positions in speech. For this reason, following a McNeillian classification of gestures, many studies have called such gestures "beat" gestures (i.e., as if marking a beat). More recently, this traditional view of non-referentials has been challenged and a more inclusive definition of non-referential gestures has been adopted that emphasizes their rhythmic, pragmatic and discursive properties. Within this view, "beat gestures" are considered as non-referential gestures that do not only act mainly as rhythmic highlighters, but also contribute clear pragmatic and discursive meanings that help frame oral discourse structure [48–51].) [52,53]. The study by [53] documented the appearance of beat gestures in French-English bilingual children in the period from 2 years to 3 years 6 months of age, and observed that the children began to produce these gestures once they were able to perform sentence-like or more linguistically complex spoken utterances (in other words, when the mean length of their utterances increased). The literature on the acquisition of these gestures is sparse and has focused on how children gesture with non-referential gestures while they are narrating. Some studies have shown that non-referential gestures start appearing in complex narrative discourses at around 4–5 to 6 years of age [6,54,55] (see also [56] for specifically language-impaired children). A cross-sectional study with French-speaking children aged 6 and 10, and adults by [54] found that, in contrast to the average number of representational (i.e., iconic) gestures, the average number of non-representational gestures (i.e., non-referential gestures) increased with age and that these gestures served both discursive functions (by accompanying connectors, highlighting important linguistic units, or performing anaphoric functions) and framing functions. Similarly, another cross-sectional study by [57] with 5- and 10-year-old French, American, and Italian children demonstrated that the older children tended to produce more non-referential gestures that helped to structure speech or mark cohesion in discourse than the younger ones. Recently, a longitudinal study assessing multimodal narrative development in children 5–6 to 7–9 years of age revealed that, in contrast with referential iconic gestures, the use of non-referential gestures increased noticeably with age, as narrative skills matured [58] (see also [59] for similar results).

As mentioned above, the literature on multimodal language development has tended to focus on the value of referential gestures in paving the way for early language development in children, with less attention being paid to the precursor and predictive role of non-referential gestures as well as their bootstrapping impact on language development. The obvious question is whether non-referential gestures have the same beneficial role in cognitive and linguistic dimensions as referential (i.e., deictic and iconic) gestures. This research gap was already noted a decade ago in a meta-analysis study conducted by [16], who called for further research to examine the nature and impact of non-referential gestures.

In our view, there are strong reasons to believe that non-referential gestures are important in children's language development. Though the classic McNeillian classification of gestures [6] has highlighted the rhythmic character of non-referential "beat" gestures and their consequent link to prosody, the fact that non-referential gestures have been relatively understudied may be due to the general theoretical claim coming from this view that these gestures lack abstract semantic content. Indeed, many studies have assumed that because these gestures lack referential meaning, their contribution to language learning and development is negligible (e.g., [20,45,60–62], and others). However, it is well known that non-referential gestures in adult discourse mark information structure and focused information, as well as new or accessible referents [62,63]. In this way, non-referential gestures can act as multimodal pragmatic cues which highlight important linguistic functions in speech that help frame complex oral discourse in later stages of language development ([6,50,51,64–67], among others) (e.g., [68]). We thus hypothesize that non-referential gestures are important in the processing and acquisition of more complex language skills such as narratives, as they play an important role in framing discourse.

The present review will systematically assess, evaluate, and compare the available research concerned with the scaffolding role of non-referential gestures in three important areas of child learning and language development, namely information recall, narrative discourse comprehension, and oral narrative discourse performance. With regard to the second and third of these areas, it should be noted that the development of narrative discourse abilities as an oral language skill is an important achievement for children, as it has been typically associated with children's complex linguistic development and successful school literacy [69,70] (see [71] for a review). Simultaneously, we will seek answers for three main research questions related respectively to the association, predictive, and causal effects of non-referential gestures, as follows.

(1) Association effects. Does observing another speaker's non-referential gestures enhance information recall and narrative discourse comprehension in TD children?
(2) Predictive effects. Does the frequency of use of non-referential gestures by TD children predict better narrative production skills later in development?
(3) Causal effects. Can training TD children with non-referential gestures bring about an improvement in narrative production scores in a subsequent posttest?

2. Methods

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement (e.g., [72,73]) guided the methodology and reporting of this systematic review.

2.1. Identification of Studies and Inclusion Criteria

As stated above, the principal aim of this systematic review was to address and compare the available research devoted to the effects of non-referential gestures on children's cognitive and linguistic skills. The process of identifying studies is summarized in the PRISMA flow diagram in Figure 1. First, a comprehensive search strategy was conducted during the second semester of 2020 using four electronic databases: Scopus, Web of Science, PubMed, and PsycInfo. The database searches were limited to English language works dated between 1970 and 2020. Search terms used in the databases included: ("children*" OR "preschooler*") AND ("beat gesture*" OR "non-referential beat gesture*" OR "non-referential gesture*" OR "non-referential beat*") AND ("memory recall" OR "recall" OR "comprehension" OR "narrative comprehension" OR "narrative performance" OR "produc*" OR "observ*"). The reference lists of articles retrieved were screened to identify any relevant additional studies on the topic. Both conference proceedings and papers that were under review were also included in the identification process (one conference proceeding paper was identified via Google Scholar).

Once the literature had been identified, the titles and abstracts of the compiled list of retrieved articles were screened for relevance by the first author and duplicated or irrelevant articles were removed using Mendeley reference software. Articles which warranted further examination were selected for full-text review.

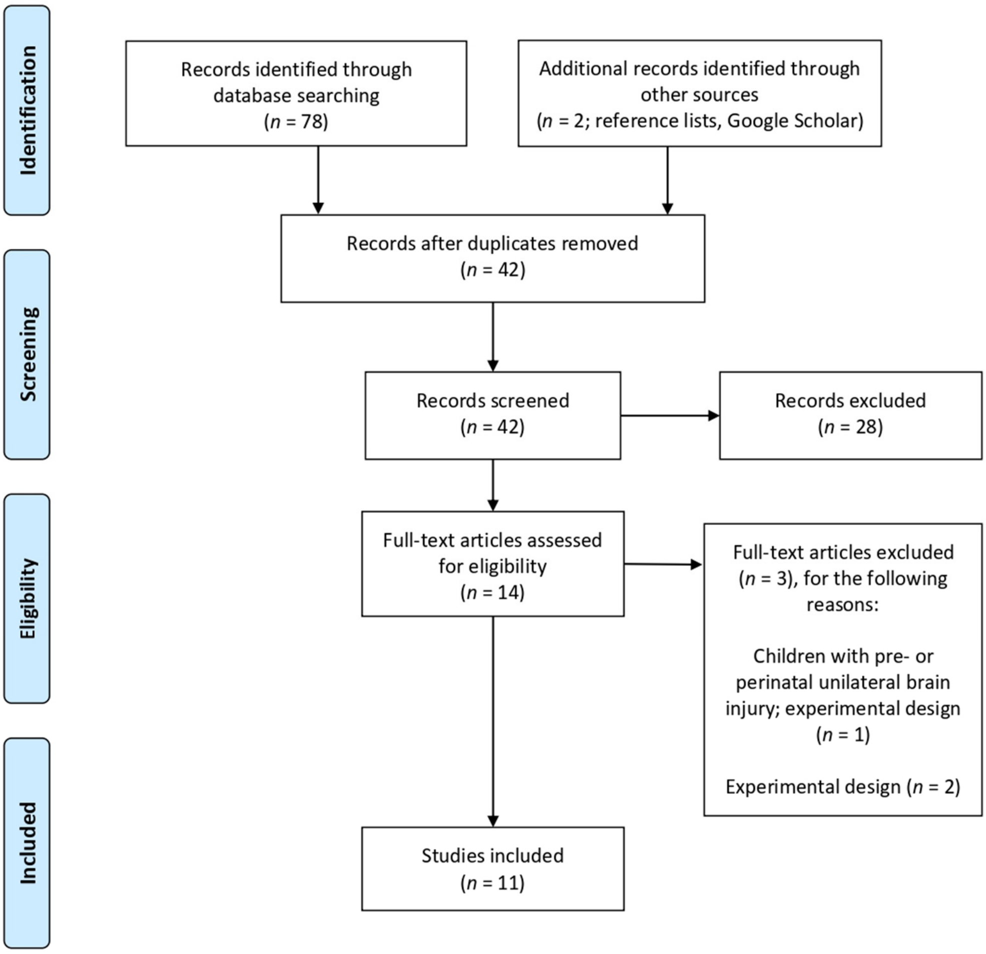


Figure 1. PRISMA Flow Diagram.

To this end, all the potentially relevant articles that were identified and that could answer the questions were assessed as full text independently by the two authors according to the following eligibility criteria, so as to reduce the risk of inclusion bias:

- The article was written in English and published (or under review) in a peer-reviewed journal or in peer-reviewed conference proceedings.
- The study was published by 1970 or later.
- The study reported in the article followed an experimental design yielding quantitative data. Thus, studies which assessed gesture use and development in a merely descriptive fashion without manipulating variables (e.g., [54,55]) were excluded.
- Participants were TD children, aged from 2 to 8 years who did not present any language or developmental disorder that affects communication. To our knowledge, there is no previous research that has dealt with the effects of non-referential gestures in children with language disorders.
- The experimental design involved speech with the presence or absence of non-referential gestures (either by investigating the children's observation of another speaker's gestures or the children's own gesture production) as an isolated variable. Thus, studies that tested the impact of all kinds of gestures simultaneously (i.e., multimodal training studies) were excluded.
- The outcome of the experiment was measured in terms of memory recall, comprehension, or language production.

2.2. Data Extraction

The assessment process yielded a total of 11 full text articles which met the inclusion criteria. All were published between 2012 and 2020 except for one, which is currently undergoing the final review process prior to publication. Likewise, in all 11 studies a non-referential gesture experimental condition was compared to other gesture or no gesture conditions. The total sample size of all the studies was 898 children between 2 and 8 years of age whose native language was English, Catalan (Catalan-Spanish bilingual), or Turkish. One study was conducted in Singapore, three in Australia, six in Catalonia, and one in Turkey. The Standard Quality Assessment criteria for evaluating primary research papers from a variety of fields (Kmet checklist Appraisal Tool) was used to assess the methodological quality of the studies [74]. The overall quality of the studies was found to be acceptable, with clearly stated research questions and appropriately used experimental methods.

For each study, the following information was extracted:

- Reference: author(s) and year of publication
- Aim of the study
- Study population, including number, gender, age range (mean and standard deviation), and language of participants
- Study design
- Control and experimental conditions
- Outcome measure
- Main results

This data (except study aims) can be seen in Table 1, with the main results framed in terms of the extent to which they answered the question "Did non-referential gestures have a positive effect on children's outcome measure?" All the information summarized in the tabular form was observed to find the main similarities between studies meeting the inclusion criteria.

Table 1. Empirical studies included in the systematic review.

Author, Year	Study Population	Study Design	Control and Experimental Conditions	Outcome Measure	Did Non-Referential Gestures Have a Positive Effect on Children's Outcome Measure?
Austin & Sweller (2014)	91 children (49 girls and 44 boys); *M* = 4 years 3 months, *SD* = 4 months; range = 3 years 4 months to 4 years 9 months; Australian-English speakers	Between-subjects experiment	(1) No gesture; (2) Beat gesture; (3) Combined gesture (5 beats, 5 deictics, 5 metaphorics, and 5 iconics)	Spatial information recall	Yes → Beat gesture condition and Combined gesture condition (vs. No gesture condition)
Austin & Sweller (2017)	172 children (original sample: 77 girls and 97 boys); *M* = 4 years 5 months, *SD* = 4 months, range = 3 years 0 months to 5 years 4 months; Australian-English speakers	Between-subjects experiment	(1) Iconic/deictic gesture; (2) Beat gesture; (3) No gesture	Spatial information recall and cued recall	No → Iconic/deictic gesture condition (vs. Beat gesture condition and No gesture condition)

Table 1. *Cont.*

Author, Year	Study Population	Study Design	Control and Experimental Conditions	Outcome Measure	Did Non-Referential Gestures Have a Positive Effect on Children's Outcome Measure?
Igualada et al. (2017)	106 children (47 girls, 59 boys); 3 years: M = 41.74, SD = 3.58; 4 years, M = 53.93, SD = 3.79; 5 years: M = 64.91, SD = 3.16; Catalan speakers	Within-subject experiment	(1) Beat; (2) No-beat	Word recall	Yes → Beat condition (vs. No-beat condition)
Kartalkanat & Göksun (2020)	67 children (original sample: 36 girls and 35 boys); M = 64.00 months, SD = 4.97; Turkish speakers	Between-subjects experiment	(1) Iconic gesture; (2) Beat gesture; (3) No gesture	Path and event information recall	No → Iconic gesture condition (vs. Beat gesture condition and No gesture condition)
Llanes-Coromina et al. (2018)	51 preschool children; M = 4.57, SD = 0.26; Catalan-Spanish bilingual speakers (Experiment 1)	Within-subject experiment (Experiment 1)	(1) Non-prominent speech; (2) Prominence in speech alone; (3) Prominence in both speech and gesture (beat gestures) (Experiment 1)	Information recall in contrastive discourse (Experiment 1)	Yes → Prominence in both speech and gesture condition (vs. Non-prominent speech and Prominence in speech alone conditions)
	55 children; M = 5.84, SD = 0.56; Catalan-Spanish bilingual speakers (Experiment 2)	Between-subjects experiment (Experiment 2)	(1) Beat; (2) No-beat (Experiment 2)	Narrative comprehension (Experiment 2)	Yes → Beat condition (vs. No-beat condition)
Macoun & Sweller (2016)	101 children (57 girls and 44 boys); girls: M = 4.62, SD = 0.40; boys: M = 4.70, SD = 0.57; total M = 4.65, SD = 0.47; Australian-English speakers	Between-subjects experiment	(1) Iconic gesture; (2) Deictic gesture; (3) Beat gesture; (4) No gesture	Narrative recall and narrative comprehension	No → Iconic gesture condition and Deictic gesture condition (vs. Beat gesture condition and No gesture condition)
So et al. (2012)	36 children (18 girls and 18 boys); 4 to 5 years; English speakers	Within-subject experiment	(1) Iconic gesture; (2) Beat gesture; (3) No gesture	Word recall	No → Iconic gesture condition (vs. Beat gesture condition and No gesture condition)

Table 1. *Cont.*

Author, Year	Study Population	Study Design	Control and Experimental Conditions	Outcome Measure	Did Non-Referential Gestures Have a Positive Effect on Children's Outcome Measure?
Vilà-Giménez et al. (2020)	83 children (43 girls, 40 boys); Time 1: *M* = 5.9, *SD* = 0.55; Time 2: *M* = 7.98, *SD* = 0.60; Catalan-Spanish bilingual speakers	Longitudinal	(1) Non-referential beat gesture; (2) Referential iconic gesture	Later oral narrative productions (narrative structure scores)	No → Referential iconic gestures (vs. Non-referential beat gestures)
Vilà-Giménez et al. (under review)	45 children; Time 1: between 14 to 58 months of age; Time 2: *M* = 6, *SD* = 0.42; American-English monolingual speakers	Longitudinal	(1) Non-referential beat gesture; (2) Non-referential flip gesture; (3) Referential iconic gesture	Later oral narrative productions (narrative structure scores)	Yes → Non-referential beat gestures (vs. non-referential flips and referential iconics)
Vilà-Giménez et al. (2019)	44 children (20 girls and 24 boys); *M* = 5.94 years; *SD* = 0.57; Catalan-Spanish bilingual speakers	Between-subjects training (pretest-posttest design)	(1) Beat; (2) No-beat	Oral narrative performance (narrative structure scores)	Yes → Beat condition (vs. No-beat condition)
Vilà-Giménez & Prieto (2020)	47 children; *M* = 5.92, *SD* = 0.54; Catalan-Spanish bilingual speakers	Between-subjects training study (pretest-posttest design)	(1) Beat encouraging; (2) Beat non-encouraging	Oral narrative performance (narrative structure and fluency scores)	Yes → Beat encouraging condition (vs. Beat non-encouraging condition)

Formal meta-analysis was not considered feasible and therefore not undertaken due to the low number of studies and to the fact that the study designs and reported outcome measures varied markedly. Thus, a narrative synthesis seemed to be the most appropriate way to compare the results of these studies and draw overall conclusions.

3. Results

Because these 11 studies did not use the same outcome measures, they were divided into three groups on that basis for purposes of comparison. It is important to mention that only comparable variables within experimental designs were extracted for all studies. Below, we first compare studies that reported results related to the association effects between non-referential gestures and children's cognitive and linguistic abilities, such as information recall and narrative discourse comprehension (Section 3.1). Next, we compare studies that examined the predictive value of non-referential gestures in children's later narrative productions (Section 3.2). Finally, we compare studies that examined the causal effects of gesture training paradigms using non-referential gestures on children's oral narrative discourse performance (Section 3.3).

3.1. Association Effects

3.1.1. Information Recall

Seven of the 11 studies of this systematic review assessed the effects of observing non-referential gestures on information recall in TD children aged between 3 and 6 years. While three of these studies examined the effects on children's recall of words [20,75,76], three others dealt with their ability to memorize spoken spatial directions or paths and event information [77–79], and the remaining one focused on their free recall of narratives [45]. Overall, 3 of the 7 experimental studies (two dealing with word recall and one with verbal spoken spatial directions) showed a positive effect of observing non-referential gestures on children's information recall, whereas the remaining studies (one on word recall, two on verbal spoken paths and event information, and one on free narrative recall) did not.

Regarding the three studies that reported non-referential gestures having beneficial effects on recall, it is noteworthy that they presented the information in contexts that were pragmatically relevant for children. Two of them followed a within-subject experimental design [75,76], while the other one had a between-subjects design [77]. The study by [77] examined whether the presence of different gesture types in the verbal descriptions of a target path (presented in a single small-scale spatial array constructed from Lego materials and with a Lego character taking a certain route through the spatial array) would have an impact on the extent to which spatial information about this route (i.e., location and movement terms) was recalled by both adults and children. Participants were given the verbal description of the target path in the assigned condition and were then asked to recall the path (e.g., "Can you tell me the path that Lego man took through the scene without taking Lego man along with you?," "Can you now show me the path that Lego man takes through the scene, taking Lego man along with you?"). A total of 95 adults (M = 28 years; SD = 7.6; range = 17 to 49) and 93 children (M = 4 years 3 months, SD = 4 months, range = 3 years 4 months to 4 years 9 months) participated in the experiment; however, the final sample consisted of 94 adults and 91 children. Participants were randomly assigned to one of three conditions: combined gesture (including 5 deictic, 5 beat, 5 iconic, and 5 metaphoric gestures), beat gesture, or no gesture. Results revealed that children in either the combined gesture condition or the beat gesture condition showed better verbal recall of spoken spatial directions than children in the no gesture condition. Overall statistical analysis using ANOVA found a large main effect of age group ($F(1, 173) = 272.73$, $p < 0.001$, partial $\eta^2 = 0.61$), showing that the children recalled less information than the adults, and a small main effect of gesture condition ($F(2, 173) = 3.17$, $p = 0.045$, partial $\eta^2 = 0.03$), revealing that gesture conditions (beat or combined) were more beneficial than the no gesture condition for information recall (ANOVA effect sizes are interpreted following the benchmarks suggested by [80].). The researchers took the average of the two gesture conditions and created a new "gesture" vs. "no gesture" variable and calculated the interaction between this new variable and age. A significant interaction between age group and the difference between the no gesture condition and the average of the two gesture conditions was found ($F(1, 179) = 5.16$, $p = 0.024$), with a small effect (partial $\eta^2 = 0.03$), revealing that the difference in recall between age groups was greater for the no gesture condition than for the gesture conditions. Simple effects showed that for the total recall of spatial information by children there was a significant difference between the no gesture condition and the average of the two gesture conditions ($F(1, 179) = 9.75$, $p = 0.002$), with a small effect (partial $\eta^2 = 0.05$). As for the adults, no significant differences were found between the no gesture condition and the average of the two gesture conditions on the total recall of spatial information ($F(1, 179) = 0.006$, $p = 0.934$, partial $\eta^2 < 0.01$). Further, no significant difference between the two gesture conditions was found ($p > 0.05$) and there was no interaction between age and the two gesture conditions ($F(1, 179) = 0.36$, $p = 0.551$, partial $\eta^2 < 0.01$).

In their within-subject study, [75] investigated whether the presence of non-referential gestures would improve word recall in 106 children aged 3 to 5 years (3-year-olds: M = 41.74 months, SD = 3.58; 4-year-olds: M = 53.93 months, SD = 3.79; 5-year-olds: M = 64.91 months, *SD* = 3.16) when presented with a list of things that Elmer, an absent-minded elephant, needed to remember before he went on a trip. Children completed the experimental task under two different audiovisual conditions, a beat condition and a no-beat condition, presented successively in counterbalanced orders. In the beat condition, each target word was accompanied by a beat gesture, whereas in the no-beat condition no beat gestures were used. Children were asked whether they could help Elmer to remember all the items on his to-do list, as he was very absent-minded and would really appreciate their help. Statistical analysis of the results using a Generalized Linear Mixed Model (GLMM) only showed a main effect of condition ($F(1, 418) = 4.01$, $p < 0.05$), indicating that the children recalled significantly more words in the beat condition than in the no-beat condition ($\beta = 0.124$, $SE = 0.062$, $p < 0.05$; no-beat condition: $M = 0.38$, $SD = 0.48$; beat condition: $M = 0.49$, $SD = 0.50$) (If the beta coefficient is positive, the interpretation is that for every 1-unit increase in the predictor variable, the outcome variable will increase by the beta coefficient value.). No effect of age ($F(2, 418) = 2.80$, $p = 0.062$) and no interaction between gesture condition and age ($F(2, 418) = 0.11$, $p = 0.849$) were found. Importantly, these findings show that observing non-referential gestures had a positive impact on word recall in children aged 3 to 5 years.

Along the same lines, the study by [76] showed that observing non-referential gestures can positively influence the memorization of contrastively focused items as well as information related to those items within contrastive discourse (i.e., containing a set of contrastively focused items). In this experiment, 51 4-year-old children ($M = 4.57$ years, $SD = 0.26$) were presented with discourse contexts in which a female human reminds Elmer the elephant about what they have done in their trips together (e.g., "Elmer, do you remember our trip to the field? In the morning, we went for a walk in the field. [. . .] We noticed that near the lake there were roses and leaves, and you picked the roses [. . .]"). These contexts contained a set of contrastively focused items (e.g., "roses and leaves") in three conditions in a counterbalanced order: non-prominent speech, prominence in speech alone, and prominence in both speech and gesture (beat gestures). The children were then asked whether they could help Elmer to remember the questions that were related to what he and his friend had done together (e.g., "Now, help Elmer remember what he picked up when he went to the field. What did he pick up?"). A GLMM analysis of the results showed a significant main effect of condition ($F(2, 288) = 5.28$, $p = 0.006$), revealing that the children recalled more contrastively focused items in the prominence in both speech and gesture condition, compared to either of the speech-alone conditions (prominence in both speech and gesture vs. prominence only in speech, $\beta = 0.241$, $SE = 0.077$, $p = 0.006$; prominence in both speech and gesture vs. non-prominent speech, $\beta = 0.191$, $SE = 0.081$, $p = 0.038$). Moreover, regarding the proportion of items recalled between the prominence only in speech vs. non-prominent speech conditions, no difference was found ($\beta = -0.050$, $SE = 0.082$, $p = 0.537$).

By contrast, the 4 other studies found that observing non-referential gestures had no beneficial effect on children's recall of information. In a study involving 30 adults (Experiment 1) and 36 4- to 5-year-old children (Experiment 2), [20] tested whether beat gestures and iconic gestures would enhance word recall by showing a video presentation of a list of verbs shown in isolation without a relevant discourse context in three counterbalanced within-subject experimental conditions: the verbs were accompanied by either an iconic gesture, a beat gesture or no gesture. Concerning the findings for the children, a main effect of gesture condition was found ($F(2, 68) = 20.16$, $p < 0.001$), with a large effect ($\eta^2 = 0.37$), revealing that the children recalled a higher proportion of words in the iconic gesture condition than in either the beat gesture condition or the no gesture condition ($p > 0.001$). No significant difference between the beat gesture condition and the control condition was found, which indicates that non-referential gestures (i.e., the beat gesture condition) did not facilitate children's word memory recall. However, results for the adults found a large main effect of condition ($F(2, 56) = 7.87$, $p = 0.001$, $\eta^2 = 0.21$), showing that the participants displayed better recall scores when words were accompanied with iconic gestures than in the no gesture condition ($p = 0.002$), and when they were accompanied with beat gestures compared with the no gesture condition ($p = 0.009$). No difference between the number of items recalled in both gesture conditions was found. Further analyses comparing results from children and adults showed a large main effect of age group, revealing that the adults recalled a higher proportion of words than the children ($F(1, 64) = 192.69$, $p < 0.001$, $\eta^2 = 0.75$), and a large main effect of condition ($F(2, 128) = 22.11$, $p < 0.001$, $\eta^2 = 0.26$). A significant interaction between condition and group was found ($F(2, 128) = 6.91$, $p = 0.001$, $\eta^2 = 0.10$). All in all, the proportion of words recalled was higher in the iconic gesture condition than in the no gesture condition for all participants. While the children recalled comparable proportions of words in the beat gesture and no gesture conditions, the adults recalled more words in the beat gesture condition than in the no gesture condition. These results suggest that non-referential gestures may entail a higher cognitive demand for children than for adults, who benefited from the presence of either iconic or beat gestures. In contrast to previous experimental designs, in this study each word was accompanied by a beat gesture, which may have reduced the highlighting function of beats. Moreover, the list of verbs was presented in isolation and not in a pragmatically natural discourse context.

Null results for non-referential gestures were also found in a study carried out by [78] in a between-subjects route direction task in which the participants were presented with a set of instructions to guide visitors through a zoo and then asked to recall and reconstruct the instructions. Like in [77], the goal was to examine the effects of gesture observation on the recall of the spatial information, but in this case the spatial direction task performed was larger in scale (i.e., when the spatial environment cannot be viewed from a single viewpoint). Participants were 172 3- to 5-year-old children (*M* = 4 years 5 months, *SD* = 4 months, range = 3 years 0 months to 5 years 4 months), who were randomly assigned to either the iconic/deictic gesture condition, the beat gesture condition, or the no gesture condition, and were presented with three videos of the head zoo-keeper verbally giving route directions through the zoo. In the two gesture conditions, key spatial or movements descriptors in the instructions were accompanied by nine gestures (e.g., "walk <u>forward</u> for a little bit," "go <u>past</u> the frogs," with underlined words indicating gesture points), which were either iconic/deictic or beat gestures, depending on the condition. A mixed-design ANOVA of the results revealed a main effect of gesture condition ($F(2, 169) = 3.85$, $p = 0.023$), with a small effect (partial $\eta^2 = 0.04$), demonstrating that the children verbally recalled more items in the iconic/deictic gesture condition than in the beat gesture condition ($F(1, 169) = 6.30$, $p = 0.013$), with a small effect (partial $\eta^2 = 0.04$). Moreover, no difference between the no gesture condition and the average of the two gesture conditions was found in terms of the number of verbally recalled items ($F(1, 169) = 1.56$, $p = 0.213$, partial $\eta^2 = 0.01$). A possible explanation given by the authors for these findings is that referential gestures "may be processed more deeply due to their semantic value, leading to great recall without the presence of environmental cues" [78] (p. 10). To explain the lack of effects of non-referential gestures, the authors point out that "it is possible that the communication of spatial information accompanied by either no gestures or beat gestures may have seemed unusual or odd to preschoolers given that they would usually experience such messages accompanied by iconic and deictic gestures" (p. 11).

Interestingly, a second experiment was performed within this study involving the same route direction task but using a more pragmatically relevant instruction for the child. In this case, the child was asked to go to a location in the zoo where he/she remembered the zookeeper giving a particular instruction, and the experimenter recorded the path of movement on a paper map of the zoo. A mixed-design ANOVA regarding cued recall (i.e., the amount of route recalled verbally and during physical route retracing) showed a medium-sized main effect of condition ($F(2, 169) = 5.72$, $p = 0.004$, partial $\eta^2 = 0.07$), indicating that the children presented with the materials in the two gesture conditions (beat condition or iconic/deictic condition) reported more at cued recall than the children presented with no gesture ($F(1, 169) = 10.06$, $p = 0.002$), with a medium effect size (partial $\eta^2 = 0.06$). No difference between the amount recalled at cued recall in the iconic/deictic and beat gesture conditions was found ($F(1, 169) = 1.60$, $p = 0.208$, partial $\eta^2 = 0.01$). Importantly, in this second experiment, although iconic gestures improved the children's recall most, beat gestures also had some positive effect, suggesting that "the benefit of beat gestures may be apparent only when recall is cued by the environment" [78] (p. 10).

In the study by [79], a total of 67 4- to 6-year-old (54–73 months) children ($M = 64.00$ months, $SD = 4.97$) and 54 adults ($M = 21.50$ years, $SD = 1.95$) were asked to listen to a story about a character who followed different paths to find her friend's house and then recount the information they had heard. The story included path descriptions (five alternative routes with various details) which were followed by event sequences with no spatial content (e.g., path: "she walked around the mountains;" event: "she saw a bank and took a rest on the bank;" again, underlined words indicate gesture points). In total, the story consisted of ten sentences each accompanied by one gesture (the underlined segment in the preceding examples). In a between-subjects design, participants were randomly assigned to one of these three conditions: iconic gesture condition, in which participants observed the stories with iconic gestures depicting the described path or action; beat gesture condition, in which speech was accompanied with rhythmic hand movements; or no gesture condition, in which the participant heard the narrative without any gesture. Participants were then asked a free recall-eliciting question (e.g., "Can you tell me everything you remember from the story?") and their answers were subsequently scored for amount of content by a researcher. A mixed-design ANOVA found a large main effect of age group ($F(1, 119) = 117.31$, $p < 0.01$, partial $\eta^2 = 0.51$), showing that the adults recalled more information than the children, and also a medium-sized main effect of gesture condition ($F(2, 119) = 3.92$, $p = 0.022$, partial $\eta^2 = 0.07$). Post hoc analyses showed that when participants observed iconic gestures, they recalled more information than in the other conditions (beat gesture and no gesture condition) (Bonferroni, *ps* < 0.05). No significant interaction between age group and gesture condition was found ($F(2, 119) = 0.39$, $p = 0.676$, partial $\eta^2 = 0.007$). Descriptive statistics showed that, for the children, the mean total free recall was 27.06 ($SD = 12.30$, min = 0, max = 45) in the iconic gesture condition, 21.63 ($SD = 15.38$, min = 0, max = 52.5) in the beat gesture condition, and 20.34 ($SD = 12.82$, min = 0, max = 45) in the no gesture condition. As for the adults, the mean total free recall was 60.29 ($SD = 14.22$, min = 40, max = 92.5) in the iconic gesture condition, 45.00 ($SD = 22.64$, min = 0, max = 95) in the beat gesture condition, and 48.44 ($SD = 18.84$, min = 15, max = 72.5) in the no gesture condition. It is important to mention that in this experiment, iconic gestures merely reinforced content and did not provide additional information. Therefore, these findings suggest that iconic gestures not only provide semantic cue, but also direct attention to certain parts of the story. By contrast, although the beat gestures were embedded in sentences to highlight target information in a naturalistic fashion, they made no contribution to participants' recall performance.

Finally, the study by [45] reported no enhancement effects of beat gestures on the free narrative recall of 101 preschoolers aged 3.25 to 5.58 years ($M = 4.65$ years, $SD = 0.47$). In this experiment with a between-subjects design, children were asked to watch a video of a storyteller telling a two-minute narrative about a girl's afternoon at the park with her family in one of four randomly assigned gesture conditions: iconic gesture, deictic gesture, beat gesture, or no gesture. In all conditions, gestures occurred at a total of ten points in the story. In the iconic gesture condition, the gestures represented the shape or action of the object described in the speech; in the deictic gesture condition, the gestures indicated the position of items referred to in the speech; and in the beat gesture condition the narrator produced rhythmic hand movements with no representational meaning and in focused positions. The different gesture types occurred at the same points in the narrative across conditions. After they had watched the video, the children were asked a free recall-eliciting question (e.g., "Please tell me everything you remember about the story you saw told on the computer"). Results demonstrated that children in the iconic and deictic gesture conditions scored higher on recall task than children in either the beat gesture or no gesture condition, between which there were no differences. A one-way between-groups ANOVA showed a main large effect of gesture condition on narrative recall ($F(3, 97) = 6.69$, $p < 0.0005$, partial $\eta^2 = 0.17$). Further analyses showed that observing iconic gestures ($F(1, 97) = 10.14$, $p = 0.010$, partial $\eta^2 = 0.09$, with a medium effect size) or deictic gestures ($F(1, 97) = 18.17$, $p < 0.0005$, partial $\eta^2 = 0.16$, with a large effect size) increased narrative recall compared with no gesture condition. No other comparisons were found to be significant ($p > 0.10$). Perhaps unsurprisingly, when the recall of information available only through gestures and not present in the content of the narrative was analyzed, pairwise comparisons in a binary logistic regression ($\chi^2(3) = 14.33$, $p = 0.002$) found that the odds of reporting this information was higher for the deictic (B = 2.23, Wald = 10.59, $p = 0.001$, odds ratio = 9.33) and iconic (B = 1.74, Wald = 6.60, $p = 0.01$, odds ratio = 5.69) conditions than for the no gesture condition (The odds of success are defined as the ratio of the probability of success over the probability of failure. An odds ratio greater than 1 is a positive association (i.e., higher number of the predictor means group 1 in the outcome), while an odds ratio less than 1 is a negative association (i.e., higher number for the predictor means group 0 in the outcome). However, results for the beat gesture condition did not differ in this regard from those for the control condition (B = 1.02, Wald = 2.24, $p = 0.14$, odds ratio = 2.78), indicating that these gestures conferred no advantage.

All in all, it is worth noting that the three abovementioned studies reporting benefits of exposure to non-referential gestures relied on naturalistic uses of non-referential gestures in their experimental materials and assessed their role within discourse contexts that were pragmatically relevant for preschool and school children (e.g., small-scale route directions, list of things that an elephant needed to do before travelling, contrastive discourse).

3.1.2. Narrative Discourse Comprehension

Two of 11 of the papers included in this systematic review addressed the potential role of observing non-referential gestures on narrative comprehension processes using a between-subjects experimental design. These two papers were also reviewed in the preceding section (see Section 3.1.1) because they analyzed the effects of observing non-referential gestures on information recall. As in that section, while the study by [76] showed positive effects of non-referential gestures in 5- and 6-year-old children's narrative comprehension, the study by [45] found no such benefits (though the age of their participants was somewhat lower at 3.25–5.58 years).

The second part of the study by [76] tested the benefits of observing beat gestures in 55 5- and 6-year-old children ($M = 5.84$ years, $SD = 0.56$) in a narrative discourse task. Each participant was randomly assigned to one of two between-subjects conditions: beat gesture condition and no-beat gesture condition. In the no-beat condition, discourses were presented with prosodic prominence and no beat gestures in the target words, while in the beat condition, discourses were with prosodic prominence and with beat gestures in the target words (i.e., both performed on discourse markers and focal content words). The children were first shown a set of videos in which a storyteller told—-with or without gestures accompanying discourse markers and focal content words—-short six one-minute stories involving some farm animals that were friends of a sheep. After viewing each video, the children were asked to help the sheep find out what had happened to each animal and were asked two comprehension questions (e.g., "Why did the pig have to go home back early?" and "So how did the pig solve his problem?"). The children's responses to the questions were then scored for comprehension. The results of a statistical GLMM analysis revealed a significant main effect of condition ($F(1, 657) = 4.21$, $\beta = 0.572$, $SE = 0.279$, $p = 0.041$, odds ratio = 1.772), indicating that the children comprehended the stories better when they were performed with beat gestures. It is important to note that in order to design the experimental materials, a preliminary study was conducted to determine what kinds of beat gestures naturally accompany child-directed narratives and at what points in the narratives they are typically used. This preliminary study guided both the form of the non-referential gestures and the placement of those gestures within the narrative discourse.

These results contrast with the findings of a second task reported in [45], in which (as described above) preschool children were asked to listen to a two-minute story in either iconic gesture, deictic gesture, beat gesture, or no gesture conditions. However, this second task was intended to test the effect of gesture conditions on the children's narrative comprehension. Thus, in this task, after they had been exposed to the story, the children were asked 15 randomized specific questions related to the content of the narrative. As previously mentioned, the gestures only occurred at ten places in the narrative. Importantly, in the beat gesture condition, rhythmic hand movements without reflecting contextual meaning of the speech were performed in focused positions within discourse. Five of these questions took into account general story content (non-gesture-related questions), another five concerned gestures that reinforced but did not add to story content (redundant gesture-related questions), and the other five concerned gestures that conveyed information not present in the verbal narrative (non-redundant gesture-related questions). Results demonstrated that while the beat gesture condition and no gesture condition yielded similar narrative comprehension scores, meaning that beat gestures in no way enhanced comprehension, iconic and deictic gestures led to higher scores. Analyses to determine the effect of condition on non-gesture-related question scores using a one-way between-groups ANOVA found no significant difference between conditions in terms of narrative comprehension ($F(3, 97) = 2.19$, $p = 0.093$, partial $\eta^2 = 0.06$). Finally, results on the effect of condition on gesture-related question scores using a one-way between-groups ANOVA showed a main large effect of gesture condition ($F(3, 97) = 6.45$, $p < 0.0005$, partial $\eta^2 = 0.17$). Further analyses found the same outcomes as in the free recall results (iconic, $F(1, 97) = 10.37$, $p = 0.009$, partial $\eta^2 = 0.10$, with a medium effect size; deictics, $F(1, 97) = 6.98$, $p = 0.047$, partial $\eta^2 = 0.07$, with a medium effect size). Moreover, children produced higher comprehension scores on gesture-related items when they were accompanied by iconic ($F(1, 97) = 12.34$, $p = 0.004$, partial $\eta^2 = 0.11$, with a medium effect size) or deictic ($F(1, 97) = 8.58$, $p = 0.022$, partial $\eta^2 = 0.08$, with a medium effect size) gestures relative to children in the beat gesture and no gesture conditions, scores from which showed no significant differences ($p = 0.994$). Differences between scores in the iconic and deictic gesture conditions were likewise not significant ($p = 0.938$).

A potential reason for the difference between the results yielded respectively by [76] and [45] lies in the experimental materials employed. While (as noted above) the former study conducted a preliminary study in order to construct a more natural set of experimental materials, this was not the case in the latter study. In our view, it is important that beat gestures in discourse are assessed in terms of both the shape of the hand during the gesture and the point in the narrative at which the gesture occur, because both factors can mediate the gesture's effect.

3.2. Predictive Effects

Two of the articles selected for this review were recent longitudinal studies that examined the predictive effects of the early frequency of use of non-referential beat gestures in children's later more complex linguistic skills [68,81]. While both studies addressed predictive effects, they differed in two aspects. While the study by [68] examined the effects of 45 children's production of non-referential gestures between 14 and 58 months of age in parent-child naturalistic interactions, the study by [81] tested the effects of the production of these gestures in older 5- to 6-year-old children while performing narrative discourses.

The main objective of the longitudinal study by [68] was to investigate whether the early production of non-referential beat and flip gestures (Non-referential flip gestures, a subtype of non-referential gestures, are performed by turning the wrist of the hand and opening it up to present the flat palm, accompanied or not with a shrug of the shoulders. They typically convey a judgmental or epistemic value of ignorance (e.g., [82])) (vs. referential iconic gestures) produced by 45 children in the total developmental window from 14 to 58 months of age predicted later narrative productions at 60 months (5 years old), measured in terms of narrative structure scores. On average, the children produced

1.19 beat gestures per session (*SD* = 1.74, range = 0 to 10.23), 1.86 flips per session (*SD* = 1.87, range = 0.15 to 9.15) and 3.58 iconic gestures per session (*SD* = 2.73, range = 0.31 to 11.46). Results from a GLMM analysis showed that the average number of beat gestures produced at baseline significantly predicted narrative skills at age 5 (β = 0.299, *SE* = 0.111, *z* = 2.689, $p < 0.01$). By contrast, the average number of flips (β = −0.163, *SE* = 0.109, *z* = −1.489, *p* = 0.137) and iconic gestures (β = 0.029, *SE* = 0.077, *z* = 0.381, *p* = 0.703) did not predict later narrative productions. This model explained 88.4% of the variance in children's narrative outcomes (R^2 = 0.884). Moreover, a second GLMM analysis also showed that the average number of non-referential beat gestures produced between 14 and 42 months of age were still predictors of children's later narrative productions at age 5 (β = 1.386, *SE* = 0.583, *z* = 2.377, *p* = 0.017), while no significant effect was found for flips (β = −0.136, *SE* = 0.112, *z* = −1.212, *p* = 0.225) or iconic gestures (β = 0.009, *SE* = 0.067, *z* = 0.137, *p* = 0.891). This model explained 80.1% of the variance in children's narrative outcomes (R^2 = 0.801).

The second longitudinal study [81] reported the predictive value of both referential and non-referential gestures produced during narrative discourse by 5- to 6-year-olds (*M* = 5.9 years, *SD* = 0.55) on their later narrative productions (measured in terms of structural wellformedness) two years later, at 7 to 9 years of age (*M* = 7.98 years, *SD* = 0.60). On average, when they were 5–6 years the children produced 0.90 referential iconic gestures (*SD* = 1.54, *n* = 149) and 0.63 non-referential beat gestures (*SD* = 0.91, *n* = 105) in their narratives. A linear stepwise regression analysis was run to predict their narrative abilities at 7–9 years old based on the number of referential iconic gestures and non-referential beat gestures the children produced in their narratives at 5–6 years of age ($F(1, 81)$ = 5.64, *p* = 0.020). Results demonstrated that the use of referential iconic gestures during narrative performance at 5–6 years predicted narrative structure scores two years later, when children were 7–9 years of age (β = 0.154, *SE* = 0.065, *p* = 0.020). However, no significant results were found for non-referential beat gestures (*p* = 0.432).

3.3. Causal Effects

Only two of the studies included in this review assessed the possible causal effects of narrative training that includes non-referential beat gestures in children's narrative performance. Both studies involved 5- and 6-year-old children and used a between-subjects pretest-posttest experimental design. However, the studies differed in the main goal of the research. While the study by [83] examined the effects of having children observe beat gestures as part of a short narrative training task on their narrative performance in a posttest, the study by [84] investigated whether encouraging children to produce beat gestures could also affect their subsequent narrative performance.

In the first of these studies [83], following a pretest measuring their ability to produce a well-formed narrative, 44 5- and 6-year-old children (*M* = 5.94 years, *SD* = 0.57) underwent training which involved watching six one-minute stories presented under two randomly assigned experimental conditions: a beat gesture condition, in which a storyteller performed a narrative with prosodic prominence and beat gestures whenever she said a discourse marker or focal content word, and a no-beat gesture condition, where narratives were performed with prosodic prominence and no beat gestures in target positions within the story. Again, a preliminary study was carried out to identify the types of beat gestures that are spontaneously produced in child-directed narratives as well as to detect at what points in the narrative discourse these beat gestures tend to occur in natural circumstances. Children were simply asked to observe the stories. Children's pretest and posttest narratives were then scored and compared by a researcher in terms of their structural wellformedness. Results of a GLMM analysis examining condition against structural wellformedness scores showed a main effect of condition ($F(1, 172)$ = 8.04, *p* = 0.005), specifically in the beat gesture condition (β = 0.441, *SE* = 0.156, *p* = 0.005); and a main effect of test ($F(1, 172)$ = 19.69, $p < 0.001$), with better posttest narrative structure scores than pretest scores (β = 0.597, *SE* = 0.135, $p < 0.001$). Moreover, the interaction between condition and test was found to be significant ($F(1, 172)$ = 4.71, *p* = 0.031). Further post hoc analyses showed that gesture conditions differed in the posttest part, showing that higher narrative structure scores were produced by children in the beat gesture condition (β = 0.733, *SE* = 0.207, $p < 0.001$) than in the no-beat gesture condition. However, differences in gesture conditions were not reflected in pretest scores (β = 0.149, *SE* = 0.205, $p = 0.467$). Significant differences between pretest and posttest narrative scores were found in the beat gesture condition, with better scores in the posttest (β = 0.889, *SE* = 0.186, $p < 0.001$) than in the pretest. Differences between pretest and posttest scores in the no-beat gesture condition were not found to be significant ($p = 0.119$).

The second study [84] used the same narrative training paradigm employed in the previous study but assessed whether having children not only observe but also encouraging them to produce beat gestures would enhance the effects seen in [83]. In this case, 47 5- to 6-year-old children (*M* = 5.92 years, *SD* = 0.54) were randomly assigned to one of two experimental conditions: beat encouraging condition and beat non-encouraging condition. Following a pretest which measured not only structural wellformedness but also fluency on their narrative output, the children were shown videos of the same six narratives used in the previous study, though in this case both groups saw the version of the video in which the storyteller performed prosodic prominence and beat gestures in target positions. Children were then asked to retell the story they had just heard. However, while children in the beat non-encouraging condition were asked to retell the stories without any instructions regarding gesture, in the beat encouraging condition they were encouraged to use hand movements (i.e., beat gestures) like those they had seen the storyteller use while recounting what they had heard. Children's pretest and posttest narratives were then scored and compared. Results from a first GLMM analysis found a main effect of test ($F(1, 184) = 25.19$, $p < 0.001$), with higher narrative structure scores in the posttest (β = 0.834, *SE* = 0.166, $p < 0.001$) than in the pretest, and a significant interaction between condition and test ($F(1, 184) = 6.17$, $p = 0.014$). Further post hoc analyses revealed that the gesture conditions differed in posttest narrative structure scores, with higher narrative structure scores in the beat encouraging condition (β = 0.697, *SE* = 0.265, $p = 0.009$) than in the beat non-encouraging condition. However, conditions did not differ in terms of pretest scores (β = 0.129, *SE* = 0.265, $p = 0.628$). Significant differences between pretest and posttest narrative scores were found in the beat encouraging condition, with higher scores in the posttest (β = 1.246, *SE* = 0.240, $p < 0.001$) than in the pretest. Differences between pretest and posttest scores in the beat non-encouraging condition were not found to be significant ($p = 0.069$). A second GLMM analysis revealed a main effect of test ($F(1, 184) = 18.28$, $p < 0.001$), with higher fluency scores in the posttest (β = 0.803, *SE* = 0.188, $p < 0.001$) than in the pretest, and an interaction between condition and test ($F(1, 184) = 4.65$, $p = 0.032$). Further post hoc analyses showed no significant difference between pretest scores (β = 0.214, *SE* = 0.468, $p = 0.647$) and also posttest scores (β = 0.596, *SE* = 0.533, $p = 0.265$) across conditions. Moreover, pretest and posttest scores for the beat non-encouraging condition did not significantly differ (β = 0.398, *SE* = 0.249, $p = 0.112$). However, pretest and posttest scores did differ for the beat encouraging condition, with higher fluency scores in the posttest (β = 1.208, *SE* = 0.281, $p < 0.001$) than in the pretest.

Overall, the two studies showed that either asking children to observe, or encouraging them to produce non-referential gestures in a short narrative training task, had immediate short-term effects on their narrative performance in terms of both narrative structure and narrative fluency.

4. Discussion and Conclusions

The aim of this systematic review was to search for and compare the findings of any experimental research that addressed the question of whether non-referential gestures can play a scaffolding role in both children's cognitive and linguistic abilities, as well as in the development of more complex language skills, like narrative performance. A total of 11 articles, all published within the last decade, met the eligibility requirements for inclusion. These studies —-some within-subject and others between–subjects in design—-

measured the effect of non-referential gestures on three different domains of cognitive or linguistic skill, namely information recall, narrative discourse comprehension, and oral narrative discourse performance. Immediate comparison of study findings was therefore only possible when the studies explored the same domains. At the same time, their findings revealed the presence or absence of three sorts of effects, namely association effects, predictive effects, or causal effects, leading to our three fundamental research questions. Importantly, it should be noted that there is a discrepancy in the number of studies concerning the different outcome measures. While seven papers are reporting recall and comprehension effects, only two articles focus on causal effects, and two more on predictive effects. In what follows, we will discuss what light this collective body of research sheds on each of these areas.

It must first be noted that the results of these 11 studies are not in full agreement. With regard to the effect of observing non-referential gestures on information recall, the contradictory findings can be explained by two factors, namely the pragmatic appropriateness and complexity of the task for child participants on the one hand; and on the other the choice of stimuli/materials used in each study. First of all, the studies that reported positive results [75–77] used ecologically valid instances of non-referential gestures in tasks that were pragmatically appropriate for children (small-scale route directions in [77]; a list of things that an elephant needs to do before travelling in [75]; contrastive discourse in [76]). On the other hand, although both studies by Austin and Sweller used pragmatically appropriate contexts, it may be that the larger scale route directions that the children had to recall in [78] nullified the potential benefit of non-referential gestures, which was not the case for the less complex and small-scale spatial array employed in [77]. The null results in [20] and [79] could also be explained by the lack of pragmatic appropriateness in the task for 4- to 6-year-olds. While [20] presented the gesture stimuli in isolation (i.e., lists of verbs accompanied by iconic gestures, beat gestures, or no gestures) and not in a pragmatically felicitous discourse context, [79] asked children to remember a list of sentences of a story that included both path descriptions and event sequences. Moreover, in relation to the naturalness of the experimental materials, and specifically the appropriateness of gesture co-occurrence with specific target words, the study by [79] and another study with null results [78] used beat gestures in co-occurrence with both path and event information (e.g., with target prepositions encoding spatial information, like "walk forward for a little bit", with underlined word indicating gesture point), which the authors themselves acknowledged might be perceived as unnatural. For instance, [78] note that "the communication of spatial information accompanied by either no gestures or beat gestures may have seemed unusual or odd to preschoolers given that they would usually experience such messages accompanied by iconic and deictic gestures" (p. 10). Therefore, it could be that beat gestures co-occurring with these target words did not seem natural to the participating children.

Regarding the influence of non-referential gestures in narrative comprehension processes, the contradictory results might again be related to the stimuli used. First, as we have noted, in order to ensure the validity and naturalness of the experimental materials, the study by [76] conducted a preliminary study prior to the experiment in order to determine precisely what kinds of non-referential gestures naturally accompany child-directed narratives and at what points they typically occur within the discourse. On the basis of this preliminary study, beat gestures were used in the experiment to highlight both focal content words and discourse markers. In [45], by contrast, gestures were simply placed at ten places in the narrative. Moreover, another issue to be considered is the number of gestures relative to the length of the narrative: while the stories in [76] were relatively short narratives containing between eight and eleven beat gestures each, the stories in [45] were four times longer and contained ten gestures each.

The two longitudinal studies that aimed to address the predictive role of non-referential gestures also yielded contradictory results. On the one hand, [68] provided evidence that the early frequency of use of non-referential beat gestures produced during naturalistic parent-child interactions in the developmental window from 14 to 58 months was predictive of higher narrative skill levels later at 60 months. These results contrasted with the lack of predictive value offered by non-referential flip gestures and referential iconic gestures. On the other hand, [81] examined the predictive value of both referential iconic gestures and non-referential beat gestures produced in narrative discourses by children aged 5–6 for the quality of their narrative production at 7–9 years. In this case, results did not show non-referential gestures having significant predictive value. These null results may be due to the higher number of referential iconic gestures produced at 5–6 years of age, which might have been triggered by the narrative retelling task. Another explanation for the different predictive results between studies could be related to the fact that in naturalistic parent-child interactions, children might have included all kinds of referential iconic gesture types, whereas referential iconics produced in narrative corpora could also include different viewpoints in narrative (e.g., CVPT or "observer-viewpoint", OVPT, gestures [6], in line with [47]). All these factors might have reduced the effect of non-referential gestures, whose use has been demonstrated to significantly increase with age in narrative development [58]. All in all, further studies should investigate the predictive effects of the use of non-referential gestures for later stages of narrative production, when such gestures occur more frequently and are thus more stably acquired in complex narrative discourses.

Finally, the two training studies in our selection revealed that training in oral narratives using non-referential gestures offers benefits, in terms of not only narrative structure but also oral fluency [83,84]. Both studies showed that a brief training session with non-referential gestures is valuable for narrative production, revealing a causal link between these gestures and narrative gains in the production of narratives —-a complex linguistic skill—-at 5 to 6 years of age.

Overall, though the findings reviewed in this manuscript are mixed regarding the effects of observing non-referential gestures for recall and comprehension, results were positive when these were used in pragmatically relevant and non-complex tasks for children and when they reflected a natural co-occurrence with target words. Six of the 11 studies assessed in this systematic review provide evidence of the positive effects of non-referential gestures in children's information recall, narrative discourse comprehension processes, and oral narrative discourse performance. Even though the empirical evidence of the value of non-referential gestures is not yet as strong as the evidence in favor of referential gestures, it is clear that there are sufficient grounds to claim that non-referential gestures play an important role in boosting children's learning and language development. It is of interest to note that the mixed findings obtained in the developmental literature resemble the contradictory patterns reported by studies assessing the role of non-referential gestures in adult speech processing (see [85] for a review). While some research has shown that non-referential gestures positively affect adults' ability to recall information [20,86,87], this has not been true of other studies [77,79,88] (see also [85]). According to [85], a potential reason for the lack of the beneficial effect of non-referential gestures reported in some studies could be related to the stimuli used, as positive results have generally been shown when non-referential gestures are used in pragmatically natural and restricted contexts, such as for marking contrastively focused information. Similar to the child experiments reviewed here, it is clear that experiments involving non-referential gestures that use ecologically valid tasks and materials have reported positive results.

In general, the present systematic review points to the need for further research to assess the role of non-referential gestures. In terms of methodology, experiment design must clearly take into account task appropriateness as well as the pragmatic function of non-referential gestures in discourse. This is because non-referential gestures emphasizing some spatial or event information can be perceived as unnatural in discourse, and thus it is important to assess which parts of the discourse the speaker should accompany with non-referential gestures. Conducting previous preliminary analyses can help to precisely define the visual features of gestures as well as their patterns of association with target words in a natural and spontaneous context (see the preliminary study used by [83]), thereby ensuring the ecological validity of the experimental materials.

All things considered, the evidence presented here would seem to support the view that the significant bootstrapping and predictive role of non-referential gestures is related to the pragmatic, discursive, and prosodic functions they perform in discourse. Non-referential gestures may serve important linguistic functions in discourse, associated with rhythmic marking, discourse structure marking, and information structure marking ([6,48,50,51,62,63,67], and others), such as new or accessible referents in discourse [62,63]. The developmental findings reported in the present review also lend support to the hypothesis that non-referential gestures develop in parallel with narrative development [54,57] (see also [58]). Importantly, non-referential gestures can help children focus on critical parts of a story, by providing them with visual markers that facilitate the parsing and processing of narrative discourse.

Though the studies here represent a first step in this direction, further research is needed to evaluate the potential of narrative training and sociopragmatic paradigms that include a strong multimodal component involving both referential and non-referential gestures. The fact that non-referential gestures are a strong discourse framing mechanism ([48,51,67], and others) is an indication that they might constitute a powerful tool for assessment in TD populations. We reviewed here two studies in which non-referential gestures were successfully used in narratives to improve the narrative production skills of TD children [83,84]. Moreover, future investigations could extend these findings to populations with language disorders. In this sense, the inclusion of non-referential gestures could be of benefit in language intervention programs for non-TD children. Reinforcing the production of these types of gestures during narratives might provide children with an important means of non-verbal discourse marking that can help them improve their narrative and interactional abilities.

Interestingly, some recent studies assessing multimodal training have suggested that such techniques can enhance children's social cognition and expressive pragmatic skills [89] and that having children observe audio-visual stimuli involving all kinds of gestures can improve their narrative productions [90]. On the one hand, the study by [89] showed that 3- to 4-year-old preschoolers improved their expressive pragmatic skill through training in which they were asked to embody mental states using prosodic and gestural cues. On the other hand, findings in [90] revealed that both 5-year-old children with early brain injury (who had difficulty in structuring narrative) and TD children were more likely to produce better-structured narrative retelling when the storyteller performed story-relevant gestures while speaking. All in all, additional research exploring the multimodal components of narrative and sociopragmatic treatments is needed. Our view is that the spontaneous use of gesture in discourse involves a combination of referential and non-referential gestures, and that non-referential gestures cannot be neglected as they have an important function in multimodal trainings.

Concerning the findings of the reported quantitative studies, we consider that future research could address the application of gesture-based narrative interventions under more specific populations, considering for example non-TD children's language and communication, which could be of help for clinicians, teachers, families, and researchers concerned with language development in such children. Various classroom training studies involving narratives have already been successfully carried out with preschool non-TD children (e.g., [91,92]). In this regard, [92] demonstrated that narrative interventions are a promising and effective strategy to teach oral narration to children with risk factors and narrative language delays, who may benefit from it in terms of their short-term and long-term narrative retelling skills (see also [93–96]). We claim that more effective interventions should include training focused on not only children's speech but also their gestures and general multimodal behavior. An example of this is the study by [91], which proposed an intervention based on activities combining voluntary storytelling with group story-acting carried out as a regular part of the preschool curriculum (see also [97], for the benefits in social competence of theatre-based intervention involving role-playing, improvisation, and play performance with autistic children). The results showed that story-acting training (i.e., story dictation and dramatization) promoted the abilities of preschool children from low-income and otherwise disadvantaged backgrounds in three major areas that contribute to their readiness for success in formal education, namely narrative and other oral language skills, emergent literacy, and social competence.

Previous systematic review and meta-analysis studies [98–100] have provided evidence that using social pragmatic, pragmatic language, and narrative interventions can support the social communication and language abilities of children with ASD or with other language disorders. However, to our knowledge, there are no studies that have assessed whether multimodal (gesture-based) training with non-TD children could also contribute to their language development. We claim that paying attention to the gestures that learners produce can help professionals determine any existing underlying delays in acquiring more complex linguistic or cognitive skills in populations with atypical development. As gestures are likely to give clues not yet evident in their speech about a learner's understanding of a task, this can help professionals determine whether learners are ready to take further steps in their learning. Also, gestures can help diagnose any existing language or cognitive difficulties (see [101–105] for reviews) that result in an atypical language profile (e.g., children with early brain injury, autism, Down syndrome, etc.; [106,107]). Because both narrative production and gesture can index individual differences in typical development profiles, a better understanding of gesture-speech development could help improve clinical practices regarding children's language assessment and intervention. All things considered, these reviewed studies may extend the findings of the TD children to non-TD children, by offering clinicians and speech-language therapists some guidance by highlighting the importance of including gestures in their cognitive and linguistic assessment tasks.

Although there are two training studies with TD children that have been reviewed in this paper [83,84], training studies conducted with non-TD children have to date not focused specifically on the role of multimodality. For instance, no previous training studies involving gestures as an empirical condition have assessed the value of narrative and sociopragmatic training in ASD or in language disorders [98–100]. The long-term effects of these interventions and the extent to which learning thus acquired is generalized to new contexts is largely unknown, and thus assessing multimodal interventions could help teachers, clinicians, speech-language therapists, and also families to adapt new teaching methodologies that emphasize the importance of the role of gestures in multimodal narrative abilities in children. This suggests that the present study would be aptly complemented by a systematic review covering research on multimodal interventions/training in both TD and non-TD populations.

In conclusion, the present systematic review should clarify the state of the art with regard to the link between non-referential gestures and children's language development. Based on this review, we feel that it is safe to claim that non-referential gestures are likely to be helpful in both children's cognitive development and their acquisition of complex linguistic skills, although further investigation is needed to confirm this conclusion. This impact could be deemed in both TD and extended to non-TD populations, as non-referential gestures can represent an important multimodal tool that can be used to build up and frame children's processing and production of complex language.

Author Contributions: I.V.-G. and P.P. contributed equally to the development of the research questions, the methodological design and the discussion of the results. I.V.-G. was in charge of writing the article, with substantive feedback from P.P. All have made a substantial, direct and intellectual contribution to the work, and approved it for publication. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Spanish Ministry of Science, Innovation and Universities, Agencia Estatal de Investigación, and Fondo Europeo de Desarrollo Regional, grant number PGC2018-097007-B-100; and by a grant awarded by the Generalitat de Catalunya, grant number 2017 SGR_971 to the Prosodic Studies Group.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: We gratefully acknowledge the anonymous reviewers and the editor for their helpful input and feedback on the content of this systematic review paper. We also thank Júlia Florit-Pons for her help in reviewing the final version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Goldin-Meadow, S. How gesture promotes learning throughout childhood. *Child Dev. Perspect.* **2009**, *3*, 106–111. [CrossRef]
2. Goldin-Meadow, S. How gesture works to change our minds. *Trends Neurosci. Educ.* **2014**, *3*, 4–6. [CrossRef]
3. Novack, M.; Goldin-Meadow, S. Learning from Gesture: How Our Hands Change Our Minds. *Educ. Psychol. Rev.* **2015**, *27*, 405–412. [CrossRef]
4. Kelly, S.D.; Özyürek, A.; Maris, E. Two Sides of the Same Coin: Speech and Gesture Mutually Interact to Enhance Comprehension. *Psychol. Sci.* **2010**, *21*, 260–267. [CrossRef] [PubMed]
5. Kendon, A. Gesticulation and speech: Two aspects of the process of utterance. In *The Relationship of Verbal and Nonverbal Communication*; Key, M.R., Ed.; De Gruyter Mouton: Berlin, Germany, 1980; pp. 207–227.
6. McNeill, D. *Hand and Mind: What Gestures Reveal about Thought*; University of Chicago Press: Chicago, IL, USA, 1992.
7. Iverson, J.M.; Goldin-Meadow, S. Gesture paves the way for language development. *Psychol. Sci.* **2005**, *16*, 368–371. [CrossRef]
8. Rowe, M.L.; Özçalişkan, Ş.; Goldin-Meadow, S. Learning words by hand: Gesture′s role in predicting vocabulary development. *First Lang.* **2008**, *28*, 182–199. [CrossRef]
9. Lüke, C.; Ritterfeld, U.; Grimminger, A.; Rohlfing, K.J.; Liszkowski, U. Integrated Communication System: Gesture and Language Acquisition in Typically Developing Children and Children With LD and DLD. *Front. Psychol.* **2020**, *11*, 1–13. [CrossRef]
10. Goldin-Meadow, S. *Hearing Gesture: How Our Hands Help Us Think*; Beknap Press of Harvard University Press: Cambridge, MA, USA, 2003.
11. Thal, D.J.; Tobias, S. Communicative gestures in children with delayed onset of oral expressive vocabulary. *J. Speech Lang. Hear. Res.* **1992**, *35*, 1281–1289. [CrossRef] [PubMed]
12. Charman, T.; Drew, A.; Baird, C.; Baird, G. Measuring early language development in preschool children with autism spectrum disorder using the MacArthur communicative development inventory (infant form). *J. Child Lang.* **2003**, *30*, 213–236. [CrossRef] [PubMed]
13. Luyster, R.J.; Kadlec, M.B.; Carter, A.; Tager-Flusberg, H. Language assessment and development in toddlers with autism spectrum disorders. *J. Autism Dev. Disord.* **2008**, *38*, 1426–1438. [CrossRef]
14. Özçaliskan, S.; Adamson, L.B.; Dimitrova, N. Early deictic but not other gestures predict later vocabulary in both typical development and autism. *Autism* **2016**, *20*, 754–763. [CrossRef]
15. Ramos-Cabo, S.; Vulchanov, V.; Vulchanova, M. Gesture and Language Trajectories in Early Development: An Overview From the Autism Spectrum Disorder Perspective. *Front. Psychol.* **2019**, *10*, 1–9. [CrossRef]
16. Hostetter, A.B. When do gestures communicate? A meta-analysis. *Psychol. Bull.* **2011**, *137*, 297–315. [CrossRef]
17. Aussems, S.; Kita, S. Seeing Iconic Gestures While Encoding Events Facilitates Children′s Memory of These Events. *Child Dev.* **2019**, *90*, 1123–1137. [CrossRef]
18. Cook, S.W.; Yip, T.K.; Goldin-Meadow, S. Gestures, but not meaningless movements, lighten working memory load when explaining math. *Lang. Cogn. Process.* **2012**, *27*, 594–610. [CrossRef]
19. Goldin-Meadow, S.; Nusbaum, H.; Kelly, S.D.; Wagner, S. Explaining math: Gesture lightens the load. *Psychol. Sci.* **2001**, *12*, 516–522. [CrossRef] [PubMed]
20. So, W.C.; Chen-Hui, C.S.; Wei-Shan, J.L. Mnemonic effect of iconic gesture and beat gesture in adults and children: Is meaning in gesture important for memory recall? *Lang. Cogn. Process.* **2012**, *27*, 665–681. [CrossRef]
21. Wagner, S.M.; Nusbaum, H.; Goldin-Meadow, S. Probing the mental representation of gesture: Is handwaving spatial? *J. Mem. Lang.* **2004**, *50*, 395–407. [CrossRef]
22. Bates, E. *Language and Context*; Academic Press: New York, NY, USA, 1976.
23. Bates, E.; Benigni, L.; Bretherton, I.; Camaioni, L.; Volterra, V. *The Emergence of Symbols: Cognition and Communication in Infancy*; Academic Press: New York, NY, USA, 1979.
24. Behne, T.; Liszkowski, U.; Carpenter, M.; Tomasello, M. Twelve-month-olds′ comprehension and production of pointing. *Br. J. Dev. Psychol.* **2012**, *30*, 359–375. [CrossRef] [PubMed]
25. Rohlfing, K.J.; Grimminger, A.; Lüke, C. An Interactive View on the Development of Deictic Pointing in Infancy. *Front. Psychol.* **2017**, *8*, 1–6. [CrossRef] [PubMed]
26. Butcher, C.; Goldin-Meadow, S. Gesture and the transition from one- to two-word speech: When hand and mouth come together. In *Language and Gesture*; McNeill, D., Ed.; Cambridge University Press: Cambridge, UK, 2000; pp. 235–258.

27. Esteve-Gibert, N.; Prieto, P. Infants temporally coordinate gesture-speech combinations before they produce their first words. *Speech Commun.* **2014**, *57*, 301–316. [CrossRef]
28. Goldin-Meadow, S.; Butcher, C. Pointing toward two-word speech in young children. In *Pointing: Where Language, Culture, and Cognition Meet*; Kita, S., Ed.; Lawrence Erlbaum Associates: Hillsdale, NJ, USA, 2003; pp. 85–107.
29. Özçalişkan, Ş.; Goldin-Meadow, S. Gesture is at the cutting edge of early language development. *Cognition* **2005**, *96*, B101–B113. [CrossRef]
30. Özçalişkan, Ş.; Goldin-Meadow, S. When gesture-speech combinations do and do not index linguistic change. *Lang. Cogn. Process.* **2009**, *24*, 190–217. [CrossRef]
31. Colonnesi, C.; Stams, G.J.J.; Koster, I.; Noom, M.J. The relation between pointing and language development: A meta-analysis. *Dev. Rev.* **2010**, *30*, 352–366. [CrossRef]
32. Baron-Cohen, S. Perceptual role taking and protodeclarative pointing in autism. *Br. J. Dev. Psychol.* **1989**, *7*, 113–127. [CrossRef]
33. Goodhart, F.; Baron-Cohen, S. How many ways can the point be made? Evidence from children with and without autism. *First Lang.* **1993**, *13*, 225–233. [CrossRef]
34. Mastrogiuseppe, M.; Capirci, O.; Cuva, S.; Venuti, P. Gestural communication in children with autism spectrum disorders during mother-child interaction. *Autism* **2015**, *19*, 469–481. [CrossRef] [PubMed]
35. LeBarton, E.S.; Goldin-Meadow, S.; Raudenbush, S. Experimentally induced increases in early gesture lead to increases in spoken vocabulary. *J. Cogn. Dev.* **2015**, *16*, 199–220. [CrossRef]
36. Acredolo, L.P.; Goodwyn, S.W. Symbolic gesturing in normal infants. *Child Dev.* **1988**, *59*, 450–466. [CrossRef]
37. Acredolo, L.P.; Goodwyn, S.W. Symbolic gesturing in language development: A case study. *Hum. Dev.* **1985**, *28*, 40–49. [CrossRef]
38. Iverson, J.M.; Capirci, O.; Caselli, M.C. From communication to language in two modalities. *Cogn. Dev.* **1994**, *9*, 23–43. [CrossRef]
39. Behne, T.; Carpenter, M.; Tomasello, M. Young Children Create Iconic Gestures to Inform Others. *Dev. Psychol.* **2014**, *50*, 2049–2060. [CrossRef] [PubMed]
40. Özçalişkan, Ş.; Goldin-Meadow, S. Is there an iconic gesture spurt at 26 months? In *Gesture Studies: Vol. 4. Integrating Gestures: The Interdisciplinary Nature of Gesture*; Stam, G., Ishino, M., Eds.; John Benjamins Publishing Company: Amsterdam, The Netherlands, 2011; pp. 163–174. [CrossRef]
41. Özçalişkan, Ş.; Gentner, D.; Goldin-Meadow, S. Do iconic gestures pave the way for children′s early verbs? *Appl. Psycholinguist.* **2014**, *35*, 1143–1162. [CrossRef]
42. Namy, L.L. Recognition of iconicity doesn′t come for free. *Dev. Sci.* **2008**, *11*, 841–846. [CrossRef]
43. Namy, L.L.; Campbell, A.; Tomasello, M. The Changing Role of Iconicity in Non-Verbal Symbol Learning: A U-Shaped Trajectory in the Acquisition of Arbitrary Gestures. *J. Cogn. Dev.* **2004**, *5*, 37–57. [CrossRef]
44. Dargue, N.; Sweller, N. Learning Stories Through Gesture: Gesture′s Effects on Child and Adult Narrative Comprehension. *Educ. Psychol. Rev.* **2020**, *32*, 249–276. [CrossRef]
45. Macoun, A.; Sweller, N. Listening and watching: The effects of observing gesture on preschoolers′ narrative comprehension. *Cogn. Dev.* **2016**, *40*, 68–81. [CrossRef]
46. Stites, L.J.; Özçalişkan, Ş. Who Did What to Whom? Children Track Story Referents First in Gesture. *J. Psycholinguist. Res.* **2017**, *46*, 1019–1032. [CrossRef]
47. Demir, Ö.E.; Levine, S.C.; Goldin-Meadow, S. A tale of two hands: Children′s early gesture use in narrative production predicts later narrative structure in speech. *J. Child Lang.* **2015**, *42*, 662–681. [CrossRef]
48. Prieto, P.; Cravotta, A.; Kushch, O.; Rohrer, P.; Vilà-Giménez, I. Deconstructing Beat Gestures: A Labelling Proposal. In Proceedings of the 9th International Conference on Speech Prosody, Poznań, Poland, 13–16 June 2018; Klessa, K., Bachan, J., Wagner, A., Karpiński, M., Śledziński, D., Eds.; pp. 201–205. [CrossRef]
49. Rohrer, P.L.; Vilà-Giménez, I.; Florit-Pons, J.; Esteve-Gibert, N.; Ren, A.; Shattuck-Hufnagel, S.; Prieto, P. The MultiModal MultiDimensional (M3D) Labelling Scheme for theAnnotation of Audiovisual Corpora. In Proceedings of the 7th Gesture and Speech in Interaction (GESPIN), KTH Speech, Music & Hearing and Språkbanken Tal, Stockholm, Sweden, 7–9 September 2020.
50. Shattuck-Hufnagel, S.; Prieto, P. Dimensionalizing Co-Speech Gestures. In Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia, 5–9 August 2019; Calhoun, S., Escudero, P., Tabain, M., Warren, P., Eds.; Australasian Speech Science and Technology Association Inc.: Canberra, Australia, 2019; pp. 1490–1494.
51. Shattuck-Hufnagel, S.; Ren, A.; Mathew, M.; Yuen, I.; Demuth, K. Non-referential gestures in adult and child speech: Are they prosodic? In *Speech Prosody 2016*; Barnes, J., Brugos, A., Shattuck-Hufnagel, S., Veilleux, N., Eds.; International Speech Communication Association: Boston, MA, USA, 2016; pp. 836–839. [CrossRef]
52. Levy, E.T.; McNeill, D. Narrative development as symbol formation: Gestures, imagery and the emergence of cohesion. *Cult. Psychol.* **2013**, *19*, 548–569. [CrossRef]
53. Nicoladis, E.; Mayberry, R.I.; Genesee, F. Gesture and early bilingual development. *Dev. Psychol.* **1999**, *35*, 514–526. [CrossRef]
54. Colletta, J.M.; Pellenq, C.; Guidetti, M. Age-related changes in co-speech gesture and narrative: Evidence from French children and adults. *Speech Commun.* **2010**, *52*, 565–576. [CrossRef]
55. Mathew, M.; Yuen, I.; Demuth, K. Talking to the beat: Six-year-olds′ use of stroke-defined non-referential gestures. *First Lang.* **2018**, *38*, 1–18. [CrossRef]
56. Blake, J.; Myszczyszyn, D.; Jokel, A.; Bebiroglu, N. Gestures accompanying speech in specifically language-impaired children and their timing with speech. *First Lang.* **2008**, *28*, 237–253. [CrossRef]

57. Colletta, J.M.; Guidetti, M.; Capirci, O.; Cristilli, C.; Demir, Ö.E.; Kunene-Nicolas, R.N.; Levine, S. Effects of age and language on co-speech gesture production: An investigation of French, American, and Italian children′s narratives. *J. Child Lang.* **2015**, *42*, 122–145. [CrossRef]
58. Florit-Pons, J.; Vilà-Giménez, I.; Rohrer, P.L.; Prieto, P. The Development and Temporal Integration of Co-Speech Gesture in Narrative Speech: A Longitudinal Study. In Proceedings of the 7th Gesture and Speech in Interaction (GESPIN), KTH Speech, Music & Hearing and Språkbanken Tal, Stockholm, Sweden, 7–9 September 2020.
59. Graziano, M. Rapporto fra lo sviluppo della competenza verbale e gestuale nella costruzione di un testo narrativo in bambini dai 4 ai 10 anni. Ph.D. Thesis, SESA—Scuola Europea di Studi Avanzati—Università degli Studi "Suor Orsola Benincasa", Napoli, Italy, Université Stendhal, Grenoble, France. 2009. Unpublished Doctoral Dissertation.
60. Abner, N.; Cooperrider, K.; Goldin-Meadow, S. Gesture for Linguists: A Handy Primer. *Lang. Linguist. Compass* **2015**, *9*, 437–451. [CrossRef] [PubMed]
61. Alibali, M.W.; Heath, D.C.; Myers, H.J. Effects of visibility between speaker and listener on gesture production: Some gestures are meant to be seen. *J. Mem. Lang.* **2001**, *44*, 169–188. [CrossRef]
62. Dimitrova, D.; Chu, M.; Wang, L.; Özyürek, A.; Hagoort, P. Beat that word: How listeners integrate beat gesture and focus in multimodal speech discourse. *J. Cogn. Neurosci.* **2016**, *28*, 1255–1269. [CrossRef] [PubMed]
63. Im, S.; Baumann, S. Probabilistic relation between co-speech gestures, pitch accents and information status. *Proc. Linguist. Soc. Am.* **2020**, *5*, 685–697. [CrossRef]
64. Graziano, M. The development of two pragmatic gestures of the so-called *Open Hand Supine family* in Italian children. In *From Gesture in Conversation to Visible Action as Utterance: Essays in Honor of Adam Kendon*; Seyfeddinipur, M., Gullberg, M., Eds.; John Benjamins Publishing Company: Amsterdam, The Netherlands, 2014; pp. 311–330. [CrossRef]
65. Graziano, M. Gestures in Southern Europe: Children′s pragmatic gestures in Italy. In *Body—Language— Communication: An International Handbook on Multimodality in Human Interaction*; Müller, C., Cienki, A., Fricke, E., Ladewig, S.H., McNeill, D., Bressem, J., Eds.; De Gruyter Mouton: Berlin, Germany, 2014; Volume 2, pp. 1253–1258.
66. Kendon, A. *Gesture: Visible Action as Utterance*; Cambridge University Press: Cambridge, UK, 2004.
67. Kendon, A. Pragmatic functions of gestures. Some observations on the history of their study and their nature. *Gesture* **2017**, *16*, 157–175. [CrossRef]
68. Vilà-Giménez, I.; Dowling, N.; Demir-Lira, Ö.E.; Prieto, P.; Goldin-Meadow, S. The predictive value of non-referential beat gestures: Early use in parent-child interactions predicts narrative abilities at 5 years of age. *Child Dev.*. under review.
69. Demir Ö, E.; Küntay, A. Narrative Development. In *Encyclopedia of Language Development*; Brooks, P., Kampe, V., Eds.; SAGE: Thousand Oaks, CA, USA, 2014; pp. 393–397.
70. Naremore, R.; Densmore, A.E.; Harman, D. *Language Intervention with School Age Children: Conversation, Narrative and Text*; San Singular Publishing Group: Diego, CA, USA, 1995.
71. Demir, Ö.E.; Levine, S.C.; Goldin-Meadow, S. Getting on track early for school success: The STEP assessment system to support effective instruction (PreK-3rd). Creating an assessment of literacy skills for three and four year olds. Foundation for Child Development 2012. Available online: http://fcd-us.org/sites/default/files/Lit%20Tech%20Report.pdf. (accessed on 12 January 2020).
72. Liberati, A.; Altman, D.G.; Tetzlaff, J.; Mulrow, C.; Gøtzsche, P.C.; Ioannidis JP, A.; Clarke, M.; Devereaux, P.J.; Kleijnen, J.; Moher, D. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *PLoS Med.* **2009**, *6*, e1000100. [CrossRef]
73. Moher, D.; Liberati, A.; Tetzlaff, J.; Altman, D.G. The PRISMA Group Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Med.* **2009**, *6*, e1000097. [CrossRef]
74. Kmet, L.M.; Lee, R.C.; Cook, L.S. *Standard Quality Assessment Criteria for Evaluating Primary Research Papers from a Variety of Fields*; Alberta Heritage Foundation for Medical Research (AHFMR) AHFMR—-HTA Initiative #13: Edmonton, AB, Canada, 2004.
75. Igualada, A.; Esteve-Gibert, N.; Prieto, P. Beat gestures improve word recall in 3- to 5-year-old children. *J. Exp. Child Psychol.* **2017**, *156*, 99–112. [CrossRef]
76. Llanes-Coromina, J.; Vilà-Giménez, I.; Kushch, O.; Borràs-Comes, J.; Prieto, P. Beat gestures help preschoolers recall and comprehend discourse information. *J. Exp. Child Psychol.* **2018**, *172*, 168–188. [CrossRef] [PubMed]
77. Austin, E.; Sweller, N. Presentation and production: The role of gesture in spatial communication. *J. Exp. Child Psychol.* **2014**, *122*, 92–103. [CrossRef]
78. Austin, E.; Sweller, N. Getting to the elephants: Gesture and preschoolers′ comprehension of route direction information. *J. Exp. Child Psychol.* **2017**, *163*, 1–14. [CrossRef]
79. Kartalkanat, H.; Göksun, T. The effects of observing different gestures during storytelling on the recall of path and event information in 5-year-olds and adults. *J. Exp. Child Psychol.* **2020**, *189*, 104725. [CrossRef] [PubMed]
80. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*; Routledge Academic: New York, NY, USA, 1988.
81. Vilà-Giménez, I.; Demir-Lira Ö, E.; Prieto, P. The Role of Referential Iconic and Non-Referential Beat Gestures in Children′s Narrative Production: Iconics Signal Oncoming Changes in Speech. In Proceedings of the 7th Gesture and Speech in Interaction (GESPIN), KTH Speech, Music & Hearing and Språkbanken Tal, Stockholm, Sweden, 7–9 September 2020.
82. Ferré, G. Functions of three open-palm hand gestures. *J. Multimodal Commun.* **2011**, *1*, 5–20. [CrossRef]

83. Vilà-Giménez, I.; Igualada, A.; Prieto, P. Observing Storytellers Who Use Rhythmic Beat Gestures Improves Children′s Narrative Discourse Performance. *Dev. Psychol.* **2019**, *55*, 250–262. [CrossRef]
84. Vilà-Giménez, I.; Prieto, P. Encouraging kids to beat: Children′s beat gesture production boosts their narrative performance. *Dev. Sci.* **2020**, *23*, 1–14. [CrossRef]
85. Rohrer, P.L.; Delais-Roussarie, E.; Prieto, P. Beat gestures for comprehension and recall: Differential effects of language learners and native listeners. *Front. Psychol.* **2020**, *11*, 1–16. [CrossRef] [PubMed]
86. Kushch, O.; Prieto, P. The effects of pitch accentuation and beat gestures on information recall in contrastive discourse. In *Speech Prosody 2016*; Barnes, J., Brugos, A., Shattuck-Hufnagel, S., Veilleux, N., Eds.; International Speech Communication Association: Boston, MA, USA, 2016; pp. 922–925. [CrossRef]
87. Morett, L.M.; Fraundorf, S.H. Listeners consider alternative speaker productions in discourse comprehension and memory: Evidence from beat gesture and pitch accenting. *Mem. Cogn.* **2019**, *47*, 1515–1530. [CrossRef]
88. Feyereisen, P. Further investigation on the mnemonic effect of gestures: Their meaning matters. *Eur. J. Cogn. Psychol.* **2006**, *18*, 185–205. [CrossRef]
89. Pronina MHübscher, I.; Holler, J.; Prieto, P. Interactional training interventions boost children′s expressive pragmatic abilities: Evidence from a novel multidimensional testing approach. *Cogn. Dev.* **2021**, *57*, 101003, in press. [CrossRef]
90. Demir, Ö.E.; Fisher, J.A.; Goldin-Meadow, S.; Levine, S.C. Narrative processing in typically developing children and children with early unilateral brain injury: Seeing gesture matters. *Dev. Psychol.* **2014**, *50*, 815–828. [CrossRef]
91. Nicolopoulou, A.; Cortina, K.S.; Ilgaz, H.; Cates, C.B.; de Sá, A.B. Using a narrative- and play-based activity to promote low-income preschoolers′ oral language, emergent literacy, and social competence. *Early Child. Res. Q.* **2015**, *31*, 147–162. [CrossRef] [PubMed]
92. Spencer, T.D.; Slocum, T.A. The Effect of Narrative Intervention on Story Retelling and Personal Story Generation Skills of Preschoolers With Risk Factors and Narrative Language Delays. *J. Early Interv.* **2010**, *32*, 178–199. [CrossRef]
93. Bunning, K.; Gooch, K.; Johnson, M. Developing the personal narratives of children with complex communication needs associated with intellectual disabilities: What is the potential of Storysharing®? *J. Appl. Res. Intellect. Disabil.* **2017**, *30*, 743–756. [CrossRef] [PubMed]
94. Fey, M.E.; Finestack, L.H.; Gajewski, B.J.; Popescu, M.; Lewine, J.D. A Preliminary Evaluation of Fast For Word-Language as an Adjuvant Treatment in Language Intervention. *J. Speech Lang. Hear. Res.* **2010**, *53*, 430–449. [CrossRef]
95. Hettiarachchi, S. The effectiveness of Colourful Semantics on narrative skills in children with intellectual disabilities in Sri Lanka. *J. Intellect. Disabil.* **2016**, *20*, 18–33. [CrossRef] [PubMed]
96. Swanson, L.A.; Fey, M.E.; Mills, C.E.; Hood, L.S. Use of narrative-based language intervention with children who have specific language impairment. *Am. J. Speech-Lang. Pathol.* **2005**, *14*, 131–143. [CrossRef]
97. Corbett, B.A.; Key, A.P.; Qualls, L.; Fecteau, S.; Newsom, C.; Coke, C.; Yoder, P. Improvement in social competence using a randomized trial of a theatre intervention for children with autism spectrum disorder. *J. Autism Dev. Disord.* **2016**, *46*, 658–672. [CrossRef]
98. Binns, A.V.; Cardy, J.O. Developmental social pragmatic interventions for preschoolers with autism spectrum disorder: A systematic review. *Autism Dev. Lang. Impair.* **2019**, *4*, 1–18. [CrossRef]
99. Favot, K.; Carter, M.; Stephenson, J. The effects of oral narrative intervention on the narratives of children with language disorder: A systematic literature review. *J. Dev. Phys. Disabil.* **2020**, 1–48. [CrossRef]
100. Parsons, L.; Cordier, R.; Munro, N.; Joosten, A.; Speyer, R. A systematic review of pragmatic language interventions for children with autism spectrum disorder. *PLoS ONE* **2017**, *12*, e0172242. [CrossRef]
101. Cartmill, E.A.; Demir, Ö.E.; Goldin-Meadow, S. Studying gesture. In *Research Methods in Child Language: A Practical Guide*; Hoff, E., Ed.; Blackwell Publishing: Oxford, UK, 2012; pp. 208–225.
102. Goldin-Meadow, S. Gesture as a window onto communicative abilities: Implications for diagnosis and intervention. *Perspect. Lang. Learn. Educ.* **2015**, *22*, 50–60. [CrossRef]
103. Goldin-Meadow, S. Taking a Hands-on Approach to Learning. *Policy insights Behav. Brain Sci.* **2018**, *5*, 163–170. [CrossRef]
104. Goldin-Meadow, S.; Alibali, M.W. Gesture′s role in speaking, learning, and creating language. *Annu. Rev. Psychol.* **2013**, *64*, 257–283. [CrossRef]
105. Goldin-Meadow, S.; Iverson, J.M. Gesturing across the life span. In *The Handbook of Life-Span Development. Cognition, Biology, and Methods*; Overton, W.F., Lerner, R.M., Eds.; John Willey & Sons Inc.: Hoboken, NJ, USA, 2010; Volume 1, pp. 754–791. [CrossRef]
106. Sauer, E.; Levine, S.C.; Goldin-Meadow, S. Early gesture predicts language delay in children with pre- and perinatal brain lesions. *Child Dev.* **2010**, *81*, 528–539. [CrossRef]
107. Iverson, J.M.; Longobardi, E.; Caselli, M.C. Relationship between gestures and words in children with Down′s syndrome and typically developing children in the early stages of communicative development. *Int. J. Lang. Commun. Disord.* **2003**, *38*, 179–197. [CrossRef] [PubMed]

children

MDPI

Article

Impact of the Use of Media Devices within the Family Context on the Language of Preteens

Lucía Acebedo [1], Lucía Buil-Legaz [2], Daniel Adrover-Roig [2] and Eva Aguilar-Mediavilla [2,*]

1 Faculty of Education, University of Balearic Islands, 07122 Palma, Spain; lucia.acebedo@ceipsaindioteria.cat
2 Applied Pedagogy and Educational Psychology, Institute of Research and Innovation in Education (IRIE), University of Balearic Islands, 07122 Palma, Spain; lucia.buil@uib.es (L.B.-L.); daniel.adrover@uib.es (D.A.-R.)
* Correspondence: eva.aguilar@uib.es; Tel.: +34-971-172-566

Received: 12 November 2020; Accepted: 6 December 2020; Published: 9 December 2020

Abstract: Several studies have found a negative impact of media use on the language of children under 5 years. This impact seems to be related to the linguistic input of their parents. However, less is known about the influence of media on language in preteens. This study aims to analyze the relationship between the use of media, the quantity of parental language input, and the linguistic level of preteens. We assessed the language level of 60 bilingual Spanish–Catalan preteens aged 11–12 years with four subtests of the Spanish version of the standardized clinical evaluation of language fundaments (CELF-5-Spanish) as well as media use at home through a multiple-choice questionnaire. Results showed lower language scores in preteens who had access to more media devices, who used them more frequently, and who talked less with their parents. Language scores were also significantly lower in preteens who used media devices to communicate compared to those who used it as a school aid or to learn new things. These results are not influenced by socioeconomic level, sex, chronological age, or family language. The present results highlight the negative impact of media use on the language level of older children, which is also related to the amount of linguistic input received from their parents.

Keywords: video games; input; education; adolescent; media; language acquisition; preteens

1. Introduction

Nowadays, children and adolescents use media and technology in almost every area of their lives, such as social life and education contexts, being thus in contact with technological devices since very early in their development. We live in a society in which media is almost omnipresent and, in the case of children, it is also part of their learning process, which makes this topic of high scientific interest. In this vein, Prensky [1] labels "digital natives" those children and adolescents who are born and grow up with media devices fully available, which they use with a distinctive naturalness in comparison to other generations.

This situation happens to an extent in which even the youngest children and adolescents are exposed to media within their family context. Recent studies show that many parents use media for entertaining their children, which leads to long exposure periods to TV, tablets, and multiple technological devices [2]. Accordingly, seven out of ten Spanish children (71%) are exposed to media while eating—watching TV, or even using a tactile screen or a smartphone—as reported in the 5th CinfaSalud Study [3], which has been certified by the *Sociedad Española de Pediatría Extrahospitalaria y Atención Primaria* (SEPEAP; Spanish Society of outpatient paediatrics and primary attention).

At older ages, the use of devices increases, especially for watching videos and playing video games, since up to 98% of children between 10 and 14 years of age have a smartphone with an internet connection [4]. As suggested by the results presented by the Spanish National Institute of Statistics, the use of these devices increases rapidly from 26.1% of use at 10 years of age to 98% at 14 years of age [5,6].

As shown, several authors recognize that the use of technological devices is progressively becoming more common in our daily life. Even in the educational setting, the Spanish *Real Decreto* (Royal Ordinance) by which the basic curriculum in Primary Education is established [7] highlights the importance of CIT (Communication and Information Technology) at this educative stage, affirming that one of the seven basic competencies that a student must acquire is digital competence.

While the use of technological devices increases, language development and literacy seems to be affected. In this sense, some authors have suggested that the use of media can act as a strong predictor of cognitive delay [8]. For young adults, recent studies [9] have pointed out that current university students present a weakened expressive language, exhibiting errors that are more typical of earlier educative stages, mostly, grammatical mistakes. Moreover, some studies emphasize two environmental factors that can cause language delays: a diminished verbal parent–child interaction and children exposure to technological media (TV, videos, computers), both having clear implications for language development [10]. However, results are inconsistent, since there are studies that found that media exposure could also have positive effects on language development. In this sense, several authors report that viewing so-called educational programs at an early age of language acquisition could contribute to increasing vocabulary [11,12], although the acquisition of vocabulary seems to occur when the use of technologies is accompanied by the interaction with parents or caregivers [10]. Studies have even shown that using media for reading-related activities can be beneficial for learning [13] and that in certain cases, the use of certain applications can improve working memory [14]. Recent studies recognize that there is an excessive use of media and point to different adverse effects that this phenomenon can have on children and adolescents who abuse it [4,15]. As Christakis [16] pointed out, exposure to TV, and other audio-visual devices, is one of the main risk factors associated with language development in children under 5 years of age. In this sense, studies point out that the time spent with technological devices could be detrimental to the children's language skills because it may displace language-enhancing activities and interaction with parents, resulting in less exposure to language input [17]. Previous studies describe a clear relation between watching TV from early ages (under 24 months of age) and language development problems [18]. For instance, as seen in a sample of 45 children, ranging from 18 months to 5 years of age, watching television from early ages, specifically before 2 years of age for more than 2 h per day, is considered as a risk factor for primary language development delays [2]. Moreover, this same study described an increment in the previous years of the number of requests for diagnostic evaluations regarding language maturation delays. In the same line, the American Academy of Pediatrics (APP) [19] found that children who start watching TV before they turn one year of age for more than 2 h per day are six times more likely to develop language problems. Likewise, Zimmerman, Christakis, and Meltzoff [12] demonstrated in a sample of children under 17 months of age that each daily hour watching television implied, on average, a diminution of 17 points in the *MacArthur-Bates Communicative Development Inventories* [20,21]. Another interesting longitudinal study, initiated in 2010 with more than 250 families, assessed the influence of screen exposure on developmental trajectories from childhood to adolescence. Their results showed that children already exposed to screens when they were 6 months old exhibited a lower cognitive and language development at 14 months of age, only 8 months later. Furthermore, no differences were found in terms of the content, educative or non-educative, to which they had been exposed [22].

Furthermore, the interaction quality, both verbal and non-verbal, between children and their parents is diminished by the simple presence of a nearby switched-on TV [23], suggesting an impact of screen presence on children even when they are not paying close attention to it. For this reason, several authors such as Pempek, Kirkorian, and Anderson [24] pointed out that television in the background can affect the quantity and quality of parent's infant-direct speech, this being closely linked to children's language development. These authors hypothesize that this negative influence can be generalized to all information and entertainment technological devices used by parents in the presence of their children. Hence, it is not only children's exposition to screens and digital devices that is harmful to their development but also their parents' exposition when they are present.

Therefore, the studies cited above suggest that media can affect language development, which is an interaction that seems mediated by the diminution of both the quantity and quality of the linguistic input provided by parents when media devices are present. Language stimulation is important since birth, and the quality of the linguistic input from caregivers, who are the first and primary interlocutors, is crucial for its adequate acquisition. As a child grows up, the relation between her or his language input and linguistic productions becomes more obvious, for the language that he or she learns depends on the input received from the environment over the years [25,26]. There is plenty of evidence regarding the importance of language input on children's overall development, particularly in terms of language [27–29], and it seems that the use of media can act as a risk factor for this input when it is either reduced or impoverished. Nevertheless, some authors indicate that moderate amounts of media exposure may not be a negative influence on children's language development. In this sense, it seems that co-viewing could act as a buffer regarding the relationship between media use and early language skills [30].

To summarize, the presence of technological devices within the family and educative settings can negatively affect language at early ages, because it can diminish the quantity and quality of linguistic input received by the child. However, studies assessing this topic in older children and adolescents are still needed.

Given the exponential growth of the use of media within the family and educative settings during adolescence, the importance of both parent's language quantity and their implication on development [31–33], and the presence of studies suggesting that young adults' language level is worsening, we sought to analyze the relationship between the use of media, the quantity of parental language input, and preteens' language development. We aimed to explore the potential relationship between media use and the linguistic level of preteens of 11 and 12 years of age. Furthermore, we evaluated the association between the quantity of parental language input, as reported by preteens, and their actual language level.

2. Materials and Methods

2.1. Participants

The total sample included 60 preteens from different schools in Mallorca (Spain). All the participants were in the 5th or 6th grade of Primary Education (age in years: $M = 11.63$, $SD = 0.486$). Twenty-two of them (3 females) were 11 years old, and 38 (14 females) were 12 years old. All preteens were Catalan–Spanish bilinguals and used Catalan and Spanish as the school language. The language used in their family context was mostly Spanish except for seven participants who used mainly Catalan with their families. Regarding the socioeconomic status of the families, and considering the Spanish socioeconomic context, 33 participants fell into the middle category, and the other 27 were in the low category [34]. Hence, it could be considered that all families could afford to acquire technological devices and could use them daily if they wanted to.

None of the participants in our sample had a diagnosis of learning difficulties nor presented educative education needs. A summary of the sociodemographic data of the sample can be consulted in Table 1.

This study is part of a project funded by FEDER/Ministerio de Ciencia, Innovación y Universidades_Agencia Estatal de Investigación/EDU2017-85909-P, and it was approved by the Ethics Committee of the University of Balearic Islands (12 September 2017). All parents of the participants included in the sample provided written informed consent at the beginning of the study, all preteens consented as well to participate, and pertinent measures have been followed to maintain their anonymity.

Table 1. Sociodemographic data of the participants.

	Frequency
Age	
11	22
12	38
Sex	
Female	17
Male	43
Socioeconomic status	
Low	27
Middle	33
High	0
Family predominant language	
Spanish	53
Catalan	7

2.2. Study Design

The present study followed a transversal design, and participants were evaluated with two tests that gathered information on the frequency and types of media use, language interaction with their parents, and participants' expressive and receptive linguistic level.

2.3. Materials

A standardized validated questionnaire was used to obtain data regarding participant's language level, and an ad hoc questionnaire served to quantify the frequency and type of media use within the family setting and the interaction frequency with their parents.

2.3.1. CELF-5

To assess language level, four subtests from the Spanish version of the standardized clinical evaluation of language fundaments (CELF-5) were administered [35].

The CELF-5 is a clinical instrument designed for individual application that is composed of different subtests that allow assessing multiple linguistic aptitudes. Specifically, it is formed of 12 subtests and includes several complementary resources. For the present study, we used the four subtests that compound the Core Language Score (CLS) of the original Spanish version designed for participants between 9 and 15 years of age, which were:

- Word classes: This subtest evaluates the participant's ability to understand relationships between words based on their semantic features. The participant must select the two words that are the most semantically related from a list of four words (e.g., *run*, *jump*, *read*, *listen*; the correct answers would be "run" and "jump"). This subtest includes 40 items.
- Recalling sentences: This subtest consists of repeating oral sentences of increasing length and complexity. It is comprised of 26 sentences, allowing the assessment of morphosyntactic aptitudes and phonological working memory capacity.
- Formulated sentences: The participant must orally formulate complete and appropriate sentences of increasing complexity using two given words. This subtest, which assesses the capacity to integrate semantic, syntactic, and pragmatic information, consists of 24 items in which an image and a related word are presented to the participant, who must elaborate a sentence that relates both items.
- Semantic relationships: This subtest evaluates the ability to understand sentences by either comparing or identifying related elements. A total of 20 items are presented. In each of them, the participant is asked to choose which two words out of a total of 4 words presented both orally and visually are semantically more suitable to answer to a given question or to complete a sentence (e.g., *One hour is longer than … (a) a minute; (b) a day; (c) a second; (d) a morning*).

The CELF-5 subtests correction was conducted following its standard guidelines. Direct scores for each subtest were transformed into scaled scores to control for the influence of chronological age on participant's results. The mean of the scaled scores is 10 with a standard deviation of 3, meaning that scores below 7 would be indicative of language difficulties.

2.3.2. Media Use Questionnaire

To collect data on media and media devices use, we applied a questionnaire elaborated ad hoc (see Appendix A). It was composed of six items that were designed to gather information about how many technological devices are used daily by the participants. Moreover, respondents are asked about when, how often, for how long they use them, and the usefulness they perceive in them. This questionnaire also included items on their frequency of linguistic interaction with their parents (see Appendix A).

2.4. Procedure

First, we contacted the schools to request their participation in the study. Before starting the assessment, we requested that the parents of all participants included in the sample sign the written consent form.

We first administered individually the four CELF-5 subtests in Spanish, since these were already available and published in this language. After the evaluator ensured that the child had understood the subtest's instructions, the assessment began with several practice items. After the CELF-5, we administered the self-reported written questionnaire of media use in-group sessions during the school hours.

Analyses were performed using SPSS-25. The relation between media use, the frequency of linguistic interaction with parents, and preteens' language level was assessed using non-parametric statistics because assumptions for parametric analyses were not met. Independent sample comparisons, Chi-squared tests, and correlations were applied depending on the types of variables considered.

3. Results

3.1. Descriptive Data

We first conducted independent sample comparisons to ensure that language outcomes were not affected by potential confounding variables, such as sex, family language, or socioeconomic status. No significant differences were found between male and female participants (Word classes: $U = 384$, $p = 0.759$; Formulated sentences: $U = 324$, $p = 0.486$; Recalling sentences: $U = 382$, $p = 0.784$; Semantic relationships: $U = 324$, $p = 0.489$). Family language did not affect language level, and no significant differences were found between participants who had Spanish or Catalan as their family language (Word classes: $U = 217.5$, $p = 0.456$; Formulated sentences: $U = 230.5$, $p = 0.288$; Recalling sentences: $U = 231.5$, $p = 0.283$; Semantic relationships: $U = 235.5$, $p = 0.242$). In addition, no significant differences were found between preteens pertaining to the middle socioeconomic status and those falling into the middle–low class in terms of language level (Word classes: $U = 398.5$, $p = 0.48$; Formulated sentences: $U = 451.5$, $p = 0.927$; Recalling sentences: $U = 405.5$, $p = 0.542$; Semantic relations: $U = 453.5$, $p = 0.903$).

Regarding media use, participants used on average 2.87 technological devices daily, TV being the most frequently used ($n = 60$), followed by smartphones ($n = 52$), computers ($n = 24$), tablets ($n = 18$), PlayStation ($n = 15$), and Nintendo Switch ($n = 2$). See Table 2 for more details about the mean scaled scores and standard deviations of CELF-5 subtests as well as the number of technological devices used.

Table 2. Mean scaled scores and standard deviations for language subtests and the number of devices used by the complete sample (n = 60).

	Mean	SD
CELF-5		
Word classes	9.05	2.740
Formulated sentences	8.72	2.762
Recalling Sentence	8.47	2.771
Semantic relationships	8.43	2.459
Number of devices used	2.87	1.186

Note: SD = standard deviation.

These devices were mostly used every day, for more than an hour, to watch TV and to play; however, nearly half of the participants did not use any device while eating. Similarly, a little under 50% of the participants stated that they talk much with their parents (see Table 3).

Table 3. Frequency and percentage of media use and communication with parents.

	Frequency	Percentage
Media use		
Daily	40	66.6%
Only weekends	20	33.3%
Usage time		
Less than half an hour	3	5%
Half an hour	11	18.3%
One hour	15	25%
More than one hour	31	51.6%
Use given to media devices		
Assistance in school tasks	14	23.3%
Learning new things	10	16.6%
Playing games	26	43.3%
Communication	10	16.6%
Devices used while eating		
None	33	55%
Television	19	31.6%
Other devices	8	13.3%
Talking a lot with their parents		
No	32	53.3%
Yes	28	46.6%

3.2. Media Use and Language Level

To explore the potential association between media use and language level, a set of correlation analyses were conducted. Results showed a significant negative correlation between the number of technological devices used daily and language scores for all the four CELF-5 subtests (see Table 4). Furthermore, we found even stronger negative associations between the time of use of these devices and the CELF-5 scores.

Following Figure 1, language scores were significantly lower for preteens who used technological devices every day (n = 40) in comparison to those who used them only during the weekends (n = 20) (Word classes: $U = 606.5$, $p < 0.0001$, $r = 0.423$; Formulated sentences: $U = 725$, $p < 0.0001$, $r = 0.674$; Recalling sentences: $U = 672$, $p < 0.0001$, $r = 0.557$; Semantic relationships: $U = 741$, $p < 0.0001$, $r = 0.701$). It is also worth noting that participants in the two groups were equivalent in terms of age ($U = 320$, $p = 0.133$).

Table 4. Spearman correlations between the number of devices, time of media use, and language scores.

	1	2	3	4	5	6
1. Number of devices	1					
2. Time of media use [a]	0.507 **	1				
3. Word classes	−0.273 *	−0.347 *	1			
4. Formulated sentences	−0.395 **	−0.714 **	0.499 **	1		
5. Recalling sentences	−0.358 **	−0.504 **	0.844 **	0.732 **	1	
6. Semantic relationships	−0.395 **	−0.776 **	0.523 **	0.945 **	0.76 **	1

[a] Time of use per day; * $p < 0.05$; ** $p < 0.01$.

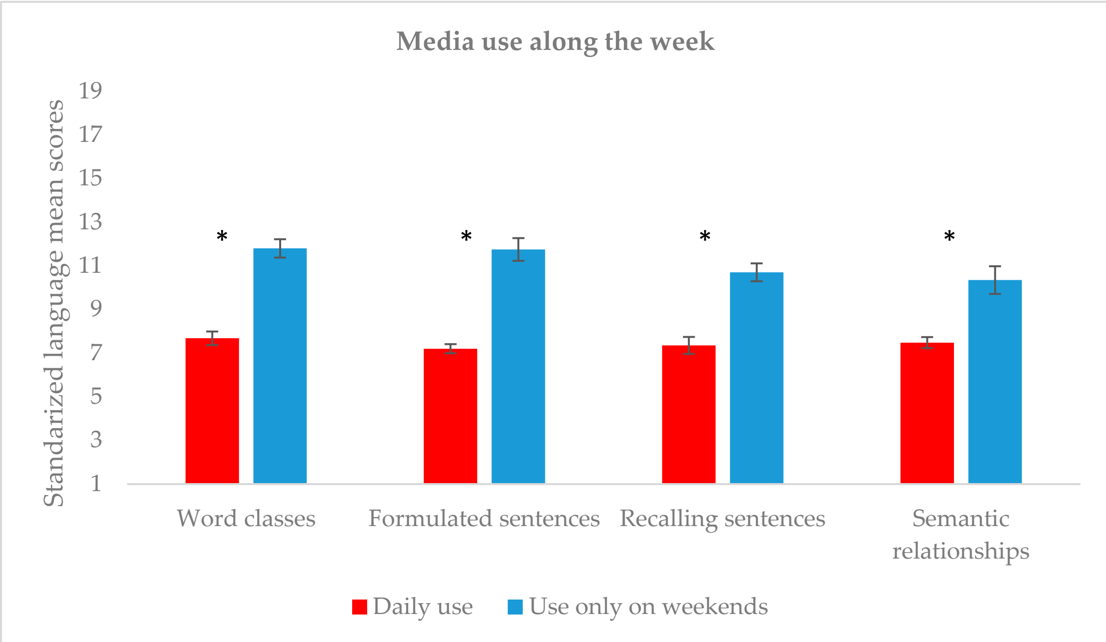


Figure 1. Mean scores on language tasks and use of media along the week. * $p < 0.001$.

Regarding the use of media while eating (see Figure 2), results revealed significantly lower language scores in participants who used media devices during this activity ($n = 27$), as compared to those who did not use them ($n = 33$) in two of the CELF-5 subtests (Recalling sentences: $U = 584.5$, $p = 0.036$, $r = 0.270$; Semantic relationships: $U = 589.5$, $p = 0.030$, $r = 0.280$). Similarly, the subtest Formulated sentences showed a tendency ($U = 570$, $p = 0.058$), and Word classes did not differ between both groups ($U = 554.5$, $p = 0.101$). Again, both groups did not differ with respect to their ages ($U = 508.5$, $p = 0.262$).

To explore the potential differences in language level according to the use given to media devices, independent sample comparisons were conducted with the use given to media as the between-subjects factor (scholar assistance, learning new things, to play, to communicate) on all CELF-5 subtests (see Figure 3). Results yielded significant differences in three of the CELF-5 scores (Formulated sentences: $H = 13.909$, $p = 0.003$, ${}_{\varepsilon}{}^{2} = 0.235$; Recalling sentences: $H = 13.909$, $p = 0.003$, ${}_{\varepsilon}{}^{2} = 0.15$; Semantic relationships: $H = 13.909$, $p = 0.003$, ${}_{\varepsilon}{}^{2} = 0.268$), except for Word classes ($H = 5.64$, $p = 0.13$). Post-hoc analyses evidenced that language scores were significantly higher in preteens who use media for educative purposes. More in detail, post-hoc comparisons revealed that preteens who use media “to help in scholar tasks” or “to learn new things” obtained larger scores in formulated sentences and semantic relationships than those who use it “to communicate” (Formulated sentences: $p = 0.008$, $p = 0.024$, respectively; Semantic relationships: $p = 0.002$, $p = 0.020$, respectively). Similarly, students who use media “to help them with scholar tasks” obtained larger scores in recalling sentences than those who use it "to communicate" ($p = 0.05$). These outcomes were not modulated by participants' ages, as it did not differ between groups according to the use given to media ($H = 1.486$, $p = 0.686$).

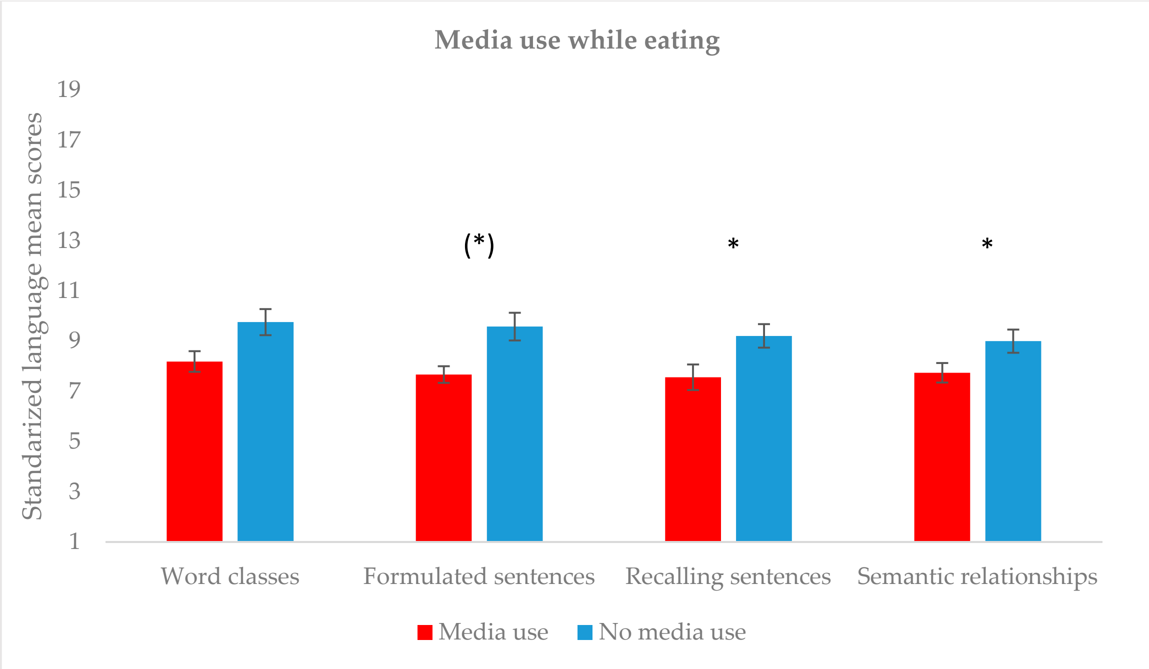


Figure 2. Mean scores on language tasks and use of technological devices while eating. * $p < 0.05$; (*) $p < 0.06$.

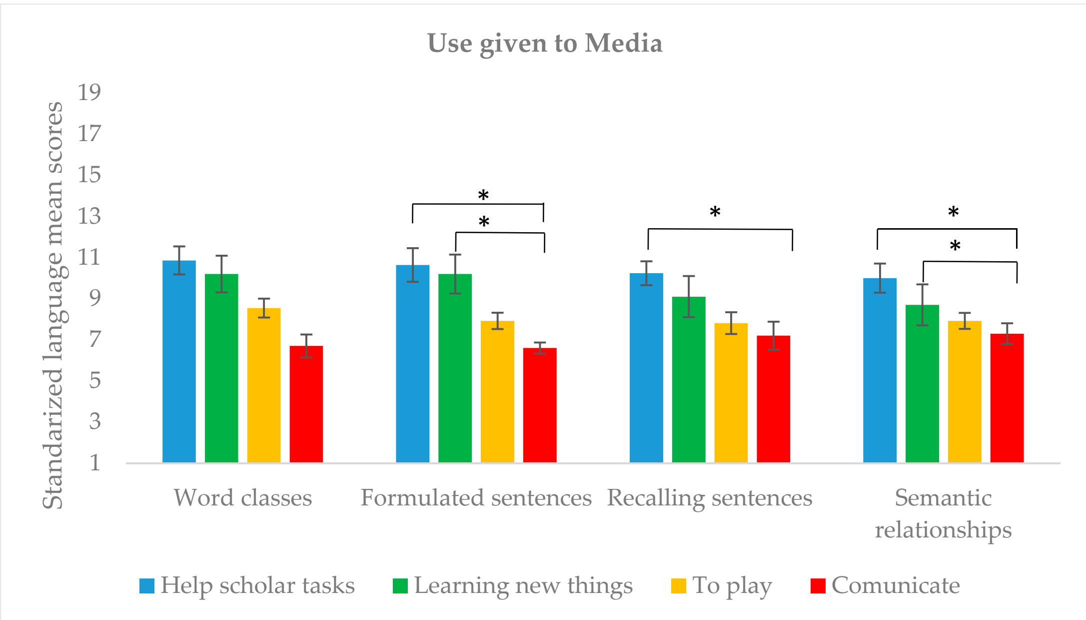


Figure 3. Mean language scores and main use of media. * $p < 0.01$.

To explore language scores between preteens who stated talking a lot with their parents in comparison to those who reported not talking much with them, independent comparisons were conducted for each language subtest (see Figure 4). More specifically, language scores were significantly higher in participants who talk a lot with their parents in all CELF-5 subtests, as compared to those who do not (Word classes: $U = 664.5$, $p < 0.001$, $r = 0.419$; Formulated sentences: $U = 840$, $p < 0.0001$, $r = 0.768$; Recalling sentences: $U = 746$, $p < 0.0001$, $r = 0.577$; and Semantic relationships: $U = 808.5$,

$p < 0.0001$, $r = 0.699$). These results were not affected by age, as participants in both groups did not differ in this regard ($U = 396$, $p = 0.356$).

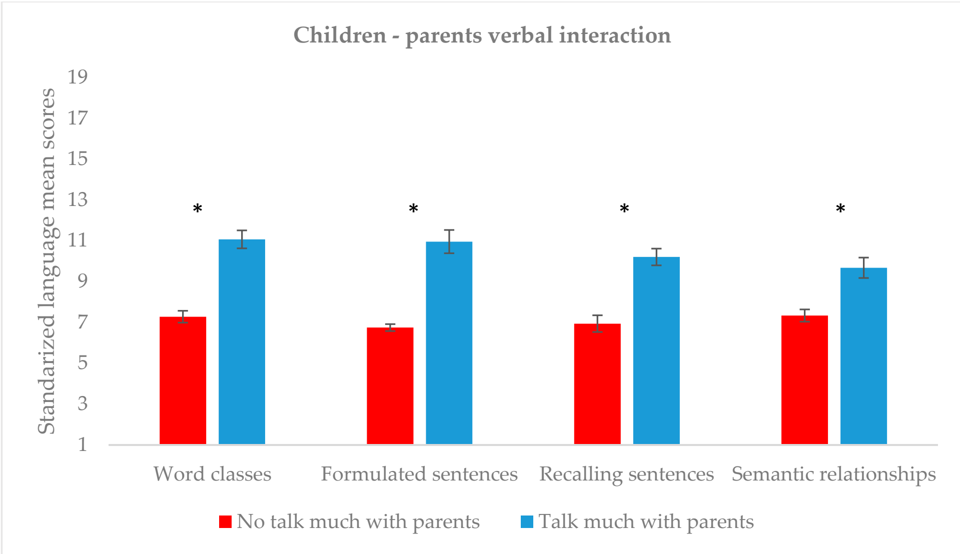


Figure 4. Mean language score and amount of verbal interaction with parents. * $p < 0.001$.

Finally, the frequency distribution analysis between the use of media in minutes and the tendency to talk with the parents (see Figure 5) showed that participants who considered that they did not talk much with their parents used technological devices in higher rates of time (χ^2 (2,60) = 19.495, $p < 0.0001$). Post-hoc tests using standardized residuals revealed that among preteens with the lowest amount of daily use of media, there were significantly more participants who talk frequently with their parents ($Z_{adj} = 3.96$, $p < 0.001$). In addition, participants who spent the highest rates of time with media every day report talking with their parents less frequently ($Z_{adj} = 3.87$, $p < 0.001$). There were no differences in terms of the frequency of self-reported interaction with their parents among preteens who spend intermediate time levels with media per day (one hour per day) ($Z_{adj} = 0.6$, $p = 0.835$).

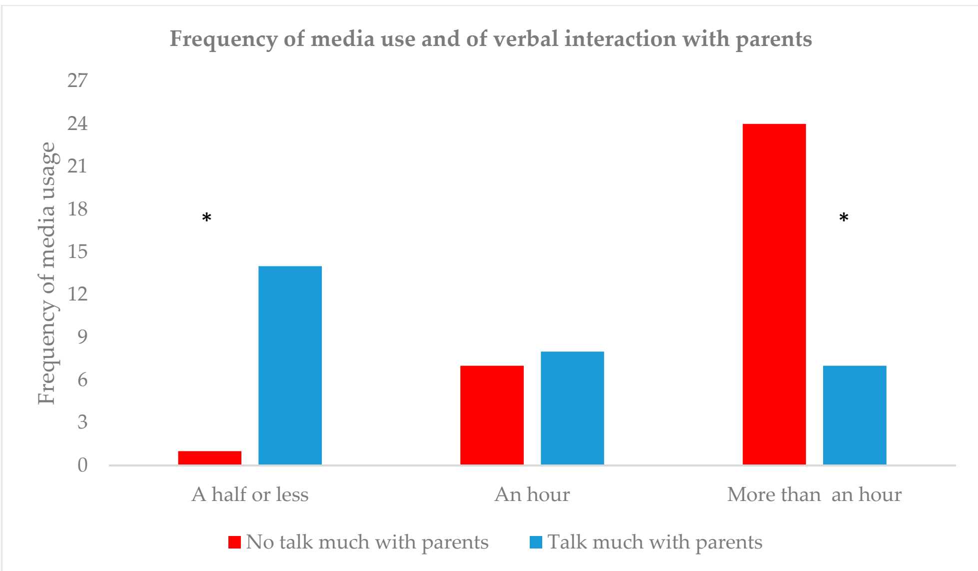


Figure 5. Frequency of media use in hours in relation to the amount of interaction with parents. * $p < 0.0001$.

4. Discussion

This study aimed to evaluate the associations between media use and language skills in a sample of preteens aged 11 and 12. As a secondary objective, this study assessed the potential impact of the quantity of parental language input on preteens' actual language skills.

Regarding technological use, most participants in our sample make daily use of technological devices. Participants stated that they used between 2 and 3 devices daily, on average. In particular, all of them had a TV at home, which was the most used device, and the second most used was the smartphone. Considering that almost 90% of the participants have a smartphone, we can consider that this sample has media as a daily element of their lives. Regarding this massive use of technological devices, recent studies have shown that the smartphone is the most used device among the youth between 10 and 14 years of age [4], which is a result that is not completely mirrored by the results of the present study, given that TV was the most frequently used one. Participants in the present study stated that they use technological devices for more than an hour and, mostly, to play. Regarding whether or not they use the devices while eating at these ages, approximately half of the participants in our sample did use them.

In terms of language skills, scores on the language subtests showed that values are slightly inferior to those expected when standardizing, which should commonly yield a mean score of 10 and a standard deviation of 3 points. We speculate that this might be because we only included bilingual participants with middle or low socioeconomic status.

We also report a negative association between the number of devices, their time of usage, and the CELF-5 scores. Considering the time of usage, results in all language subtests were lower for participants who made daily use of technological devices, as compared to those who made a more restricted use, mostly on weekends. These results are in consonance, but nuanced, with recent studies showing that children who use a moderate amount of media show the largest language gains, whereas both the lowest and the highest levels of media use are associated with smaller language gains [30]. In addition, other studies have demonstrated that sustained exposure to screens at an early age could even increase the risk of language delay [36]. Nevertheless, it has also been shown that when parents or caregivers are involved in the use of technological devices (frequent media joint engagement), the negative relationship between media exposure and language development has not been found [10].

Another important aspect of this study concerns the self-reported use of technological devices during meals. Recent studies have shown that children who regularly use technological devices during mealtime show a delay in language development, as long as this media use may displace language-enhancing activities, such as the interaction with their parents and siblings [30]. In our study, and considering that all participants are preteens, this interaction still shows an effect, as participants who eat while using technological devices show lower results in language subtest than those who do not combine these two activities. In this vein, several experts recommend limiting the use of technologies due to its negative effect on child development, not only in terms of language and cognition but also on their daily habits and routines, such as eating or sleeping [19,36].

In terms of the different uses given to media, results showed that those participants who stated to use media mainly to assist with their homework or to learn new things had higher scores on most of the language subtests as compared to the participants who stated to use media mainly to communicate with other people. In this vein, a meta-analysis examining the effects of children's exposure to international co-productions of Sesame Street, a program with clear learning content, showed significant positive effects of exposure to the program in cognitive outcomes [37]. Other studies have also shown that media used with educational purposes can improve language development, especially in economically disadvantaged children [11]. Therefore, the use given to media can influence language skills, which might mitigate their negative effect or even boost language development.

Another factor seems to mediate the relationship between media and language development. In this study, participants were also asked about their perceptions of the amount of interaction with their parents. Approximately half of the sample did not perceive to have much linguistic interaction with their parents or caregivers, and results showed that their language scores were lower as compared to those who reported having frequent interaction with them. More in detail, participants who reported having less communication with their parents spent more time using electronic devices. Thus, the interaction between parental language input and children's language level arises as a fundamental factor for the adequate development of language. Previous studies support the view that both the interaction with and the exposure to media are decisive factors for language development and suggest that they are directly related [10]. There is also evidence showing that when the interaction between parents and children is conditioned by the simple presence of a technological device in the background, the quantity and quality of the language used seems to be diminished [24].

We suggest that the preadolescent population might be making an abusive use of media and, adding the use given to it together with a low interaction with their parents or caregivers, can lead to an impoverishment in their language skills. Hence, we can conclude that the linguistic skills of preteens who have been considerably exposed to technological devices exhibit deficiencies and are below the average level. Nevertheless, we cannot claim that this relationship is causal, since other studies have described a larger motivation to use media for communicative purposes in children with language difficulties. In this regard, adolescents with language difficulties use media to establish social relationships and communicate, given that these communication formats (instant messaging and e-mails) are more impersonal and tolerant with linguistic errors [38]. However, this same study did not find differences in the frequency of media usage between children with language problems and their normative peers. Instead, other studies suggest that parents of children with language difficulties tend to talk less with them [39] and produce a language of worse quality [40]. Therefore, it seems that the relation between the use of media, the interaction with parents, and adolescent's language level is rather complex. A bidirectional cyclic relation might work as a more suitable explanation, in which the different factors influence each other. In this sense, we speculate that the use of media would diminish the time of parent–adolescent interaction and would worsen its quality. In a complementary way, having language difficulties might generate a weaker tendency by others to carry out appropriate communicative behavior. In turn, this behavior would lead to increased use of media by adolescents with language difficulties in an attempt to palliate this effect.

In general terms, the present results confirm the generalized spread of the use of media and, thus, the need to study its effect on children and adolescents, especially in those with prolonged exposures. Nevertheless, to the extent that parents or caregivers use technological devices with their children to learn, the negative effects of media on language development can be reduced, as the joint participation in learning activities with these devices can act a mitigating factor for the potential risks of exposure to media [30,41].

However, the results of this study should be taken with caution due to several limitations. The first limitation is the small sample size. Second, the Questionnaire on the Usage of Media in the Family Setting has some questions that offer little variability in their response items. In particular, for questions regarding the time spent daily using media—with the higher score (more than one hour) including around 50% of participants—and the question regarding frequency of talking with parents—with a dichotomous answer. Third, only parent's language input has been analyzed in this study not considering other agents such as siblings or peers. Finally, data collection was merely based on self-perceptions on the use given to media, as well as the linguistic input that preteens consider receiving from their parents. Consequently, the data used in this study are based on the associations between self-reports and objective linguistic outcomes.

Hence, future investigations might benefit from evaluating the quality and quantity of language input from parents, siblings, and peers through an observational approach, being advisable to include objective measures on the time spent using media devices to explore its potential associations with the language development of preteens in a larger sample. A more detailed differentiation regarding the duration of use of media devices and the frequency of talking with parents and other agents is needed in future research.

5. Conclusions

To summarize, the data obtained in this study point to a generalized use of media devices among preteens of 11 and 12 years of age, which is related to lower language scores and a perception of less frequent communication with their parents. Therefore, it can be considered that media has an influence on language skills depending on both the time spent using it and the type of use given to it. However, it must be kept in mind that media use and language level do not have a direct causal relation but rather a cyclic relation in which the involved factors mutually affect each other.

Considering our results, we recommend limiting the use of media devices to less than an hour per day and with an academic purpose, and to use them for communicative and play only on weekends.

Author Contributions: Conceptualization, L.B.-L., D.A.-R. and E.A.-M.; Data curation, L.A.; Formal analysis, D.A.-R. and E.A.-M.; Funding acquisition, D.A.-R. and E.A.-M.; Project administration, L.A.; Supervision, E.A.-M.; Writing—original draft, L.A.; Writing—review and editing, L.B.-L., D.A.-R. and E.A.-M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by FEDER/Ministerio de Ciencia, Innovación y Universidades_Agencia Estatal de Investigación / EDU2017-85909-P.

Acknowledgments: We thank all school centers and all the participants for their willingness to participate in the present study.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. English Translate Version of the Questionnaire on the Usage of Media in the Family Setting

1. Which technological devices do you use every day (you can choose more than one option)?
    a. Tablet
    b. Computer
    c. Television
    d. Mobile phone
    e. Nintendo Switch
    f. PlayStation
    g. Others: ________________________
2. How many hours do you think you approximately spend using these devices daily?
    a. Less than half an hour
    b. Half an hour
    c. One hour
    d. More than one hour
3. You use media devices . . .
    a. Everyday
    b. Weekends or festivities only
4. Do you usually use these devices while eating?
    a. Yes, the television
    b. Yes, other technological devices
    c. No, I don't

5. Do you consider that you talk a lot with your parents?

   a. Yes, I do
   b. No, I don't

6. Which is the most frequent usage you give to these devices? (Choose only one option)

   a. Assistance in school tasks
   b. To learn new things
   c. To play
   d. To communicate with friends or relatives

Thanks for your participation!

References

1. Prensky, M. Digital natives, digital immigrants. *Horizon* **2001**, *9*, 1–6.
2. Valdivia Álvarez, I.; Gárate Sánchez, E.; Regal Cabrera, N.; Castillo Izquierdo, G.; Sáez, Z.M. Exposición a televisión y retardo primario del lenguaje en menores de 5 años [Television exposure and primary language delay in children under 5 years of age]. *Rev. Cuba. Pediatr.* **2014**, *86*, 18–25.
3. CinfaSalud Siete de Cada Diez Niños y Niñas Españoles Comen con una Pantalla Delante [Seven Out of Ten Spanish Boys and Girls Eat with a Screen in Front]. Available online: http://bit.ly/2SQ5E9q (accessed on 18 February 2020).
4. Masip, M.; Balagué, I. *El uso del Teléfono Móvil en España [The Movile Phone Use in Spain]*; Instituto Psicológico Desconect@: Barcelona, Spain, 2015.
5. National Institute of Statistics Summary of Data on Children 10 to 15 Years Old by Sex, Age, Home Size, Habitat, Monthly Income and Main Variables. Available online: http://bit.ly/2vGQD1M (accessed on 18 February 2020).
6. National Institute of Statistics Survey on Equipment and Use of Information and Communication Technologies in Homes. Available online: http://bit.ly/2V1KojJ (accessed on 18 February 2020).
7. Ministerio de Educación, Cultura y Deporte. Royal Ordinance 126/2014, 28 February, Where Basic Primary Education Curricula are Established; BOE, 2014. Available online: https://www.boe.es/buscar/pdf/2014/BOE-A-2014-2222-consolidado.pdf (accessed on 18 February 2020).
8. Hillemeier, M.M.; Morgan, P.L.; Farkas, G.; Maczuga, S.A. Perinatal and socioeconomic risk factors for variable and persistent cognitive delay at 24 and 48 months of age in a national sample. *Matern. Child Health J.* **2011**, *15*, 1001–1010. [CrossRef] [PubMed]
9. Rico Martín, A.M.; Níkleva, D. Análisis de la competencia lingüístico-discursiva escrita de los alumnos de nuevo ingreso del Grado de Maestro en Educación Primaria [Analysis of the written linguistic-discursive competence of new students of the Degree in Primary Education]. *Rev. Signos* **2016**, *49*, 48–70. [CrossRef]
10. Mendelsohn, A.L.; Brockmeyer, C.A.; Dreyer, B.P.; Fierman, A.H.; Berkule-Silberman, S.B.; Tomopoulos, S. Do verbal interactions with infants during electronic media exposure mitigate adverse impacts on their language development as toddlers? *Infant Child Dev.* **2010**, *19*, 577–593. [CrossRef] [PubMed]
11. Al-Harbi, S. The Influence of Media in Children's Language Development. *J. Educ. Dev. Psychol.* **2015**, *5*, 1. [CrossRef]
12. Zimmerman, F.J.; Christakis, D.A.; Meltzoff, A.N. Associations between media viewing and language development in children under age 2 years. *J. Pediatr.* **2007**, *151*, 364–368. [CrossRef]
13. Taylor, G.; Monaghan, P.; Westermann, G. Investigating the association between children's screen media exposure and vocabulary size in the UK. *J. Child. Media* **2018**, *12*, 51–65. [CrossRef]
14. Huber, B.; Yeates, M.; Meyer, D.; Fleckhammer, L.; Kaufman, J. The effects of screen media content on young children's executive functioning. *J. Exp. Child Psychol.* **2018**, *170*, 72–85. [CrossRef]
15. Malo-Cerrato, S.; Martín-Perpiñá, M.-M.; Viñas-Poch, F. Excessive use of social networks: Psychosocial profile of Spanish adolescents. *Comunicar* **2018**, *26*, 101–110. [CrossRef]
16. Christakis, D.A. The effects of infant media usage: What do we know and what should we learn? *Acta Paediatr.* **2009**, *98*, 8–16. [CrossRef] [PubMed]

17. Vandewater, E.A. Time well spent? Relating television use to children's free-time activities. *Pediatrics* **2006**, *117*, e181–e191. [CrossRef] [PubMed]
18. Chonchaiya, W.; Pruksananonda, C. Television viewing associates with delayed language development. *Acta Paediatr.* **2008**, *97*, 977–982. [CrossRef]
19. American Academy of Pediatrics. Media Use by Children Younger Than 2 Years. *Pediatrics* **2011**, *128*, 1040–1045. [CrossRef] [PubMed]
20. Fenson, L.; Dale, P.S.; Reznick, J.S.; Donna, T.; Elisabeth, B.; Hartung, J.P.; Pethick, S.J.; Reilly, J.; Thal, D.J.; Bates, E.; et al. *The MacArthur Communicative Development Inventories: User's Guide and Technical Manual*; Singular Publishing Group: San Diego, CA, USA, 1993.
21. Jackson-Maldonado, D.; Thal, D.J.; Marchman, V.; Newton, T.; Fenson, L.; Conboy, B. *McArthur Inventarios del Desarrollo de Habilidades Comunicativas. User's Guide and Technical Manual*; Brookes: Baltimore, MD, USA, 2003.
22. Tomopoulos, S.; Dreyer, B.P.; Berkule, S.; Fierman, A.H.; Brockmeyer, C.A.; Mendelsohn, A.L. Infant media exposure and toddler development. *Arch. Pediatr. Adolesc. Med.* **2010**, *164*, 1105–1111. [CrossRef] [PubMed]
23. Kirkorian, H.L.; Pempek, T.A.; Murphy, L.A.; Schmidt, M.E.; Anderson, D.R. The impact of background television on parent-child interaction. *Child Dev.* **2009**, *80*, 1350–1359. [CrossRef] [PubMed]
24. Pempek, T.A.; Kirkorian, H.L.; Anderson, D.R. The effects of background television on the quantity and quality of child-directed speech by parents. *J. Child. Media* **2014**, *8*, 211–222. [CrossRef]
25. Serra, M.; Serrat, E.; Solé, R.; Bel, A.; Aparici, M. *La Adquisición del Lenguaje [Language Acquisition]*; Ariel: Barcelona, Spain, 2000.
26. Rivero García, M. Influencia del habla materna en los inicios de la adquisición del lenguaje: Primeras palabras y primeros enunciados de más de una palabra [Influence of mother's speech in the beginning of language acquisition: First words and first sentences of more than one word]. *Rev. Logop. Foniatr. Audiol.* **1994**, *14*, 148–155. [CrossRef]
27. Aguaded Gómez, M.C. La participación de la familia en el desarrollo del lenguaje y en el aprendizaje escolar [Family involvement in language development and school learning]. *Rev. Educ. XXI* **2000**, *2*, 313–318.
28. Gilkerson, J.; Richards, J.A.; Warren, S.F.; Montgomery, J.K.; Greenwood, C.R.; Kimbrough Oller, D.; Hansen, J.H.L.; Paul, T.D. Mapping the Early Language Environment Using All-Day Recordings and Automated Analysis. *Am. J. Speech Lang. Pathol.* **2017**, *26*, 248–265. [CrossRef]
29. Hirsh-Pasek, K.; Adamson, L.B.; Bakeman, R.; Owen, M.T.; Golinkoff, R.M.; Pace, A.; Yust, P.K.S.; Suma, K. The Contribution of Early Communication Quality to Low-Income Children's Language Success. *Psychol. Sci.* **2015**, *26*, 1071–1083. [CrossRef] [PubMed]
30. Dore, R.A.; Logan, J.; Lin, T.-J.; Purtell, K.M.; Justice, L. Characteristics of children's media use and gains in language and literacy skills. *Front. Psychol.* **2020**, *11*. [CrossRef] [PubMed]
31. Einarsdóttir, J.T.; Björnsdóttir, A.; Símonardóttir, I. The predictive value of preschool language assessments on academic achievement: A 10-year longitudinal study of Icelandic children. *Am. J. Speech Lang. Pathol.* **2016**, *25*, 67–79. [CrossRef] [PubMed]
32. Murillo, F.J.; Hernández-Castilla, R. ¿La implicación de las familias influye en el rendimiento? Un estudio en educación primaria en América Latina [Does family involvement influence performance? A study in primary education in Latin America]. *Rev. Psicodidáctica* **2020**, *25*, 13–22. [CrossRef]
33. Valera-Pozo, M.; Adrover-Roig, D.; Pérez-Castelló, J.A.; Sanchez-Azanza, V.A.; Aguilar-Mediavilla, E. Behavioral, emotional and school adjustment in adolescents with and without developmental language disorder (DLD) is related to family involvement. *Int. J. Environ. Res. Public Health* **2020**, *17*, 1949. [CrossRef]
34. Asociación para la Investigación de Medios de Comunicación Nuevo Sistema de Clasificación Socioeconómica en el EGM [New Socioeconomic Classification System in EGM]. Available online: https://www.aimc.es/otros-estudios-trabajos/clasificacion-socioeconomica/ (accessed on 18 February 2020).
35. Wiig, E.H.; Semel, E.; Secord, W.A. *Evaluación Clínica de los Fundamentos del Lenguaje CELF-5*; Pearson: Madrid, Spain, 2019.
36. Collet, M.; Gagnière, B.; Rousseau, C.; Chapron, A.; Fiquet, L.; Certain, C. Case–control study found that primary language disorders were associated with screen exposure. *Acta Paediatr.* **2019**, *108*, 1103–1109. [CrossRef]
37. Mares, M.L.; Pan, Z. Effects of Sesame Street: A meta-analysis of children's learning in 15 countries. *J. Appl. Dev. Psychol.* **2013**, *34*, 140–151. [CrossRef]

38. Durkin, K.; Conti-Ramsden, G.; Walker, A.J. Computer-mediated communication in adolescents with and without a history of specific language impairment (SLI). *Comput. Human Behav.* **2010**, *26*, 176–185. [CrossRef]
39. Hammer, S.; Tomblin, J.B.; Zhang, X.; Weiss, A.L. Relationship between parenting behaviours and specific language impairment in children. *Int. J. Lang. Commun. Disord.* **2001**, *36*, 185–205. [CrossRef]
40. Ruser, T.F.; Arin, D.; Dowd, M.; Putnam, S.; Winklosky, B.; Rosen-Sheidley, B.; Piven, J.; Tomblin, B.; Tager-Flusberg, H.; Folstein, S. Communicative competence in parents of children with autism and parents of children with specific language impairment. *J. Autism Dev. Disord.* **2007**, *37*, 1323–1336. [CrossRef]
41. Zimmerman, F.J.; Gilkerson, J.; Richards, J.A.; Christakis, D.A.; Xu, D.; Gray, S.; Yapanel, U. Teaching by listening: The importance of adult-child conversations to language development. *Pediatrics* **2009**, *124*, 342–349. [CrossRef] [PubMed]

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

children

Article

Identifying Bilingual Children at Risk for Language Impairment: The Implication of Children's Response Speed in Narrative Contexts

Pui Fong Kan *, Anna Miller and Sierra Still

Department of Speech, Language, and Hearing Sciences, University of Colorado Boulder, Boulder, CO 80309, USA; anmi8839@colorado.edu (A.M.); Sierra.Still@colorado.edu (S.S.)
* Correspondence: puifong.kan@colorado.edu

Citation: Kan, P.F.; Miller, A.; Still, S. Identifying Bilingual Children at Risk for Language Impairment: The Implication of Children's Response Speed in Narrative Contexts. *Children* **2021**, *8*, 62. https://doi.org/10.3390/children8020062

Academic Editor: Brad Farrant
Received: 7 December 2020
Accepted: 18 January 2021
Published: 20 January 2021

Abstract: The purpose of the study was to examine whether monolingual adults can identify the bilingual children with LI on the basis of children's response speed to the examiner. Participants were 37 monolingual English-speaking young adults. Stimuli were 48 audio clips from six sequential bilingual children (48 months) who were predominately exposed to Cantonese (L1) at home from birth and started to learn English (L2) in preschool settings. The audio clips for each child were selected from an interactive story-retell task in both Cantonese and English. Three of the children were typically developing, and three were identified as having a language impairment. The monolingual adult participants were asked to judge children's response times for each clip. Interrater reliability was high (Kalpha = 0.82 for L1; Kalpha = 0.75 for L2). Logistic regression and receiver operating characteristic curves were used to examine the diagnostic accuracy of the task. Results showed that monolingual participants were able to identify bilingual children with LI based on children's response speed. Sensitivity and specificity were higher in Cantonese conditions compared to English conditions. The results added to the literature that children's response speed can potentially be used, along with other measures, to identify bilingual children who are at risk for language impairment.

Keywords: processing speed; bilingual; impairment; screening

1. Introduction

Developmental language disorder (DLD) [1] affects approximately 7–11% of children [2,3]. Children with DLD exhibit significant language deficits that cannot be attributed to sensory, motor, neurological, or socio-emotional impairments [1,3,4]. Clinically, one pressing problem is that young children who learn a minority language (L1) at home from birth and start to learn a community language (L2) in school settings are at particular risk for misdiagnosis with DLD [5–9]. There are three factors related to the over-identification and under-identification: (1) the fluctuation of typically developing (TD) bilingual children's language skills as a function of L1 and L2 input and use [10–12], (2) the lack of valid language assessment tools and norms for bilingual children [13,14], and (3) the shortage of bilingual or bicultural speech-language pathologists who are trained to assess bilingual children [15,16].

In the past decades, many alternative assessment approaches (e.g., processing tasks) have been proposed to screen bilingual children who might be at risk for language impairment [17–22]. One approach that involves examining bilingual children's processing speed has gained traction as the indicator for language impairment [20–24]. According to the processing-based accounts, the slow processing speed in children with DLD might be related to their limited ability to process linguistic information [25,26] and attention deficits [21,27]. Convergent evidence indicates that monolingual children with DLD demonstrate slower processing speed than TD monolingual children on linguistic and nonlinguistic tasks [23,25,28–30]. For example, children with DLD are slower to name pictures [31], judge whether a sentence is grammatically correct [24,32], recall visuospatial information [28], distinguish nonlinguistic tones [21,22,26], and rotate shapes mentally [23,30]. Park and colleagues [33] examined whether linguistic and nonlinguistic processing speed measures can be used as clinical markers for monolingual children with DLD. The binary logistic regression results showed that a combination of linguistic and nonlinguistic processing speed tasks moderately predict monolingual children's DLD status. However, slow processing speed appears to be more predictive of the presence of DLD, but not the absence of DLD.

Bilingual children's language experience is an important factor in the investigation of processing speed. Some studies found that bilingual experience could enhance TD bilingual children's executive function, resulting in faster processing speed in certain nonlinguistic processing tasks that involve conflict resolution (e.g., card sort task) [34,35]. However, some studies do not find such an advantage in other processing tasks (e.g., visually detecting colors) [22,24]. These findings suggest that the variability in bilingual children's response time could be associated with their bilingual experience, and the type of tasks could affect the diagnostic accuracy of processing speed. Ebert and Pham (2019) compared the processing speed of Spanish-English school-aged bilingual children with DLD (n = 92; 6; 0–10; 11) and aged-matched TD bilingual children (n = 109) using a nonlinguistic task, called visual detection. The task required the child to press a button that corresponded to the red/blue circle on the computer screen. They found that bilingual children with DLD were slower than their TD peers across all age groups. However, the sensitivity (ranging from 0.41 to 1) and specificity (ranging from 0.27 to 0.9) varied across age groups. Another consideration is the implementation of processing speed tasks for 3- to 5-year-old preschool children. Many processing tasks used in previous studies are designed for school-aged children and require the press of a button or strike of a key on a computer [21–23,27]. These tasks, which require attention control, motor, and perceptual skills to encode auditory and/or visual stimulus, might be difficult for young preschool-aged children [22,23,36]. The implementation of such processing speed tasks for young bilingual preschool children could lead to larger variability, which could negatively affect its diagnostic accuracy.

In this study, we present an alternative method for identifying young bilingual preschool children who are slower than their peers. Specifically, we examined the feasibility of using a judgment task by adults to identify the slow bilingual children who might be at risk for language impairment. Parental and teachers' concerns, or ratings have been an important early indicator of developmental issues in clinical settings [37–42]. Many prescreening and screening tools for bilingual children involve parents or teachers rating the amount of L1 and L2 input [10,38] and/or rating bilingual children's language skills [43]. To our best knowledge, no previous studies have examined whether the rating of children's response speed could be used as a tool for identifying at-risk bilingual preschool children who are at risk for language impairment in the screening process.

Previous research has utilized auditory-perceptual judgment tasks to examine speech characteristics such as respiration, voice quality, intelligibility, and fluency in individuals with speech disorders [44–50]. In this study, we explored whether judging children's response speed to adults' prompts in narrative contexts could be used to identify slow bilingual children who might be at risk for language impairment. Methodologically, two aspects should be noted in this investigation. First, the stimuli were extracted from the audio clips of three typically developing bilingual children who speak Cantonese as L1 and English as L2 and three Cantonese-English bilingual children who have been clinically identified as having language impairment. The interactive story-retelling task was implemented in two sessions for each child: one in Cantonese (children's L1) and one in English (L2). Second, the response-speed judgment task is done by monolingual English-speaking adults. The use of monolingual adults was motivated by the shortage of bilingual clinicians in the U.S. The primary objective of this study was to examine whether monolingual adult speakers identify bilingual children with LI on the basis of their response speed to the examiner. The results would contribute to our understanding of the identification of at-risk bilingual children by monolingual clinicians. We specifically asked the following questions:

1. What is the interrater reliability of the response speed ratings?
2. What are the classification accuracies of the response speed ratings at the audio clip level? Are there any differences between L1 and L2 audio clips?
3. How well do the response speed ratings differentiate bilingual children with LI from TD bilingual children?

2. Materials and Methods

This project has been approved by the Institutional Review Board of the University of Colorado Boulder on 9 November 2018 (Protocol #: 18-0277).

2.1. Participants

Participants were 37 monolingual English-speaking adults (28 females and 9 males) between 18 and 41 years old (Mean age = 23.35; SD = 5). They were recruited from the Department of Speech, Language, and Hearing Sciences (SLHS) at the University of Colorado, U.S. To qualify for this study, the individuals must meet the following criteria: (1) primarily use English in his/her daily lives, (2) have no knowledge of Cantonese; (3) must have completed at least two courses in SLHS. The participants reported that they had between 15 and 20 years of formal education, from first grade to their current educational year (Mean = 16.08 years; SD = 1.38). Of the 37 participants, 27 were undergraduate students, 8 were in the post-baccalaureate or master's program, and 2 were in the doctoral program. Most of the participants (n = 27) were White; 3 were African American; 2 were Asian American, and 5 were mixed race. None of the participants had exposure to Cantonese. None of them reported that they had language, hearing, or vision problems.

2.2. Response-Speed Judgement Task

Stimuli of the Response-Speed Judgement Task were 48 short audio-clips (Mean = 23.7 seconds; SD = 9.83) of the adult-child interactions of 6 children (4 clips per child $\times$ 2 languages) during a story-retell task (see Table 1). These samples were selected from 248 audio recordings of a larger study led by the first author. All children had audio recordings in L1. The third author used *Praat* [51] to identify the examiner–child interactions in the audio recordings of the story-retell tasks in L1 and L2. The selection criteria of the clips included at least three continuous exchanges between the examiner and the child. Children who only had recordings in L1 or L2 were excluded. The children with LI and TD children were age-matched. Because of the variation of the adult–child interactions, the clips varied in times. The identity of the selected clips was blind to the rest of the research team; only the third author, who was not involved in data collection, had the key to all the selected clips. The 48 selected audio clips, including those from TD and LI groups, were randomly combined into two large audio files: one in Cantonese (24 clips) and one in English (24 clips). There was a five-second interval of silence between each clip.

The six children were exposed to Cantonese (L1) at home from birth and started to learn English (L2) when they started preschool. Three of the six children were typically developing (TD) children (2 females, 1 male). The other three children (two females, one male) were clinically identified as having language impairment (LI) and had an individual educational program (IEP). The clinical diagnoses were based on clinicians' interpretation of children's language performance on criterion-reference tasks, parents' concerns, and teachers' reports, and clinical observations. All children had a standard score of 80 or above on the brief IQ screening of the Leiter-R [52]. There were no significant differences between the two groups $F(1, 4) = 1{,}29$, $p > 0.05$. Since there are no valid measures and norms to make DLD diagnosis for Cantonese-English bilingual children [11], we use a broad term, language impairment (LI), to describe the children who received language intervention in this study. The story-retell task involved each child retelling a story, *Frog, Where Are You?* [53] after the examiner told him/her the story. The story retell task was administered in both Cantonese and English. The prompts by the examiner were open-ended and minimal, including phrases such as "tell me more", "uh-huh", "and then what happens?" to encourage the child to continue the story. The order of the language tested first was counterbalanced. Table 1 summarizes the information of the audio clips. To reduce biases, the measurements were done after the monolingual participants completed the response-speed judgment tasks.

Table 1. Audio-clips information: LI and typically developing (TD) children by language.

	LI (12 Clips)		**TD (12 Clips)**	
	Cantonese	English	Cantonese	English
Clip length (in seconds)	19.92 (1.26)	21.42 (7.51)	19.83 (5.54)	26.50 (11.07)
Turns	6.00 (4.16)	6.75 (4.41)	7.08 (3)	8.00 (2.95)
Examiner syllables per second	4.77 (0.73)	4.07 (1.26)	5.02 (0.43)	5.17 (1.12)
Examiner response-to-child interval (in seconds)	0.30 (0.28)	0.30 (0.26)	0.24 (0.14)	0.36 (0.22)
Child syllables per second	2.04 (2.24)	2.05 (1.05)	4.85 (1.57)	2.14 (.86)
Child response-to-adult interval	3.24 (2.55)	2.25 (1.41)	0.67 (0.05)	1.2 (0.41)

Note. Child response-to-examiner interval = the interval between the end of the examiner's prompt and the onset of the child's first syllable; Examiner response-to-child interval = the interval between the end of the child's utterance and the onset of the examiner's first syllable.

Repeated measures analysis of variance (ANOVA) indicated that there was no significant effect of group (LI vs. TD clips) on clip length, $F(1, 44) = 1.37$, $p > 0.05$, or on turns $F(1, 44) = 1.2$, $p > 0.05$. There was no effect of language (Cantonese vs. English) on clip length, $F(1, 44) = 1.96$, $p > 0.05$, or on turns $F(1, 44) = 0.61$, $p > 0.05$. The findings suggest that the clips were comparable across the two groups and across languages. In terms of the responses by examiners, the repeated measures ANOVA results also showed that there was no significant group effect (LI vs. TD clips) on examiner syllables per second, $F(1, 44) = 3$, $p > 0.05$, or on examiner response to child interval $F(1, 44) = 0.003$, $p > 0.05$. There was no significant language effect on examiner syllables per second, $F(1, 44) = 2.28$, $p > 0.05$, or on examiner response-to-child interval $F(1, 44) = 0.87$, $p > 0.05$. The findings suggest that the prompts by the examiner and the amount of time the examiner took to respond to the child were comparable across the two groups and across languages. There was a significant group effect (LI vs. TD clips) on children's syllables per second, $F(1, 44) = 1.77$, $p < 0.05$, suggesting there were fewer syllables per second in the clips of children with LI than those of TD children. There was a significant language effect on children's syllables per second, $F(1, 44) = 9.38$, $p < 0.05$, suggesting children had more syllables per second in Cantonese than in English. The results are consistent with the teachers' report that children had stronger Cantonese skills (L1) than English at the time of testing. Repeated measures ANOVA results showed that there was a significant group effect on response-to-examiner interval, $F(1, 44) = 37.85$, $p < 0.001$, suggesting children from the LI group took longer to respond to the examiner than their TD peers. There was no language effect, $F(1, 44) = 0.44$, $p > 0.05$ or group $\times$ language interaction on response-to-examiner intervals, $F(1, 44) = 1.37$, $p > 0.05$.

To illustrate the variability of individual children's response-to-examiner intervals (in seconds), we summarize the means and standard deviations of the response-to-examiner intervals of each child in Table 2. There are four clips for each child for each language; and there were six to eight turns within each clip.

Table 2. Mean and Standard Deviation of child response-to-examiner interval (in seconds).

	Group	Cantonese (L1)	English (L2)
Child 2	LI	2.75 (0.75)	2.56 (1.05)
Child 3	LI	1.48 (0.44)	1.95 (0.65)
Child 5	LI	5.5 (0.45)	2.25 (0.26)
Child 1	TD	0.75 (0.71)	1.66 (0.47)
Child 4	TD	0.64 (0.38)	1.06 (0.77)
Child 6	TD	0.63 (0.53)	0.88 (0.14)

Note: TD = Typically-developing; LI = Language impairment.

2.3. Procedures

Each participant was tested separately in a quiet room in the laboratory. It took the participant between 25 and 35 min to complete the practice trials and rate the audio clips in the testing phase. Practice trials were administered before testing to ensure that the participants understood the testing procedure. The practice trials involved four audio clips, which were different from those used for the response-speed judgment task. Two of the practice clips contained interactions of a child who had slow response speed, while the other two practice clips were interactions of a child who had normal response speed. The examiner read the following script to each participant: "You will listen to a series of audio-clips, where you will hear an adult and a child's voice. Please rate the speed that you believe it takes the child to respond to the examiner. The scale ranges from 1 to 4. "1" is a very slow response time, "2" is a slow response, "3" is a slightly slow response, and "4" is a normal response time." After the presentation of each clip, the participants were instructed to mark the child's response speed on a 4-point scale. To advance the response-speed judgment task, the participants had to respond correctly to all four practice items and verbally indicate that they understood the procedures. All participants reached the criteria.

During the response-speed judgment task, the 48 target audio clips were presented to each adult participant. The participants were not told that some clips were from children with language impairment, and some were from TD children. The order of the Cantonese and English clips was counterbalanced. Nineteen participants were presented to the 24 Cantonese clips before the 24 English clips; 18 participants were presented to the 24 English clips prior to the 24 Cantonese clips. Before each clip was presented, a number was shown on the computer screen in front of the participant to confirm the clip number, which corresponded with the rating form. After the examination of interrater reliability (see Section 3.1), the ratings of 1 and 2 were coded as "slow speed" and 3 and 4 as "normal speed" for analysis.

3. Results

3.1. Interrater Agreement

Krippendorff's α was computed to examine the reliability across the 37 raters for items in each language. The Krippendorff's α was developed for more than two raters and various data types, including ordinal data [54]. In this analysis, each rater's ratings, ranging from 1 to 4, were examined. The 95% confidence intervals (CIs) were calculated by bootstrapping (n = 10,000). As shown in Table 3, Kalpha was 0.82 (95% CI = 0.81, 0.82) for the Cantonese items, suggesting high interrater agreement about the response speed of the children in the clips in Cantonese, whereas Kalpha was 0.75 (95% CI = 0.75, 0.76) for the English items, suggesting moderate interrater agreement about the response speed of the children in the clips in English.

Table 3. Interrater agreement: Krippendorff's α (kalpha) for ratings across 37 raters.

	Kalpha	95% CI	*p* (Kalpha < 0.60)
Cantonese (L1)	0.82	[0.81, 0.82]	<0.001
English (L2)	0.75	[0.75, 0.76]	<0.001

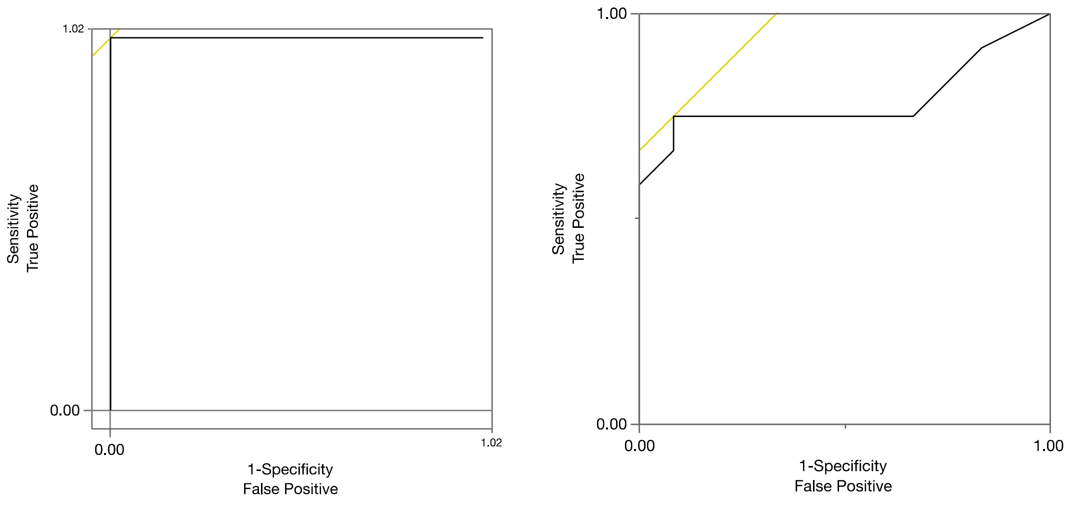
3.2. Slow Speed Ratings and Audio Clips from Children with LI

The ratings of 1 and 2 were coded as "slow speed" and 3 and 4 as "normal speed" for analysis. The distribution of slow response speed ratings is summarized in Table 4. Overall, 95% of the clips from the children with LI in the Cantonese condition were identified as slow speed, whereas 77% of the clips from the children with LI in the English condition were identified as slow speed.

Table 4. Distribution of slow response speed ratings by audio clip category (LI vs. TD) by language.

	LI	TD
Cantonese (L1)	420 (95%)	22 (5%)
English (L2)	332 (77%)	99 (23%)

Logistic regression analyses showed that the clips from children with LI in the Cantonese conditions were likely to be rated as "slow speed," χ^2 (1) = 33.27, $p < 0.001$; and children with LI in the English conditions were likely to be rated as "slow speed," χ^2 (1) = 9.31, $p < 0.01$. The ratings were assessed as a metric for determining the clip categories (i.e., LI or TD). Receiver operating characteristic (ROC) curves are plotted in Figure 1 (1a for the Cantonese and 1b English) and the areas under the ROC Curve (AUC) were calculated. For the Cantonese conditions (24 clips), the sensitivity was 1, and the specificity was 1, with AUC = 1 (see Table 5). The results indicated that the ratings were excellent at separating the audio clips of children with LI from those of TD children (see Figure 1a). For the English conditions (24 clips), the sensitivity was 0.7 and the specificity was 0.92, with AUC = 0.79 (see Table 5). The results indicated that the ratings were good at separating the clips of children with LI from those of TD children (see Figure 1b).



(**a**) (**b**)

Figure 1. Receiver operating characteristic (ROC) curve for slow rating for all audio clips, (**a**) Cantonese; (**b**) English.

Table 5. Ratings associated with the best combination of sensitivity and specificity.

	X	Sensitivity	Specificity	False Positive	False Negative
Cantonese (L1)	13.00	1.0	1.0	0	0
English (L2)	34.00	0.75	0.92	0.08	0.25

3.3. Identification of Children with LI Using Response Speed Ratings

Figure 2 displays the total number of audio clip ratings for each child (37 participants × 4 clips for each language). For the Cantonese conditions with four clips nested within each child, the ratings (slow vs. normal) predict LI, χ^2 (4) = 33.27, $p < 0.001$. The findings suggest that the ratings of the Cantonese samples are excellent in differentiating the children with LI from TD children. For the English conditions with four clips nested within each child, the ratings (slow vs. normal) predict LI, χ^2 (4) = 19.14, $p < 0.01$. The findings suggest that the ratings of the English samples are likely to differentiate the children with LI from TD children. However, there are some individual differences across children. As shown in Figure 2, the ratings for Child 3 (with LI) and Child 4 (TD) for the English condition, were at chance or almost at chance, respectively, although the ratings for these children's Cantonese clips identify Child 3 as LI and Child 4 as TD.

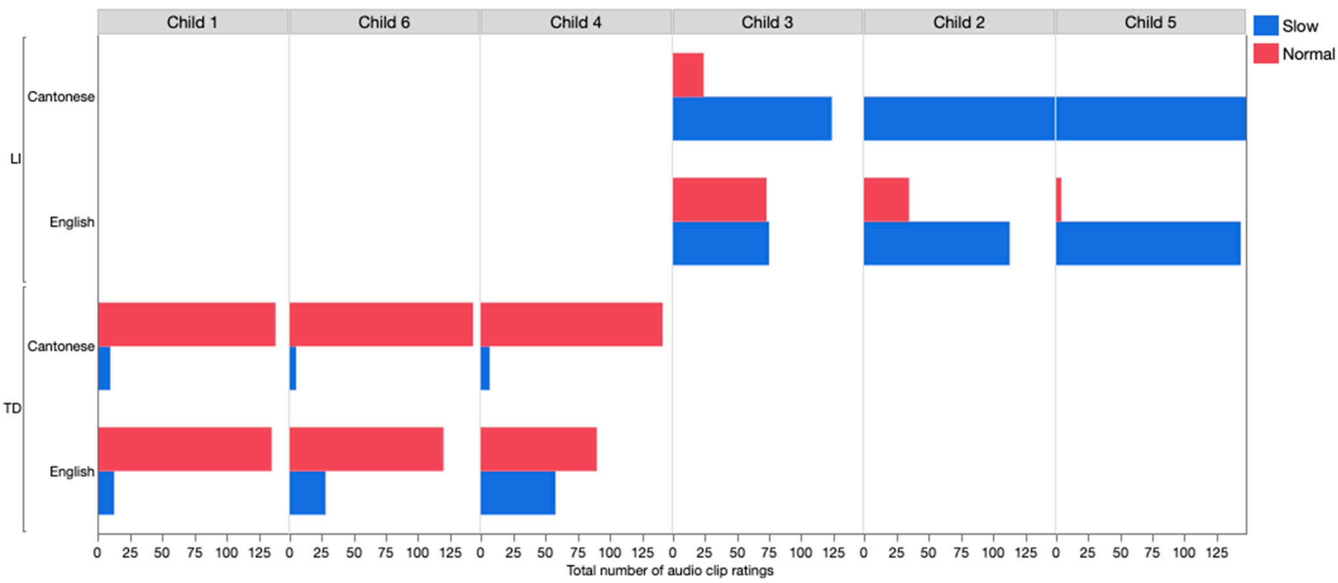


Figure 2. Total number of audio clip ratings (slow vs. normal) for each child. TD = Typically-developing and LI = Language impairment

4. Discussion

The present study was built from the literature about the slow processing speed in children with DLD [21,25,26,33]. The study was designed to examine young bilingual children who cannot complete the processing tasks. Rather than directly testing young children, this study examined whether monolingual adults' judgment of children's response speed in a narrative context could be used to identify bilingual children with LI (mean age = 48 months). Thirty-seven normal monolingual English-speaking adults (28 females and 9 males), who were enrolled in the SLHS Department, completed a response-speed judgment task. The task stimuli were 48 audio clips of 6 bilingual preschool children who began learning Cantonese (L1) from birth and English (L2) in a preschool setting. Three of these children had been identified as having language impairment, and three were typically developing bilingual preschoolers. The audio clips for each child were selected from an interactive sample between an examiner and a child, where he/she was asked to retell a story called, *Frog, Where Are You?*. Several key findings emerged from the analyses of this study. First, reliability is particularly notable across the 37 raters for both Cantonese and English audio clips. Second, both sensitivity and specificity were high for audio clips (Figure 1). Importantly, although the raters do not know Cantonese, they were able to identify the audio clips of the children who were slow in responding to the examiners. Their ratings were more consistent when rating clips in Cantonese than in English. Third, variability was noted across raters in the English clips nested within individual children. The English samples of two children, in particular, were rated at chance or near chance level.

One important finding in this investigation is the high agreement among raters. The participants were asked to rate the child's responses to the examiner on a 4-point scale. It is important to note that because the audio samples were taken from interactive narrative samples, there was great variability in the response speed among the audio clips from each group of children (see Table 2). The 37 monolingual English-speaking participants received limited training about response speed before the task began. Yet, high interrater agreements about the response speed of the children in the clips were found for the Cantonese items (Kalpha = 0.82; 95% CI = 0.81, 0.82) and for the English items (Kalpha = 0.75; 95% CI = 0.75, 0.76). This finding indicates monolingual clinicians or teachers could rate the response speed with a high agreement level. Future research is needed to replicate the high interrater agreement in the response-speed judgment task using stimuli from bilingual children who learn other languages as a home language.

In this study, the primary question is whether the response speed rating (slow vs. normal speed) by monolingual adults predict the audio clip type (LI vs. TD). The 48 audio clips were from 6 Cantonese-English bilingual children: 3 with LI and 3 TD children. Consistent with prior findings [22,24,26], children with LI, as a group, had significantly slower response speed than TD children in both Cantonese and English conditions (see Table 2). Although the monolingual English-speaking participants did not speak Cantonese (L1), they could still identify children who were slower to respond in Cantonese. For the Cantonese clips, the monolingual adults' ratings appear to have excellent sensitivity and excellent specificity (100% and 100%, respectively; Figure 1a). For the English clips, the ratings by the monolingual adults appear to have good sensitivity and excellent specificity (75% and 92%, respectively; Figure 1b). There are three possible explanations for the difference between the Cantonese and English conditions. The first explanation is related to the monolingual participants' knowledge of English. The participants are monolingual English-speaking and have no knowledge of Cantonese. When making a judgment on the English clips, they could be distracted by the linguistic contexts. Although the participants were instructed to use response time in making their judgements solely, other linguistic cues (e.g., prosody, vocabulary, grammar, or intonation) in the audio clips could have implicitly affected their ratings. In contrast, because the participants did not have any exposure to Cantonese, they were likely to focus solely on children's response speed. The second explanation is that the audio clips were sampled from children who were at the beginning stage of learning L2. These children, as a group, had more L1 experience and stronger L1 skills at the time of testing. As a result, the variability cross children's response-to-examiner in the L2 clips were high (*SD* = 4.13 for the TD children) compare to the L1 clips (*SD* = 0.069 for the TD children). It is likely the variability for the clips in L2 contributes to the variability of ratings in L2. The third explanation might be related to the response speed difference between the Cantonese and English samples. For the Cantonese samples, the mean response speed of the TD children ranged from 0.64 to 0.75 s, but the response speed of the children with LI was between 1.48 to 5.5 s. For the English samples, the response speed of the TD children was from 0.88 to 1.66 s, but the mean response speed of the children with LI was from 1.96 to 2.25 s. The larger LI–TD contrast in the Cantonese clips may have contributed to the high sensitivity and specificity. Future studies are needed to investigate the LI-TD contrast across L1 and L2 samples in diagnostic accuracy.

The present study examined the response speed ratings of the audio clips nested within each child (four clips × two languages). One limitation is that the stimuli were developed using a small number of children's narrative samples ($n = 6$). Although we had a larger database of language samples from Cantonese-English children with LI, many children with LI did not meet the selection criteria (e.g., three exchanges between the examiner and child for both Cantonese and English). Some children did not have three exchanges, while some children's language samples in Cantonese met the selection criteria, but their English samples did not. Future work should use less restrictive criteria to include samples from more children to validate the response-speed judgment task. A second limitation is that the clips were selected from children's narrative samples. The response speed to the examiner varies within each child. For the Cantonese samples, the overall ratings appear to differentiate children with LI from TD children accurately (Figure 2). However, the ratings for the English samples appear to be less accurate. In particular, the ratings for Child 3 (with LI) and Child 4 (TD) were at chance or almost at the chance, respectively. As noted in Table 2, Child 3, although had a diagnosis of LI, was faster than the other two children in the LI group in the English conditions (Child 2 and Child 5). In contrast, Child 4 did not appear to be significantly faster or slower than the other two TD children. Future investigation with larger samples is needed in order to examine the response speed threshold's effect on listeners' judgment.

In the search for screening tools across a wide variety of languages, this study explores the methodology that allows clinicians to identify at-risk children on the basis of their response speed. The findings in this study provide some preliminary evidence for including a response-speed judgment task as a screening tool for monolingual English-speaking speech-language pathologists who work with bilingual preschool children. Future work needs to be done to examine how judging children's response speed is incorporated in classroom observation and parent or teacher reports.

Author Contributions: Conceptualization, P.F.K. and A.M.; methodology, P.F.K., A.M. and S.S.; software, P.F.K.; formal analysis, P.F.K., A.M. and S.S.; investigation, A.M.; data curation, P.F.K. and A.M.; writing—P.F.K. and A.M.; writing—review and editing, P.F.K., A.M. and S.S.; visualization, P.F.K.; supervision, P.F.K.; project administration, P.F.K. and A.M.; funding acquisition, P.F.K. and A.M. All authors have read and agreed to the published version of the manuscript.

Funding: The stimuli development of this research was funded by National Institute of on Deafness and Other Communication Disorders (NIDCD; R03 DC14059-01), awarded to Pui Fong Kan. The data collection portion of this study was funded by the Undergraduate Research Opportunities Program at the University of Colorado Boulder.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board of The University of Colorado Boulder (Protocol #: 18-0277, 9 November 2018).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to original informed consent provisions.

Acknowledgments: The authors wish to thank the parents and children whose participation made the development of the stimuli possible. Appreciation is extended to Erika Soderholm for her help in the initial stages of this project.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bishop, D.V.M.; Snowling, M.J.; Thompson, P.A.; Greenhalgh, T. Phase 2 of CATALISE: A multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *J. Child Psychol. Psychiatry* **2017**, *58*, 1068–1080. [CrossRef] [PubMed]
2. Norbury, C.F.; Gooch, D.; Wray, C.; Baird, G.; Charman, T.; Simonoff, E.; Vamvakas, G.; Pickles, A. The impact of nonverbal ability on prevalence and clinical presentation of language disorder: Evidence from a population study. *J. Child Psychol. Psychiatry* **2016**, *57*, 1247–1257. [CrossRef] [PubMed]
3. Tomblin, J.B.; Records, N.L.; Buckwalter, P.; Zhang, X.; Smith, E.; O'Brien, M. Prevalence of specific language impairment in kindergarten children. *J. Speech Lang. Hear. Res.* **1997**, *40*, 1245–1260. [CrossRef]

4. Leonard, L. *Children with Specific Language Impairment*, 2nd ed.; MIT Press: Cambridge, UK, 2014; ISBN 978-0-262-02706-9.
5. Bedore, L.M.; Pena, E.D. Assessment of bilingual children for identification of language impairment: Current findings and implications for practice. *Int. J. Biling. Educ. Biling.* **2008**, *11*, 1–30. [CrossRef]
6. Grimm, A.; Schulz, P. Specific language impairment and early second language acquisition: The risk of over- and underdiagnosis. *Child Indic. Res.* **2014**, *7*, 821–841. [CrossRef]
7. Paradis, J. The interface between bilingual development and specific language impairment. *Appl. Psycholinguist.* **2010**, *31*, 227–252. [CrossRef]
8. Morgan, P.L.; Farkas, G.; Hillemeier, M.M.; Mattison, R.; Maczuga, S.; Li, H.; Cook, M. Minorities are disproportionately under-represented in special education: Longitudinal evidence across five disability conditions. *Educ. Res.* **2015**, *44*, 278–292. [CrossRef]
9. Rethfeldt, S.W. Speech and language therapy services for multilingual children with migration background: A cross-sectional survey in Germany. *Folia Phoniatr. Logop.* **2019**, *71*, 116–126. [CrossRef]
10. Bedore, L.M.; Peña, E.D.; Griffin, Z.M.; Hixon, J.G. Effects of age of English exposure, current input/output, and grade on bilingual language performance. *J. Child Lang.* **2016**, *43*, 687–706. [CrossRef]
11. Cheung, S.; Kan, P.F.; Winicour, E.; Yang, J. Effects of home language input on the vocabulary knowledge of sequential bilingual children. *Biling. Lang. Cogn.* **2019**, *22*, 986–1004. [CrossRef]
12. Kan, P.F.; Huang, S.; Winicour, E.; Yang, J. Vocabulary growth: Dual language learners at risk for language impairment. *Am. J. Speech Lang. Pathol.* **2020**, *29*, 1178–1195. [CrossRef] [PubMed]
13. Dollaghan, C.; Horner, E. Bilingual language assessment: A meta-analysis of diagnostic accuracy. *J. Speech Lang. Hear. Res.* **2011**, *54*, 1077–1088. [CrossRef]
14. Boerma, T.; Blom, E. Assessment of bilingual children: What if testing both languages is not possible? *J. Commun. Disord.* **2017**, *66*, 65–76. [CrossRef] [PubMed]
15. American Speech-Language-Hearing Association (ASHA). *Demographic Profile of ASHA Members Providing Bilingual Services, Year-End 2019*; ASHA: Rockville, MD, USA, 2019.
16. Caesar, L.G.; Kohler, P.D. The state of school-based bilingual assessment: Actual practice versus recommended guidelines. *Lang. Speech Hear. Serv. Sch.* **2007**, *38*, 190–200. [CrossRef]
17. Laing, S.P.; Kamhi, A. Alternative assessment of language and literacy in culturally and linguistically diverse populations. *Lang. Speech Hear. Serv. Sch.* **2003**, *34*, 44–55. [CrossRef]
18. Girbau, D.; Schwartz, R.G. Phonological working memory in Spanish–English bilingual children with and without specific language impairment. *J. Commun. Disord.* **2008**, *41*, 124–145. [CrossRef]
19. Guiberson, M.M.; Rodríguez, B.L. Working memory and linguistic performance of dual language learners with and without developmental language disorders. *Am. J. Speech Lang. Pathol.* **2020**, *29*, 1301–1306. [CrossRef]
20. Thordardottir, E.; Brandeker, M. The effect of bilingual exposure versus language impairment on nonword repetition and sentence imitation scores. *J. Commun. Disord.* **2013**, *46*, 1–16. [CrossRef]
21. Ebert, K.D.; Pham, G. Including nonlinguistic processing tasks in the identification of developmental language disorder. *Am. J. Speech Lang. Pathol.* **2019**, *28*, 932–944. [CrossRef]
22. Kohnert, K.; Windsor, J. The search for common ground: Part II. Nonlinguistic performance by linguistically diverse learners. *J. Speech Lang. Hear. Res.* **2004**, *47*, 891–903. [CrossRef]
23. Windsor, J.; Kohnert, K.; Loxtercamp, A.L.; Kan, P.-F. Performance on nonlinguistic visual tasks by children with language impairment. *Appl. Psycholinguist.* **2008**, *29*, 237–268. [CrossRef]
24. Park, J.S.; Miller, C.A.; Sanjeevan, T.; van Hell, J.G.; Weiss, D.J.; Mainela-Arnold, E. Bilingualism and processing speed in typically developing children and children with developmental language disorder. *J. Speech Lang. Hear. Res.* **2020**, *63*, 1479–1493. [CrossRef] [PubMed]
25. Leonard, L.B.; Ellis Weismer, S.; Miller, C.A.; Francis, D.J.; Tomblin, J.B.; Kail, R.V. Speed of processing, working memory, and language impairment in children. *J. Speech Lang. Hear. Res.* **2007**, *50*, 408–428. [CrossRef]
26. Montgomery, J.W.; Windsor, J. Examining the language performances of children with and without specific language impairment: Contributions of phonological short-term memory and speed of processing. *J. Speech Lang. Hear. Res.* **2007**, *50*, 778–797. [CrossRef]
27. Ebert, K.D.; Kohnert, K. Sustained attention in children with primary language impairment: A meta-analysis. *J. Speech Lang. Hear. Res.* **2011**, *54*, 1372–1384. [CrossRef]
28. Archibald, L.M.D.; Gathercole, S.E. The complexities of complex memory span: Storage and processing deficits in specific language impairment. *J. Mem. Lang.* **2007**, *57*, 177–194. [CrossRef]
29. Kail, R. A method for studying the generalized slowing hypothesis in children with specific language impairment. *J. Speech Lang. Hear. Res.* **1994**, *37*, 418–421. [CrossRef]
30. Miller, C.; Kail, R.; Leonard, L.; Tomblin, J. Speed of processing in children with specific language impairment. *J. Speech Lang. Hear. Res.* **2001**, *44*. [CrossRef]
31. Lahey, M.; Edwards, J. Why do children with specific language impairment name pictures more slowly than their peers? *J. Speech Lang. Hear. Res.* **1996**, *39*, 1081–1098. [CrossRef]
32. Noonan, N.B.; Redmond, S.M.; Archibald, L.M.D. Contributions of children's linguistic and working memory proficiencies to their judgments of grammaticality. *J. Speech Lang. Hear. Res.* **2014**, *57*, 979–989. [CrossRef]

33. Park, J.; Miller, C.A.; Mainela-Arnold, E. Processing speed measures as clinical markers for children with language impairment. *J. Speech Lang. Hear. Res.* **2015**, *58*, 954–960. [CrossRef] [PubMed]
34. Bialystok, E. Global–local and trail-making tasks by monolingual and bilingual children: Beyond inhibition. *Dev. Psychol.* **2010**, *46*, 93–105. [CrossRef]
35. Martin-Rhee, M.M.; Bialystok, E. The development of two types of inhibitory control in monolingual and bilingual children. *Biling. Lang. Cogn.* **2008**, *11*, 81. [CrossRef]
36. Thomas, K.M.; Nelson, C.A. Serial reaction time learning in preschool- and school-age children. *J. Exp. Child Psychol.* **2001**, *79*, 364–387. [CrossRef] [PubMed]
37. Glascoe, F.P. Can clinical judgment detect children with speech-language problems? *Pediatrics* **1991**, *87*, 317–322. [PubMed]
38. Gutiérrez–Clellen, V.F.; Kreiter, J. Understanding child bilingual acquisition using parent and teacher reports. *Appl. Psycholinguist.* **2003**, *24*, 267. [CrossRef]
39. Klee, T.; Pearce, K.; Carson, D.K. Improving the positive predictive value of screening for developmental language disorder. *J. Speech Lang. Hear. Res.* **2000**, *43*, 821–833. [CrossRef] [PubMed]
40. Glascoe, F.P. Evidence-based approach to developmental and behavioral surveillance using parents' concerns. *Child Care Health Dev.* **2000**, *26*, 137–149. [CrossRef]
41. Wallisch, A.; Little, L.; Dean, E.; Dunn, W. Parent concerns: Differentiating developmental diagnoses in young children. *Res. Dev. Disabil.* **2020**, *103*, 103684. [CrossRef]
42. Bedore, L.M.; Pena, E.D.; Joyner, D.; Macken, C. Parent and teacher rating of bilingual language proficiency and language development concerns. *Int. J. Biling. Educ. Biling.* **2011**, *14*, 489–511. [CrossRef]
43. Pua, E.P.K.; Lee, M.L.C.; Rickard Liow, S.J. Screening bilingual preschoolers for language difficulties: Utility of teacher and parent reports. *J. Speech Lang. Hear. Res.* **2017**, *60*, 950–968. [CrossRef] [PubMed]
44. Coppens-Hofman, M.C.; Terband, H.; Snik, A.F.M.; Maassen, B.A.M. Speech characteristics and intelligibility in adults with mild and moderate intellectual disabilities. *Folia Phoniatr. Logop.* **2016**, *68*, 175–182. [CrossRef]
45. Einarsdóttir, J.; Ingham, R.J. Does language influence the accuracy of judgments of stuttering in children? *J. Speech Lang. Hear. Res.* **2009**, *52*, 766–779. [CrossRef]
46. Formby, C.; Yang, X.; Payne, J.; Parton, J. Group means and intersubject and intrasubject variability estimates for absolute and relative (categorical) loudness judgments of typical young adult listeners. *J. Speech Lang. Hear. Res.* **2019**, *62*, 768–778. [CrossRef]
47. Lee, A.; Potts, S.; Bressmann, T. Speech-language therapy students' auditory—Perceptual judgements of simulated concurrent hypernasality and articulation disorders. *Clin. Linguist. Phon.* **2020**, *34*, 479–492. [CrossRef]
48. Kim, Y.; Choi, Y. A cross-language study of acoustic predictors of speech intelligibility in individuals with Parkinson's disease. *J. Speech Lang. Hear. Res.* **2017**, *60*, 2506–2518. [CrossRef] [PubMed]
49. Bunton, K.; Kent, R.D.; Duffy, J.R.; Rosenbek, J.C.; Kent, J.F. Listener agreement for auditory-perceptual ratings of dysarthria. *J. Speech Lang. Hear. Res.* **2007**, *50*, 1481–1495. [CrossRef]
50. Schölderle, T.; Haas, E.; Ziegler, W. Age norms for auditory-perceptual neurophonetic parameters: A prerequisite for the assessment of childhood dysarthria. *J. Speech Lang. Hear. Res.* **2020**, *63*, 1071–1082. [CrossRef]
51. Boersma, P.; Weenink, D. *Praat: Doing Phonetics by Computer*; Version 6.0.37; Institute of Phonetic Sciences, University of Amsterdam: Amsterdam, The Netherlands, 2018; Available online: http://www.praat.org/ (accessed on 25 March 2019).
52. Roid, G.H.; Miller, L.J. Leiter international performance scale-revised: Examiner's manual. In *Leiter International Performance Scale-Revised*; Roid, G.H., Miller, L.J., Eds.; Stoelting: Wood Dale, IL, USA, 1997.
53. Mayer, M. Frog, Where Are You? Dial Books: New York, NY, USA, 2003; ISBN 978-0-8037-2881-3.
54. Hayes, A.F.; Krippendorff, K. Answering the call for a standard reliability measure for coding data. *Commun. Methods Meas.* **2007**, *1*, 77–89. [CrossRef]

children

MDPI

Article

Is There an Association between Executive Function and Receptive Vocabulary in Bilingual Children? A Longitudinal Examination

Vanessa Diaz [1,*], Maria Borjas [2] and M. Jeffrey Farrar [3]

[1] Department of Psychology, Virginia Tech, Blacksburg, VA 24061, USA
[2] Department of Psychology, The University of Houston, 3695 Cullen Boulevad, Houston, TX 77204, USA; mborjas021@gmail.com
[3] Department of Psychology, University of Florida, Gainesville, FL 32611, USA; farrar@ufl.edu
* Correspondence: vdiaz@vt.edu; Tel.: +1-540-231-0973

Abstract: Dual language management has been proposed as the reason for bilingual children's sometimes enhanced executive functioning (EF). We sought to identify the directionality of the relation between language proficiency and EF, using measures of receptive vocabulary, inhibitory control, and cognitive flexibility. Data were collected twice, a year apart, on 35- to 66.8-month-old bilingual (n = 41, M = 49.19 months) and monolingual preschool children (n = 37, M = 47.82 months). The longitudinal results revealed that while the monolingual children's vocabulary at Time 1 predicted EF at Time 2, EF at Time 1 did not predict vocabulary at Time 2. In contrast, for bilingual children the relation was not present at all. The results were similar after the one-time analyses. The absence of relations between EF and language in bilinguals, while present in monolinguals, challenges the current conceptualization of the EF advantage in bilinguals, and emphasizes the need for more research on the development of bilingual children.

Keywords: bilingualism; executive functioning; receptive vocabulary; language development; longitudinal

Citation: Diaz, V.; Borjas, M.; Farrar, M.J. Is There an Association between Executive Function and Receptive Vocabulary in Bilingual Children? A Longitudinal Examination. *Children* **2021**, *8*, 44. https://doi.org/10.3390/children8010044

Received: 14 December 2020
Accepted: 8 January 2021
Published: 13 January 2021

1. Introduction

The specific nature of the relation between language development and executive functioning (EF) has been much examined and greatly debated (e.g., [1–3]). EF refers to higher-order cognitive processes such as attention management, planning, monitoring, and inhibition of habitual responses [1]. Previous longitudinal studies on this topic on monolingual children have found varying associations in terms of directionality, suggesting a number of possible underlying mechanisms. These suggested mechanisms include simultaneous and interactive development [3]; primacy of EF relative to language development [4]; the influence of a potential third factor, such as processing speed [5,6]; and that language proficiency may act as a mediator between EF and other factors, such as phonological awareness. Other studies have found that language proficiency mediates EF group differences but not the reverse [7], or in the case of Slot and Suchodoletz (2018) [3], that even when the relationship is bidirectional, earlier language proficiency is a stronger predictor of later EF rather than the reverse. We will review these findings in greater detail, as well as the available findings on bilingual populations.

For bilingual children, determining the directionality of the relation between receptive vocabulary skills, a measure of language proficiency, and EF is particularly interesting in light of the accumulating evidence showing differential development between bilingual and monolingual children on a number of related socio-cognitive and psycholinguistic capacities. For example, we see performance differences between bilingual and monolingual children in theory of mind development (i.e., [8]), metalinguistic awareness [9], as well as on EF [10] and language development themselves [11]. In this paper, we are interested in the following question: Does potentially differential development in EF and receptive vocabulary compared to monolingual children result in a comparatively different relation between them in bilingual children?

Specifically, a number of studies have found that bilingual children have an advantage when it comes to EF compared to monolingual children [12–16], although this has been recently challenged (see [17,18]). This potential EF advantage in bilinguals has been proposed as resulting from the nature of bilingual language use (i.e., [10]), which requires efficient language monitoring, selection, inhibition of the non-selected language, and cognitive flexibility to switch as appropriate, thus resulting in a strengthened EF system [19]. A bilingualism effect on EF has been reported in children of different ages and cultural backgrounds, such as 8-year-old Canadian and Indian bilingual compared to monolingual Canadian children [10], Spanish–English bilinguals compared to English monolinguals, and English speakers enrolled in a second-language immersion kindergarten [1], as well as in toddlers [15] and infants [20]. On the other hand, bilingual children have also been shown to underperform in language proficiency, such as measures of receptive vocabulary, a difference that has been shown to be stable over time [11], though these assessments are usually limited to receptive vocabulary in their dominant language. Given the characterization in previous research that the bilingual EF enhancement on cognitive flexibility and inhibitory control comes from using these capacities to develop proficiency in two languages, here we ask instead if the reverse could also be the case, could a strengthened EF lead to improved language outcomes?

Previous longitudinal studies on the relation between EF and language proficiency in monolingual children have produced several findings. In a recent longitudinal study with measurements taken over a one-year period, language skills, assessed through receptive vocabulary and the comprehension and imitation of grammatical structures, were a significant predictor of later EF performance for monolingual three- to four-year-old children [3]. The authors used the Pencil Tapping task, Dimensional Change Card Sort (DCCS) task, Forward Digit Span, and Copy Hand Movement task as their EF measures to assess inhibitory control, attention shifting, and working memory. Although this relation was bidirectional, language skills were a stronger predictor of later EF performance than EF was of later language skills [3]. Based on this finding, the authors proposed that the nature of a child's language development may suggest that language skills and EF skills develop simultaneously and interactively. In contrast, Weiland and colleagues (2014) [4] reported a different pattern of relations in a sample of monolingual children. In their longitudinal study, the authors found that EF skills at Time 1 predicted receptive vocabulary at Time 2 and, unlike Slot and Suchodoletz [3], they did not find this relation to be bidirectional. The authors argued that the direction of this relationship is intuitive due to the fact that preschool curricula tend to target receptive vocabulary, with not much emphasis on EF, thus allowing EF to develop on its own [4]. It is worth noting that Weiland et al. (2013) used the Pencil Tap task, Forward Digit Span task, and Backward Digit Span task to measure EF, and the Peabody Picture Vocabulary Test to measure receptive vocabulary, similarly to Slot and Suchodoletz [3]. Kaushanskaya and colleagues [21] reported a similar finding with an older sample of eight- to ten-year-old monolingual children. They assessed EF using the Flanker task, Go/No-Go task, n-back task, Corsi Block task, and the DCCS to measure inhibition, working memory, and task shifting; and comprehensive language assessments, including receptive vocabulary and the Clinical Evaluation of Language Fundamentals (CELF). The authors reported a predictive relationship between working memory and language proficiency, but there was an overall weak relation between language scores and other aspects of EF, such as inhibition and shifting [21].

For a different pattern of results, another longitudinal study on monolingual children in which data were collected at ages 4, 5, 6, and 7, a stable relation between language and EF was found at each time point, but interestingly, no directional association over time was found between the two factors [5]. The tasks used to assess EF included the Head Toes Knees and Shoulders task, Block Recall task, and the Go/No-Go task to measure inhibition, selective attention, and visuo-spatial memory. Language was assessed using the Receptive One Word Picture Vocabulary Test (ROWPVT), Experimental Sentence Imitation Task, and two subsets (sentence structure and expressive vocabulary) of the CELF. The authors argue that these results may be due to a third factor, such as processing speed, influencing both. In contrast, a study with a sample composed of deaf and hearing children found that language mediated non-verbal EF differences between the groups, but that non-verbal EF performance did not mediate language group differences [7]. The authors used a long array of executive function measures, including the Odd One Out Span task, Backward Spatial Span task, Design Fluency task, Children's Color Trails Tests 1 and 2, Tower of London task, and (computerized) Simon task to assess visuospatial working memory and cognitive fluency, shifting, planning, and inhibitory control. The Expressive One-Word Picture Vocabulary Test (EOWPVT) was used to assess language skills. These results suggest that in a sample with non-normative language development, language proficiency became more influential. For a different kind of language measure, Wilbourn and colleagues [6] looked at the relation between phonological awareness, receptive vocabulary, and EF (assessed using the Lexical Stroop task and the DCCS task to assess cognitive flexibility and attention) in a sample of five- to eight-year-old English monolinguals. They proposed that vocabulary proficiency may act as a mediator in the relation between performance on an EF task and phonological awareness, suggesting a much more intricate relation between language development and EF.

Overall, studies on monolingual children have found support for all three longitudinal relationships: EF predicting language proficiency, language proficiency predicting EF, and bidirectional relations. Taken together, we propose that this likely points to the relationship between EF and receptive vocabulary skills, an aspect of language proficiency, at least in monolingual children, as being simultaneous, interactive, and also bidirectional.

As reviewed, the majority of our current understanding of the longitudinal relation between EF and language development comes from monolingual samples. When it comes to bilingual children, a number of researchers have examined this question with the purpose of finding out whether and when there is a bilingual advantage on EF [12–15]. For example, in the only longitudinal study of this kind, the authors found an inhibitory control EF advantage for bilinguals when testing two- to three-year-old monolinguals and bilinguals at six months apart [22]. Specifically, knowledge of translation equivalents assessed through a parental report vocabulary checklist was associated with better performance on the EF tasks assessing cognitive flexibility, working memory, inhibitory control, and response suppression (Reverse Categorization task, Shape Stroop task, Gift Delay task, and the Multilocation task), suggesting a connection between proficiency on both languages and EF [22]. Importantly, however, the directionality of this relationship was not examined.

When it comes to single time point studies examining the relation between language development and EF in bilinguals, Carlson and Meltzoff [1] found enhanced EF abilities for bilinguals related to specific language exposure. The authors found an EF advantage for five- to seven-year-old bilingual children compared to both monolinguals and children in an immersion program on the battery of conflict EF tasks compared to the battery of EF delay tasks. In turn, Iluz-Cohen and Armon-Lotem [23] found that four- to seven-year-old bilingual children with high language proficiency performed better at tasks assessing inhibition and shifting than participants with low language proficiency, as measured by the Goralnik Diagnostic Test for Hebrew and the CELF. So, while some studies have focused on the nature of the relation between language proficiency and EF for this group, there are no longitudinal investigations that draw their attention to the directionality of these relations in bilingual children.

In the current investigation we examine the directionality of this relationship across time in bilinguals and monolinguals to answer the following questions: Does more efficient EF at one point predict improved language proficiency a year later as measured by receptive vocabulary? Does higher language proficiency, as measured by receptive vocabulary, instead predict a more efficient EF later on? Is the relation bidirectional? To our knowledge, our study is the first to examine the longitudinal relation between EF and receptive vocabulary skills in bilingual preschool children. Given the variations that exist between language and EF for bilinguals and monolinguals, a longitudinal analysis may enhance our understanding of the relation between EF and language for the two groups.

As described, the motivation for examining this question in bilingual children regards the cognitive flexibility and inhibitory control required to manage dual linguistic representations by inhibiting one language and actively selecting another. What would the pattern of a single time point and longitudinal associations be in this group given this particular need, as well as the pattern of relations described above for monolingual children? For bilingual children we see two possibilities. On the one hand, effective learning of two languages as reflected in vocabulary scores at one time may predict later enhanced EF through a generalization of the control mechanism, or language learning as a tool for attention control. On the other hand, stronger EF at one point may in turn facilitate subsequent bilingual language development (demonstrated in vocabulary scores) by facilitating the management of dual linguistic representations, thus enhancing cognitive flexibility and inhibition, as well as receptive vocabulary skills. Both of these possibilities carry theoretical significance, firstly for our understanding of the relation between language development and EF, and secondly for our understanding of bilingual development.

Current Study

In the present study, we aim to identify the directionality of the relation between language proficiency, as measured through receptive vocabulary, and EF using measures of inhibitory control and cognitive flexibility, in samples of monolingual and bilingual children using two time points a year apart. We are interested in analyzing whether receptive vocabulary skills predict EF in monolingual and bilingual children and/or whether EF, in turn, predicts receptive vocabulary skills. The absence in the literature on this relationship, focusing on bilingual children specifically, leaves an important gap in our understanding of bilingual children's development. We are particularly interested in possible differences in these relations for bilingual and monolingual children, which may suggest differences in the mechanisms underlying development in these two groups.

Our specific research questions are as follows: (1) Does a longitudinal relation exist between language proficiency, as measured by receptive vocabulary, and EF, similar to the literature on monolingual children? If so, does this measure of language proficiency at Time 1 predict EF at Time 2, and/or does EF at Time 1 predict receptive vocabulary assessed at a Time 2? (2) Are there identifiable differences in the directionality of the relation between receptive vocabulary skills and EF for monolingual and bilingual children? We use receptive vocabulary as our language proficiency measure because it captures the understanding of linguistic labels that young language learners may not yet feel comfortable producing, which is particularly relevant for dual language learners. Moreover, we theorize that bilingual learners would engage EF resources to manage dual linguistic representations as denoted by their total vocabulary knowledge, which includes understanding of labels in both languages. We selected EF tasks as well-established conflict measures tapping into inhibitory control and cognitive flexibility.

2. Materials and Methods

2.1. Participants

Our sample was composed of 41 Spanish–English bilingual children and 37 English monolinguals. Their ages ranged from 35 to 66.8 months at Time 1 (Bilinguals M = 49.19 months, SD = 7.32; Monolinguals M = 47.82 months, SD = 6.94). Because of this range, we controlled for age on all of our analyses. Participants were recruited from predominantly bilingual and monolingual preschools in the Southeastern United States, and this protocol was approved by the University of Florida's Institutional Review Board (UFIRB #2011-U-451). Bilingualism was defined by regular exposure to both English and Spanish, and parental reports of fluency in both languages. This was assessed through parent language questionnaires where parents were asked "1. Which language/languages does your child speak (may list more than one)?"; "2. Which one is your child's preferred language?"; "3. Which language/languages are regularly spoken in your home?"; and "4. Which language/languages is the child exposed to in Preschool?"—among other language-exposure questions. Only children whose parents listed two languages for Question 1 and for either Questions 3 or 4 were considered bilinguals. All such parents listed Spanish and English for Question 1. Of this sample of Spanish–English bilinguals, a large majority (77.5%) were exposed to both languages since birth (see Table 1). The non-crib bilinguals had all been exposed to their second language for one year or more.

Table 1. Bilingual children's language exposure parental questionnaire.

	% English Only	% Spanish Only	% Both
Language exposed at birth	6	15	77.5
Child's preferred language	46.2	48.7	5.1
Spoken at home	12.5	55	27.5
Spoken at preschool	25.6	7.7	61.5
Language mother speaks to child	15.4	59	25.6
Language father speaks to child	10.3	53.8	30.8
Language other adult in the home speaks to child	-	66.7	33.3

No significant differences were found between the bilingual and monolingual groups on gender and SES (parental marital status, and the education, employment, and occupational status of both parents), nor for the primary caregiver's level of education (all p-values > 0.05; presented in Table 3). A year later, the children were assessed again. The final sample was composed of 25 bilinguals and 27 monolingual children (Bilinguals M = 56.97 months, SD = 5.25; Monolinguals M = 57.22 months, SD = 6.68). Because testing was conducted over the summer months, attrition was due to the children moving away or changing schools, and was not significantly different for the bilingual and monolingual groups, thus not systematic for our main comparisons. In addition, we ran comparisons on the demographic variables and our main variables of interest at Time 1 between those children who returned for Time 2 and those who did not. We found no significant differences between the return and no-return sample on gender and parental education. We did find differences for age ($t(76) = 2.84$, $p < 0.01$) and SES ($t(69) = -2.47$, $p < 0.05$). The no-return sample was older (return M = 46.82, SD = 5.80 vs. no-return M = 51.98, SD = 8.32) and had lower SES. These differences lead us to speculate that the older and lower-income children were more likely to start attending Voluntary Prekindergarten (VPK), a free educational program for 4-year-olds in the region. In terms of our variables of interest (EF and vocabulary), once controlling for differences in age, comparisons between the return and no-return sample revealed no differences. These results give us confidence to report on the entire Time 1 sample, as well as on the return sub-sample.

2.2. Vocabulary Measure

Receptive One Word Picture Vocabulary Test [24]: The ROWPVT is normed for examinees 2–80 years of age and has a median reliability of 0.97 across all ages. Children are shown test plates with four pictures each and asked to point to the picture that best represents a word spoken by the experimenter. Testing begins with words normed on the child's chronological age and increases in difficulty until eight consecutive errors are made. The test is scored by deducting errors from the number of the final item administered.

The English monolingual version was used for the monolingual children at the beginning of the testing session, followed by the EF measures. For bilingual children a Spanish–English bilingual version of the test was used to assess proficiency in both languages using the same test plates. The test was first administered in the child's non-preferred language, as reported in the parent language questionnaire (Table 1), because the test stimuli were the same for both languages. The test was then re-administered in their preferred language at the end of the testing session, after the EF measures. For children whose parents listed that the child had no preference, the child was asked for their preference to conduct the session at the time of testing and followed the procedure above. The testing then proceeded with the EF measures in the child's preferred language. At the end of the session, the vocabulary test was re-administered in the children's preferred language by the same research assistant who had native-like proficiency in both languages. Results for the bilingual children were tallied for both a conceptual score and a total score, as described below.

Conceptual Score: A conceptual score was computed by adding the total number of concepts the children pointed to correctly in either one or both languages. This effectively assessed the number of "concepts" the children recognized, regardless of the language used by the investigator. This is thought to more adequately represent their language competence versus only analyzing their dominant language [25] compared to monolinguals.

Total Score: We computed a total language score by adding the bilingual children's separate scores in both Spanish and English. While it may not always be appropriate to use this score in comparisons with monolinguals, because it counts translation equivalents as two individual items, we were interested in whether competency in both languages may be a good measure of EF's involvement in the bilingual's effective acquisition of two lexicons. These values are presented in Table 2.

2.3. Executive Functioning Measures

Executive functioning was assessed using the following several measures focusing on inhibitory control and cognitive flexibility. A composite score was created by obtaining the z-score of each task due to differing number of items in each and adding these z-scores together. Individual EF task group scores and correlations with vocabulary are reported in Tables 2 and 3.

Bear/Dragon Task [26]: For this inhibitory control and cognitive flexibility Simon-Says like measure, children were asked to follow the commands of a bear puppet and ignore the commands of a dragon puppet. The children received points for ignoring the dragon's command, and for following the bear's commands. There were 10 trials split evenly and alternating between the bear commands and dragon commands. This task had a total of 10 possible points.

Happy/Sad Task [27]: For this inhibitory control measure, children were asked to say "happy" when shown a picture of a "sad" face, and "sad" when shown a picture of a "happy" face. The child was given a point for every correct response. There were 16 trials split evenly between "happy" and "sad" faces in random order. This task has a total of 16 possible points.

Dimensional Change Card Sort Task (DCCS) [28]: For this cognitive flexibility measure, the child was asked to sort pictures of red or blue boats and rabbits by color and then by shape. The child was given a point for every picture that was correctly sorted. This task had a total of 6 possible points. Only the switched trials were scored.

Table 2. Partial correlations controlling for age for bilinguals and monolinguals between receptive vocabulary and executive functioning (total receptive vocabulary score between parentheses) at Time 1 and Time 2.

Receptive Vocabulary	Time 1		Time 2	
	Monolinguals	**Bilinguals**	**Monolinguals**	**Bilinguals**
Executive Function Composite	0.41 *	0.13 (0.15)	0.50 *	38 (0.37 ^)
Happy/Sad task	−0.04	−0.06 (−0.02)	0.52 **	0.27 (0.32)
Bear/Dragon task	0.53 *	0.19 (0.24)	0.44 *	0.08 (0.02)
Card Sort task	0.25	0.17 (0.14)	0.21	0.33 (0.25)

Note: ^ $p < 0.10$; * $p < 0.05$, ** $p < 0.01$.

2.4. General Procedure

The data presented here were obtained as part of a larger study assessing the cognitive development in bilingual children. At Time 1, consent was obtained from the parents of the participants through signed letters returned to the children's school with the demographic and language exposure questionnaires. Parents were compensated for filling out the questionnaires. The testing sessions lasted about 45 min and were conducted in quiet offices or extra rooms at the children's preschools. Testing was conducted in English for English monolinguals and in the language selected by the parent as the preferred language for bilingual children. This preference was corroborated by both the children and their teachers. Time 2 data were collected about a year later, using the same procedure.

3. Results

Analyses were first conducted to detect possible demographic differences between the monolingual and bilingual children. No significant differences were found between bilinguals and monolinguals on age, gender, parental level of education, and SES for the total sample at Time 1, nor for the return sample (all p-values > 0.05). In addition, we performed regression analyses for the receptive vocabulary scores based on SES for both groups combined, and separately for monolinguals and bilinguals using both the total and conceptual scores controlling for age (Time 1 data). The analyses revealed a non-significant effect of SES for both groups together, using the conceptual score for bilinguals ($\beta = 0.20$ $t(68) = 1.85$, $p = 0.068$) and for each group separately (monolinguals $\beta = 0.07$ $t(31) = 0.45$, $p = 0.65$; bilingual conceptual $\beta = 0.20$ $t(34) = 1.37$, $p = 0.17$; bilingual total $\beta = 0.22$ $t(34) = 1.50$, $p = 0.14$).

We then examined the receptive vocabulary and EF performance differences between the groups, and the relation between receptive vocabulary and EF performance at Time 1, Time 2, and between Time 1 and Time 2, using partial correlations controlling for age and for autoregressive effects.

3.1. Time 1 and Time 2

Independent sample t-tests between bilinguals and monolinguals on EF at both time points revealed only differences on the DCCS task at Time 2, with monolinguals outperforming bilinguals at $t(50) = 3.53$, $p = 0.001$, $\eta 2 = 0.20$ (see Table 3). For vocabulary, in contrast, we found that monolinguals outperformed bilinguals at both Time 1 ($t(76) = 5.49$, $p = 0.000$, $\eta 2 = 0.28$) and Time 2 ($t(50) = 4.36$, $p = 0.000$, $\eta 2 = 0.27$) using the conceptual score for bilinguals. In turn, bilinguals outperformed monolinguals when using the total score, which added their Spanish and English vocabulary scores (Time 1 ($t(76) = -2.42$, $p = 0.018$ and Time 2 ($t(50) = -3.13$, $p = 0.004$). The total scores are a holistic reflection of language competence in bilinguals and demonstrated no language deficiencies when knowledge of their both languages is considered. These comparisons remained the same when only considering the Time 2 return sample at Time 1.

The main focus of this investigation was the relation between language proficiency and EF for each group. Starting with each time point separately, at Time 1, controlling for age, receptive vocabulary was correlated with the EF composite ($r(34) = 0.41$, $p = 0.014$), but only for the monolingual children (see Table 2). For bilinguals, the correlation between receptive vocabulary and EF was not significant, using both the conceptual score ($r(37) = 0.15$, $p = 0.158$) and the total score. For the return sample only, the results were similar with only a significant correlation between receptive vocabulary and EF for the monolingual sample ($r(24) = 0.47$, $p = 0.018$). Similar to Time 1, for Time 2, also controlling for age, there was a significant relation between receptive vocabulary and EF for the monolingual group ($r(23) = 0.50$, $p < 0.05$), but not for the bilingual group using either the conceptual or total language score for bilinguals. All Time 2 and longitudinal analyses include the return sample only.

Table 3. Bilingual and monolingual children's mean performance (standard deviation in parentheses) and group differences at Time 1 and Time 2.

	Time 1		**Time 2**	
	Monolingual (*n* = 37)	**Bilingual** (*n* = 41)	**Monolingual** (*n* = 27)	**Bilingual** (*n* = 25)
Age (months)	47.82 (6.94)	49.19 (7.32)	57.22 (6.68)	56.97 (5.25)
Gender	16 F, 21 M	19 F, 22 M	10 F, 15 M	9 F, 13 M
SES	29.64 (4.5)	28.02 (5.5)	-	-
Parental Education ^	5.24 (0.98)	5.0 (1.14)	-	-
Executive Function Composite	−0.27 (1.99)	0.26 (2.20)	0.48 (2.17)	−0.53 (1.76)
Happy/Sad task	10.30 (3.57)	11.56 (3.82)	12.26 (3.21)	12.96 (3.48)
Bear/Dragon task	8.65 (1.84)	8.83 (1.96)	9.70 (0.99)	9.44 (1.05)
Card Sort task	4.05 (1.79)	4.20 (1.57)	5.48 (0.98) *	4.20 (1.56) *
Receptive Vocab (Conceptual)	62.89 (12.29) **	47.17 (12.91) **	73.59 (14.19) **	57.44 (12.34) **
English	-	41.24 (16.40)	-	52.70 (13.63)
Spanish	-	34.05 (13.72)	-	43.30 (18.62)
Bilingual Total Vocabulary	-	73.44 (23.74)	-	92.92 (28.10)

Note: * $p < 0.01$; ** $p < 0.001$; ^ Primary caregiver level of education.

3.2. Longitudinal Relations

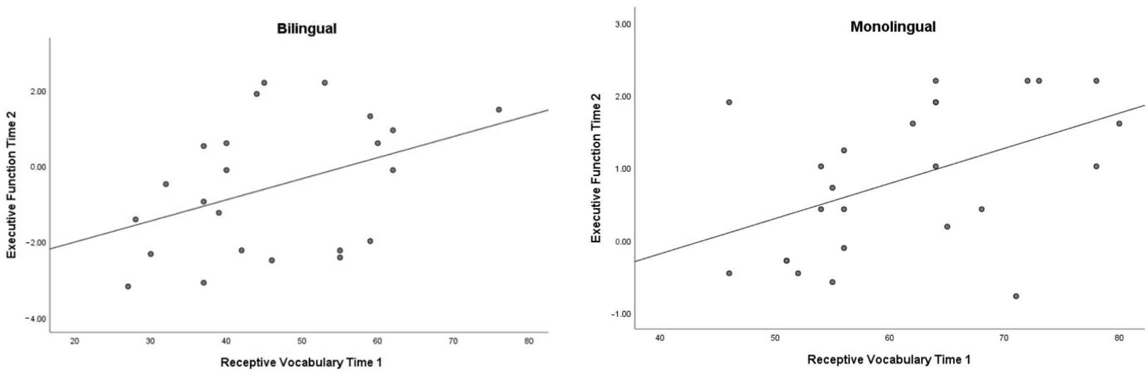
We were particularly interested in analyzing the longitudinal directionality of the relation between receptive vocabulary and EF by using cross-lagged correlation analyses. Specifically, we were interested in which factor (vocabulary, EF) at Time 1 predicted performance on the other factor at Time 2, and in identifying any differences between the groups in these patterns. As in the previous analyses, we controlled for age at both time points. In addition, we also controlled for the effect of the target Time 2 variable at Time 1 to account for autoregressive effects (see Table 4).

Table 4. Partial correlations controlling for age and autoregressive effects between Time 1 and Time 2 executive functioning (EF) and receptive vocabulary for the bilinguals.

	Executive Functioning 2	**Receptive (Conceptual) Vocabulary 2**	**Receptive Total Vocabulary 2**
Bilinguals			
Executive Functioning 1	0.22	0.09	0.13
Receptive Conceptual Vocabulary 1	0.14	0.45 ^	0.15
Receptive Total Vocabulary 1	0.09	0.44 ^	0.47 *
Monolinguals			
Executive Functioning 1	0.15	0.24	-
Receptive Vocabulary 1	0.58 **	0.67 **	-

Note: ^ $p < 0.10$; * $p < 0.05$; ** $p < 0.01$.

For monolinguals, similar to their single time point results, we found that receptive vocabulary at Time 1 was related to EF at Time 2, ($r(21) = 0.58$, $p = 0.004$). However, EF at Time 1 was not related to receptive vocabulary at Time 2 ($r(21) = 0.16$, $p = 0.602$). This suggests that, while for monolinguals language proficiency is significantly predictive of later EF, this relation was not bidirectional. For bilinguals, in contrast, using the receptive conceptual vocabulary score no longitudinal relation was significant; that is, in contrast to our monolingual sample, Time 1 receptive vocabulary did not predict EF at Time 2 for bilinguals ($r(18) = -0.08$, $p = 0.728$), nor did EF at Time 1 predict receptive vocabulary at Time 2 ($r(17) = 0.09$, $p = 0.708$). Using the total vocabulary scores for the bilingual sample did not change the findings (see Table 4). These differences can be observed in Figure 1, where a linear trend between receptive vocabulary at Time 1 and EF can be observed for monolinguals but not for bilinguals. We corroborated these results by running repeated measures ANCOVAs for each group and the EF Time 2 composite as the outcome variable. Receptive vocabulary at Time 1 significantly predicted EF at Time 2, controlling for EF at Time 1 and age at both time points for the monolingual group only (F (1, 25) = 8.56, p = 0.004, η2p = 0.95).



(**a**) (**b**)

Figure 1. Relation between receptive vocabulary at Time 1 and EF at Time 2 for bilinguals (**a**) and monolinguals (**b**).

4. Discussion

For this investigation, we were interested in the relation between two centrally situated cognitive development skills: EF and language proficiency. We analyzed receptive vocabulary, a measure of language proficiency, and EF across two time points a year apart for English monolingual and Spanish–English bilingual children. Previous studies have identified an EF advantage for bilinguals, often hypothesized to emerge from their effective language management [1,13–15,22]. Following this line of reasoning, we hypothesized that a relation should exist between receptive vocabulary and EF for bilingual children. One such relation may involve higher EF, leading to more extensive vocabularies through the effective management of dual linguistic representations as reflected in our total language score, where translation equivalents were counted individually. We also hypothesized that more receptive vocabulary skills could in turn lead to higher EF through a strengthening of control mechanisms that bilinguals may employ during dual language processing and acquisition. A third possibility involved both of these relations through bidirectional longitudinal relations. In the monolingual literature there is support for all three directions of effect between EF and vocabulary acquisition.

For our bilingual sample, we did not find any of the above scenarios. Results of the single time point and longitudinal cross-lagged analysis revealed a relation between receptive vocabulary and EF measures, but only for monolingual children. At both Time 1 and Time 2, receptive vocabulary was significantly correlated with EF only for monolinguals. Importantly, receptive vocabulary at Time 1 predicted EF at Time 2. When controlling for autoregressive effects, the relation was not found to be bidirectional for the monolingual group. These findings support those of previous studies that have identified a significant relation between receptive vocabulary and EF in monolingual children [3–5], especially those finding that receptive vocabulary predicts EF in monolingual children rather than the reverse (i.e., [3,7]). In our sample, monolingual children's language proficiency as measured by receptive vocabulary was predictive of EF performance a year later. A number of possibilities may account for this. In our case, as it is for most EF tasks, there is an important verbal component to the tasks, such as understanding the instructions and the prompts. Another possibility regards the connection between the internalization of private speech, language skills, self-regulation, and cognitive abilities [29,30]. It is possible, for example, that in our sample of monolingual children language skills at Time 1 were instrumental in the development of self-regulation and EF through Time 2. Regarding the non-significant relation between Time 1 EF and later language skills, we posit that perhaps as the instructional environment becomes more rigorous, EF skills become more instrumental for other types of knowledge acquisition than during the preschool years.

In contrast to the consistent findings on monolingual children, no significant relations were found for the bilingual group between receptive vocabulary and EF. This stark difference raises questions about both the established relation between linguistic processes and EF found in monolinguals in previous research in terms of its applicability for bilingual populations, as well as how to characterize the advantage that bilinguals have been reported to have when performing EF tasks (even though that was not the case in our sample). As described, this bilingual advantage is often credited to their effective language management. Given the lack of relation between EF and receptive vocabulary, even for the two languages combined for bilinguals in the current study, there remains a need to better characterize this phenomenon; that is, how effective bilingual children were at acquiring and recognizing labels for normed vocabulary words of the same objects in their two languages combined was not related to their EF skills, as it was for monolingual children in their one language. These results suggest the need for a closer analysis of the relation between language proficiency and EF, and the role these two variables play in children's cognitive development.

In line with our results, a recent study by Nicoladis et al. [31] also failed to find a relation between EF cognitive flexibility, using the DCCS, and measures of either language dominance (parental reports of dominance, relative scores on vocabulary tests, and knowledge of translation equivalents) or language use (living in a bilingual or monolingual community and a language separation task) in French–English bilingual children in Montreal three to six years of age. In addition to the present study, this reiterates the need for future research on the development of bilingual children. We propose that, given the lack of relations detected in this sample, other variables related to language development and EF should be analyzed as part of alternative ways of conceptualizing the development of EF and language proficiency in bilingual children. For example, theory of mind (e.g., [8]) and metalinguistic awareness, such as syntactic awareness [9], have been shown to develop differently in bilingual children and may be related to how bilingual children are making sense of the particularly complex linguistic contexts they experience, instead of EF. Other cognitive skills that may be particularly relevant to bilinguals' receptive vocabulary development (e.g., working memory, phonological awareness) may reveal different learning routes for bilingual vocabulary skills compared to those of monolinguals, and should be addressed by future studies. It follows, for example, that since the mapping between concepts and lexical items is one-to-one in monolinguals and one-to-two in bilinguals, different cognitive skills could be involved. A related issue is that some EF skills, such as inhibition, may be more related to productive measures and not receptive ones, due to the inhibition needed to select the language and produce a word. It is also important to expand this topic of research across different languages and regions. While our monolingual data was collected in Central Florida, the bilingual data was collected from Spanish–English bilinguals in South Florida. South Florida has its own contextual particularities, such as a high number of bilingual speakers (see Table 1), versus other bilingual contexts where the languages may be more compartmentalized. It is possible, for example, that in this context the children do not have to inhibit their language as often, thus accounting for the lack of bilingual advantage in EF; but also that their language learning environment is particularly challenging. This kind of sociolinguistic context may lead to a diminished interaction between EF and linguistic processes compared to other kinds of bilingual contexts where the languages are more compartmentalized than in this largely bilingual community. For example, if one language is spoken at home and a different language is spoken at school, children would need to recruit more EF resources in order to inhibit the "wrong" language, and switch to the appropriate one. In this bilingual sample, however, only 25% of parents reported that only English was spoken at their children's school (see Table 1). This does not explain, however, why there would be a lesser relation compared to monolingual children. Other limitations of the current study include its relatively small sample size compounded by attrition, as addressed in our "Participants" section. In addition, there may be significant variation within our bilingual sample's language exposure and language use, thus making correlations harder to detect. This may be particularly relevant coming from the parental language questionnaires, which may not be entirely accurate at capturing the important distinction between language use versus language exposure in bilinguals [32].

5. Conclusions

The current study further elucidates the relation between language proficiency, as measured by receptive vocabulary, and EF in monolingual children by demonstrating that future EF performance is predicted by previous language proficiency. This study also demonstrates that relations between EF and language in monolingual children, such as the one described, may not necessarily apply to bilingual children. This latter point brings awareness to the importance of continued research on the development of bilingual children. Our results were not congruent with the current conceptualization of the EF advantage in bilingual children found in other samples. In turn, we suggest that future research should examine alternative factors that may be supporting bilingual children's linguistic development, perhaps using a larger sample than the one we were able to retain for the longitudinal analyses. What are bilingual children using to make sense of linguistic input and contexts where two languages are used simultaneous by either the same or different individuals, in the same or different contexts, and with more than likely different levels of proficiency?

Author Contributions: The data presented here come from a study conducted by V.D.; under the supervision of M.J.F.; The data analysis and development of this manuscript were conducted by V.D. and M.B.; with input from M.J.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board of the University of Florida (protocol code UFIRB #2011-U-451, approved on May 2011).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to original informed consent provisions.

Acknowledgments: We would like to thank the Multicultural Academic Opportunities Program (MAOP) at Virginia Tech for allowing our team to come together, the Jacqueline Goldman Graduate Student Research Award at the University of Florida for financial support, and Virginia Tech's Open Access Subvention Fund.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Carlson, S.M.; Meltzoff, A.N. Bilingual experience and executive functioning in young children. *Dev. Sci.* **2008**, *11*, 282–298. [CrossRef]
2. Mazuka, R.; Jincho, N.; Oishi, H. Development of executive control and language processing. *Lang. Linguist. Compass* **2009**, *3*. [CrossRef]
3. Slot, P.L.; Suchodoletz, A. Bidirectionality in preschool children's executive functions and language skills: Is one developing skill the better predictor of the other? *Early Child Res. Q.* **2018**, *42*, 205–214. [CrossRef]
4. Weiland, C.; Barata, M.C.; Yoshikawa, H. The co-occurring development of executive function skills and receptive vocabulary in preschool-aged children: A look at the direction of the developmental pathways. *Infant Child. Dev.* **2014**, *23*. [CrossRef]
5. Gooch, D.; Thompson, P.; Nash, H.M.; Snowling, M.J.; Hulme, C. The development of executive function and language skills in the early school years. *J. Child Psychol. Psychiatry* **2016**, *57*, 180–187. [CrossRef]
6. Wilbourn, M.P.; Kurtz, L.E.; Kalia, V. The Lexical Stroop Sort (LSS) picture-word task: A computerized task for assessing the relationship between language and executive functioning in school-aged children. *Behav. Res. Methods* **2012**, *44*, 270–286. [CrossRef]

7. Botting, N.; Jones, A.; Marshall, C.; Denmark, T.; Atkinson, J.; Morgan, G. Nonverbal executive function is mediated by language: A study of deaf and hearing children. *Child Dev.* **2017**, *88*, 1689–1700. [CrossRef]
8. Diaz, V.; Farrar, M.J. Do bilingual and monolingual preschoolers acquire false belief understanding similarly? The role of executive functioning and language? *First Lang.* **2018**, *38*, 382–398. [CrossRef]
9. Davidson, D.; Rainey, V.R.; Pervez, J. Syntactic awareness in young monolingual and bilingual (Urdu-English) children. *Cogn. Dev.* **2010**, *25*, 166–182. [CrossRef]
10. Bialystok, E.; Viswanathan, M. Components of executive control with advantages for bilingual children in two cultures. *Cognition* **2009**, *112*, 494–500. [CrossRef]
11. Bialystok, E.; Luk, G.; Peets, K.F.; Yang, S. Receptive vocabulary differences in monolingual and bilingual children. *Biling. (Camb. Engl.)* **2010**, *13*, 525–531. [CrossRef]
12. Barac, R.; Bialystok, E. Bilingual effects on cognitive and linguistic development: Role of language, cultural background, and Education. *Child Dev.* **2012**, *83*, 413–422. [CrossRef]
13. White, L.J.; Greenfield, D.B. Executive functioning in Spanish- and English-speaking Head Start preschoolers. *Dev. Sci.* **2017**, *20*. [CrossRef]
14. Gathercole, V.C.M.; Thomas, E.M.; Jones, L.; Guasch, N.V.; Young, N.; Hughes, E.K. Cognitive effects of bilingualism: Digging deeper for the contributions of language dominance, linguistic knowledge, socio-economic status and cognitive abilities. *Int. J. Biling. Educ. Biling.* **2010**, 617–664. [CrossRef]
15. Poulin-Dubois, D.; Blaye, A.; Coutya, J.; Bialystok, E. The effects of bilingualism on toddler's executive functioning. *J. Exp. Child Psychol.* **2010**, *108*, 567–579. [CrossRef]
16. Martin-Rhee, M.M.; Bialystok, E. The development of two types of inhibitory control in monolingual and bilingual children. *Biling. (Camb. Engl.)* **2008**, *11*, 81–93. [CrossRef]
17. Lehtonen, M.; Soveri, A.; Laine, A.; Järvenpää, J.; de Bruin, A.; Antfolk, J. Is bilingualism associated with enhanced executive functioning in adults? A meta-analytic review. *Psychol. Bull.* **2018**, *144*, 394–425. [CrossRef]
18. Paap, K.R.; Johnson, H.A.; Sawi, O. Bilingual advantages in executive functioning either do not exist or are restricted to very specific and undetermined circumstances. *Cortex* **2015**, *69*, 265–278. [CrossRef]
19. Bialystok, E. Bilingualism: The good, the bad, and the indifferent. *Biling. (Camb. Engl.)* **2009**, *12*, 3–11. [CrossRef]
20. Kovacs, A.M.; Mehler, J. Cognitive gains in 7-month-old bilingual infants. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 6556–6560. [CrossRef]
21. Kaushanskaya, M.; Park, J.S.; Gangopadhyay, I.; Davidson, M.M.; Weismer, S.E. The relationship between executive functions and language abilities in children: A latent variable approach. *J. Speech Lang. Hear. Res.* **2017**, *60*, 912–923. [CrossRef]
22. Crivello, C.; Kuzyk, O.; Rodrigues, M.; Friend, M.; Zesiger, P.; Poulin-Dubois, D. The effects of bilingual growth on toddlers' executive function. *J. Exp. Child Psychol.* **2016**, *141*, 121–132. [CrossRef]
23. Iluz-Cohen, P.; Armon-Lotem, S. Language proficiency and executive control in bilingual children. *Biling. (Camb. Engl.)* **2013**, *16*, 884–899. [CrossRef]
24. Brownell, R. *Receptive One-Word Picture Vocabulary Test*; Academic Therapy Publications: Novato, CA, USA, 2001.
25. Bedore, L.M.; Peña, E.D.; García, M.; Cortez, C. Conceptual versus monolingual scoring: When does it make a difference? *Lang. Speech Hear. Serv. Sch.* **2005**, *36*, 188–200. [CrossRef]
26. Reed, M.A.; Pien, D.L.; Rothbart, M.K. Inhibitory self-control in preschool children. *Merrill Palmer Q. (Wayne State Univ. Press)* **1984**, *30*, 131–147.
27. Lagattuta, K.H.; Sayfan, L.; Monsour, M. A new measure for assessing executive function across a wide age range: Children and adults find happy-sad more difficult than day-night. *Dev. Sci.* **2011**, *14*, 481–489. [CrossRef]
28. Frye, D.; Zelazo, P.D.; Palfai, T. Theory of mind and rule-based reasoning. *Cogn. Dev.* **1995**, *10*, 483–527. [CrossRef]
29. Aro, T.; Poikkeus, A.M.; Laakso, M.L.; Tolvanen, A.; Ahonen, T. Associations between private speech, behavioral self-regulation, and cognitive abilities. *Int. J. Behav. Dev.* **2015**, *39*, 508–518. [CrossRef]
30. Bono, K.E.; Bizri, R. The role of language and private speech in preschoolers' self-regulation. *Early Child Dev. Care* **2014**, *184*, 658–670. [CrossRef]
31. Nicoladis, E.; Hui, D.; Wiebe, S.A. Language dominance and cognitive flexibility in French-English bilingual children. *Front. Psychol.* **2018**, *9*. [CrossRef]
32. Bohman, T.; Bedore, L.; Peña, E.; Mendez-Perez, A.; Gillam, R. What you hear and what you say: Language performance in Spanish-English bilinguals. *Int. Jour. Biling. Edu. Biling.* **2010**, *13*, 325–344. [CrossRef]

children

MDPI

Article

Syntactic Gender Agreement Processing on Direct-Object Clitics by Spanish-Speaking Children with Developmental Language Disorder: Evidence from ERP

Paloma Roa-Rojas [1], John Grinstead [2,*], Juan Silva-Pereyra [3,*], Thalía Fernández [4] and Mario Rodríguez-Camacho [3]

[1] Dirección de Investigación, Instituto Nacional de Geriatría, Ciudad de Mexico 10200, Mexico; paloma_roa@hotmail.com
[2] Department of Spanish & Portuguese, The Ohio State University, Columbus, OH 43210, USA
[3] Proyecto de Neurociencias, Unidad de Investigación Interdisciplinaria en Ciencias de la Salud y Educación, Facultad de Estudios Superiores Iztacala, Universidad Nacional Autónoma de México, Tlalnepantla 54090, Mexico; marcizta@gmail.com
[4] Laboratorio de Psicofisiología, Instituto de Neurobiología, Universidad Nacional Autónoma de México, Juriquilla, Querétaro 76230, Mexico; thaliafh@yahoo.com.mx
* Correspondence: grinstead.11@osu.edu (J.G.); jsilvapereyra@gmail.com (J.S.-P.); Tel.: +1-6142928856 (J.G.); +52-5513932344 (J.S.-P.)

Citation: Roa-Rojas, P.; Grinstead, J.; Silva-Pereyra, J.; Fernández, T.; Rodríguez-Camacho, M. Syntactic Gender Agreement Processing on Direct-Object Clitics by Spanish-Speaking Children with Developmental Language Disorder: Evidence from ERP. *Children* **2021**, *8*, 175. https://doi.org/10.3390/children8030175

Academic Editors: Eva Aguilar Mediavilla, Miguel Pérez Pereira, Elisabet Serrat-Sellabona and Daniel Adrover-Roig

Received: 25 January 2021
Accepted: 19 February 2021
Published: 25 February 2021

Abstract: Children with developmental language disorder (DLD) have a psycholinguistic profile evincing multiple syntactic processing impairments. Spanish-speaking children with DLD struggle with gender agreement on clitics; however, the existing evidence comes from offline, elicitation tasks. In the current study, we sought to determine whether converging evidence of this deficit can be found. In particular, we use the real-time processing technique of event-related brain potentials (ERP) with direct-object clitic pronouns in Spanish-speaking children with DLD. Our participants include 15 six-year-old Mexican Spanish-speaking children with DLD and 19 typically developing, age-matched (TD) children. Auditory sentences that matched or did not match the gender features of antecedents represented in pictures were employed as stimuli in a visual–auditory gender agreement task. Gender-agreement violations were associated with an enhanced anterior negativity between 250 and 500 ms post-target onset in the TD children group. In contrast, children with DLD showed no such effect. This absence of the left anterior negativity (LAN) effect suggests weaker lexical representation of morphosyntactic gender features and/or non-adult-like morphosyntactic gender feature checking for the DLD children. We discuss the relevance of these findings for theoretical accounts of DLD. Our findings may contribute to a better understanding of syntactic agreement processing and language disorders.

Keywords: specific language impairment; clitics; ERP; gender agreement; Spanish

1. Introduction

1.1. The Phenomena

The development of direct-object clitics in Spanish and other languages has been the subject of much study. This stems from the fact that typically developing Spanish-speaking children pass through a stage during which they neither produce them in contexts where adults obligatorily would, nor do they mark them with gender and number agreement corresponding to their antecedents, as adults obligatorily would [1–4]. In this study, we will be most interested in non-adult-like agreement in children diagnosed with developmental language disorder (DLD); however, we will briefly describe both object omission and agreement errors in typically developing children. In this way, we aim to distinguish the kind of error we are interested in from the kind that is less directly relevant to our theoretical claims.

1.1.1. Null Objects

Typically developing children across a range of languages show evidence of using direct objects of the verb in non-adult-like ways. One way in which their language is non-adult-like is their failure to produce a direct object of a transitive verb (where the adult grammar would require it), as in the following examples.

1. Spanish (Simon-Cerejido and Gutiérrez-Clellen [5] p. 336)

   Child: El niño agarró.
   the boy grabbed
   "The boy grabbed."

2. Bulgarian (Ivanov [6] p. 196)

   Child: Ritna.
   kicked
   "He kicked."

3. Greek (Marinis [7], p. 11)

   Child: Aniki Ula.
   open Ula
   "Ula shall open."

Sometimes these object omissions appear to take place in contexts that would require a pronominal in the adult language, either a clitic pronoun, or a tonic, free morpheme pronoun, according to the language. Other times, the omission occurs in a context in which pronominalization would not seem to be called for. The pronominal contexts are those in which the antecedent is made prominent in the immediately preceding discourse, while the non-pronominal contexts are those that lack such an antecedent. Each is illustrated in the following French examples from Pirvulescu and Roberge [8] from the Champaud Corpus of the CHILDES (Child Language Data Exchange System) Data Base [9].

4. Pronominal Null Object Context (Grégoire, 1; 11.22)

   Adult: La pièce elle est dedans, oui.
   the coin it is inside yes
   "The coin it is inside, yes."
   Child: Enlever.
   take out
   "Take out."

5. Non-pronominal Null Object Context (Grégoire, 2; 5.13)

   Child: Remonte tout seul (he is trying to pull up his pants).
   pull up all alone
   "Pull up myself."

This generalization is true in languages that require referential objects to be overt, such as Spanish (Here we refer to varieties of Spanish that do not allow referential null objects, though other varieties of Spanish, including at least those that are in contact with Euskera [10], Guaraní [11] and Quechua [12], do allow such null objects. See Schwenter [13] for a review), and it is true in languages in which referential null objects are a legitimate grammatical form in the adult language, including Portuguese [14] and Mandarin [15]. That is, children seem to use even more null objects in child Portuguese and child Mandarin than do adults.

In their study of direct-object omission, Castilla and Pérez-Leroux [1] showed in a sample of 103 monolingual Spanish-speaking children in Colombia that object omission occurs in 25% of elicitations by 3-year-olds, 15% in 4-year-olds and 13% in 5-year-olds (p. 14). Though other studies, using elicited production and other methodologies, find different percentages (cf. [2,3]), at least in this large monolingual sample, using a standard elicited production methodology, object omission seems to be well attested in child Spanish.

Though there is an active debate regarding the cause of non-adult-like object omission in child language, perhaps the most compelling evidence for lexical development as the principal cause comes from Pérez-Leroux et al. [4], who showed, using a structural equation model (SEM), that lexical development predicts object drop, while it does not, for example, predict determiner drop. Determiners, unlike direct objects, do not appear as a function of verb-specific lexical properties (i.e., transitive vs. intransitive), but rather are fully morphosyntactically productive. This predictive relationship of lexicon on object drop supports their claim that children have not yet learned which verbs are obligatorily transitive (e.g., devour), which are optionally transitive (e.g., eat) and which are obligatorily intransitive (e.g., laugh). This kind of information is, after all, lexically specific and would have to be learned on an item-by-item basis. In contrast, such is not the case for definite determiners, which are used as a function of regular, productive morphosyntax across the board. We will not have anything more to say about this debate in the remainder of this article, but rather limit ourselves to the simple observation that lexical development would seem to be a conceptually critical component of the development of this aspect of child language.

1.1.2. Agreement Errors

Failing to produce direct objects of verbs is not the only non-adult-like dimension of this phenomenon, however. When children attempt to produce direct-object clitic pronouns, they often produce errors of gender and number agreement with the 3rd person antecedent. In Castilla and Pérez-Leroux's [1] (p. 14) typically developing monolingual sample, gender errors were produced in 3% of elicitations with 3-year-olds, 2% with 4-year-olds and 4% with 5-year-olds. Similarly, with number errors, there were 11% number errors with 3-year-olds, 8% with 4-year-olds and 5% with 5-year-olds.

Critically, Castilla and Pérez-Leroux [1] reported that object omissions and agreement errors did not correlate in their sample, which is consistent with the idea that they arise from different causes (While some studies concern themselves with person agreement errors, also, we will limit ourselves, following Castilla et al. [1] and others, in focusing on 3rd person clitics). Among the possible errors that children can make with direct-object clitics, then, gender agreement errors between the direct-object clitic pronoun and its antecedent are what we will be concerned with in this study.

1.2. Interface Delay, Interface Deficit and Definites

What is theoretically interesting about agreement errors with object clitics in the developing language of monolingual Spanish-speaking children? Does this kind of error fit into a larger pattern of errors produced by children in general and children with DLD? To answer these questions, it could help to think about the type of construction that we are considering. Notice that direct-object pronominal clitics refer, via anaphora or deixis, to an antecedent that is prominent in the preceding discourse or that is made prominent via ostension in the physical context. In the following utterance, we see an anaphoric context. An indefinite Determiner Phrase (DP) is presented in subject position of the first sentence, and is followed in the second sentence as a direct-object clitic ("la"), which agrees in person, number and gender with the antecedent, third person, singular, feminine "una niña".

6. Anaphoric Context

    Una niña de nuestro equipo marcó un gol. La debes felicitar.
    a girl of our team scored a goal. pronoun 3rd sg. should-2nd sg. pres. congratulate-inf.
    "A girl from our team scored a goal. You should congratulate her."

In the following deictic context, a waiter brings freshly foraged mushrooms to a table to show the customers. The chef is going to make them into ravioli, but one of the customers is very hungry and says the following.

7. Deictic Context

    Los quiero comer ya.

pro-acc. want-1st sg. pres. eat-inf. already
"I want to eat them right now."

In this physical context, the pronoun "los" agrees in person, number and gender with the third person, singular, masculine noun "hongos" or mushrooms. Given the physical context, the speaker could either point or otherwise gesture towards the mushrooms or discourse pragmatics might allow the "unheralded" use of the pronoun, presupposing the interlocutor is as familiar with the unspoken referent of the pronoun as the speaker is.

In both cases, the referents of the pronouns have become prominent in either the previous discourse or by inference from the physical context, what Stalnaker [16] referred to as the Conversational Common Ground. Roberts [17,18] argued convincingly that both such pronouns should be considered definite expressions, inasmuch as they are licensed in contexts in which the speaker presupposes that interlocutors are familiar with them and that they are unique in context, in the same way that noun phrases modified by definite articles are. Following this account, other definites would include tonic pronouns (e.g., él—he, ustedes—you-pl, etc.), names (e.g., Ramón, Melissa, Ximena, etc.), definite noun phrases (e.g., el auto tuyo—your car) and null subjects (e.g., Ya llegó Juan. Ø se había ido hace una hora.—John just arrived. (He) had left an hour ago.). This semantic natural class of definites is interesting in that they, as a class, seem to develop later in child language than do constructions that do not require access to the Conversational Common Ground.

There is a sense in which the Optional Infinitive verb phenomenon (e.g., [19]) forms part of this natural class, as well, inasmuch as verbal anaphora, in the sense of Bittner [20], denotes a relationship between speech-time and event-time, following Reichenbach [21], that is taken to be familiar to both speakers and interlocutors. Thus, both nominal and verbal anaphora take time to develop in typically developing children. We contrast these constructions with more syntactically local relationships, such as nominal plural marking, noun-determiner agreement or noun–adjective agreement, which seem to develop relatively earlier in child language than do definites (e.g., [22–24]).

Given this pattern, one might be tempted to hypothesize that it is human discourse-pragmatic abilities, as instantiated, for example, in the components of Theory of Mind [25,26] that develop, and not syntax itself. More than one insightful cognitive scientist has in fact made just such a proposal (e.g., [27,28]). However, there is also evidence that multiple components of Theory of Mind are quite adult-like at an early age, even in infancy. Belief tracking—one Theory of Mind component—seems well-developed in 15-month-old pre-linguistic infants [29]. Furthermore, work on intention tracking, another subcomponent of the Theory of Mind construct, in Woodward et al. [30], shows that 12-month-old infants appear able to track the intentions of those around them. Similarly, outside the domain of Theory of Mind, but still in the domain of discourse pragmatics, Baker and Greenfield [31] gave evidence that 2-year-old children have knowledge of new versus old information in spontaneous production, before adult-like morphosyntax was being used in their English. Taken together, what is known about plausible non-linguistic cognitive abilities that could constitute important dimensions of discourse-pragmatic knowledge suggests that this knowledge, to the degree that we can separate it from language, appears well-developed in infants, who are not yet able to use definites.

On this basis, it would seem ill advised to conclude that definites are slow to develop in typically developing children as a function of discourse pragmatics itself being slow to develop. Given the fact that so much else in morphosyntax has developed to relatively adult-like levels (e.g., declarative and interrogative word orders, plural marking, preposition-object word order, etc.) at the point at which definites are still a struggle for children, researchers have proposed the hypothesis they refer to as Interface Delay [32,33] for typically developing children and Interface Deficit [34,35] for children diagnosed with DLD. These hypotheses attempt to account for the difficulty of acquisition of these constructions in terms of the inability of linguistic and discourse-pragmatic domains of cognition to interact with one another. The relationship between these domains, assuming a modular cognitive architecture (e.g., [36–38]), appears to become more robust and facile gradually over the course of development. A perhaps useful analogy for this development could be the relationship between the two language systems of bilinguals, if they study to become interpreters. Any bilingual has both systems, which can more or less interact, but only interpreters go back and forth quickly and accurately between the two. Similarly, the domain of language may be well developed on its own and the domain of discourse pragmatics may be well developed on its own in the developing mind of a child, but to become a competent, neurotypical adult user of definites, both systems must work together in an agile, robust fashion.

1.3. The Unique Checking Constraint (UCC), the Computational Complexity Hypothesis (CCH), Interface Deficit and Clitic Agreement Errors

In previous work, Grinstead et al. [39] have argued that Interface Deficit can account for at least the linguistic phenomena that appear difficult for children with DLD. Another prominent proposal, the UCC (see [19,40,41]), has produced some excellent empirical work and raised the bar for linguistic study in the domain. It reports on both expressive and receptive dimensions of the tense deficit in a child English DLD sample, studied longitudinally. The theory-specific Minimalist [42] formulation of the UCC, however, would seem to over-exclude constructions that should be problematic in the grammars of Spanish-speaking children with DLD on the UCC account, such as noun–adjective agreement (e.g., dos gatos negros—two cats-masc. pl. black-masc. pl.), which do not in fact appear to be problematic in monolingual Spanish-speaking children in Mexico [43] (Though Bedore and Leonard [44,45] found different results with Spanish-speaking children in the US context). That is, the UCC claims that multiple features occurring in the derivation of a construction (number and gender in this case; tense and agreement in the original formulation designed to account for the Extended Optional Infinitive Stage in DLD) should be sufficient to make a construction subject to prolonged and severe difficulty in the language of children with DLD. However, this does not seem to be the case for noun–adjective agreement in child Spanish DLD. The same could be said of, for example, plural marking in Spanish, at least on Picallo's [46] account of the adult morphosyntax of plurals, which, again, is not problematic for monolingual child Spanish-speakers in Mexico, diagnosed with DLD [43].

Noun plural marking in Spanish DLD would also seem to be predicted to be problematic by another prominent account of DLD, namely the Computational Complexity Hypothesis (CCH) of Jakubowicz and Nash [47], which claims that constructions will be difficult for children with DLD as a function of their relative necessity and as a function of whether they are required by syntactic versus semantic motivations. These criteria are designed to explain why the present tense in French does not seem to be difficult for children with DLD, that is, person morphology is used to mark present tense and is required in all tenses, and is therefore necessary, while past tense is more complex, in that it is required by semantics and not always necessary. One could argue that plural marking on nouns is required by semantics, but not always present, as on singular nouns, and thus should be classified as computationally complex by the CCH. On this basis, plural nouns should consequently be difficult for Spanish-speaking children with DLD, which, as we have discussed, does not appear to be the case in monolingual Spanish-speaking children in Mexico. It is for this reason that a linguistic account of DLD would seem to need to take into consideration the discourse-sensitive versus discourse-insensitive distinction, which neither the UCC nor the CCH do, but which Interface Deficit does.

Beyond these constructions that seem difficult for other frameworks to handle, there is a range of constructions that are problematic for children with DLD that directly illustrate the leading idea of Interface Deficit: anaphora is difficult. Definite noun phrases, for example, are problematic, according to work by Anderson and Souto [48] and Restrepo and Gutiérrez-Clellen [49], null subjects are problematic [50], as is tense marking in Spanish [35]. It has, of course, been demonstrated to be a highly specific and sensitive clinical marker of DLD in English (e.g., [51,52]) and has been shown to be problematic in other languages of children with DLD, including French [53], Dutch [54] and Hebrew [55]. Finally, and most relevant to our study here, direct-object clitic pronouns are difficult in monolingual child Spanish DLD, and agreement errors in particular, are prominent [3,56,57].

To address the construction that we propose to study here, how should agreement errors in clitic production be treated theoretically? The proposal in Wexler et al. [2], following the UCC, predicted that clitic errors should not occur in typically developing Spanish-speaker or in Spanish-speaking children with DLD, given that, on their account of the syntax of clitics, only 1 relevant feature is involved, which should exempt the construction from non-adult-like grammar. From the perspective of the CCH, one might predict that clitic agreement errors should occur in that they are not always present, given that not all verbs take direct-object arguments, and because they are required semantically to complete the theta grid of the transitive verb with which they occur. From the perspective of Interface Delay, direct-object clitics are definites, and consequently occur as a function of anaphora and should therefore be problematic. Furthermore, gender and number agreement, specifically, are going to depend on anaphora to a previously mentioned antecedent, making an interface between syntax and discourse pragmatics critical. Though our study of clitic agreement errors will not adjudicate between Interface Delay and the CCH, as they appear to make identical predictions, the UCC does appear to predict that they should not be problematic. We note, for completeness, that a distinct interpretation of the UCC, in which gender and number were taken to be the relevant features, and not participial agreement generally, could make the UCC fall in line predictively with the CCH and Interface Deficit.

1.4. The Surface Hypothesis and Clitic Agreement in DLD

Finally, another popular account of the DLD deficit is the Surface Hypothesis of Leonard [58], inter alia, which is explored in Aguilar-Mediavilla, Sanz-Torrent and Serra-Raventós [59] as an account of function word and weak syllable omission in bilingual Spanish-Catalan-speaking children. The core claim of the Surface Hypothesis is that children struggle to perceive words and morphemes with low phonetic salience, which results in them creating grammatical systems that do not incorporate these elements or incorporate unstable or otherwise defective versions of them. The intervening years since the Surface Hypothesis was proposed, however, have produced a number of strong counter-arguments to this core claim. First, and perhaps most compellingly, child English-speakers diagnosed with DLD produce plural /s/ with high levels (83% correct) of accuracy [60,61]. This is problematic for the Surface Hypothesis inasmuch as it predicts across-the-board difficulty with morphemes as a function of their phonological properties, which make plural /s/ in English identical to the /s/ that marks third person singular present tense. Rice and Wexler [62], convincingly, report that, in both elicited production and spontaneous production, child English-speakers diagnosed with DLD produced noun plural /s/ correctly 88% of the time, while third singular present tense /s/ was produced correctly 35% of the time, at best. This comparison is critical because “Three rocks.” and “He walks.” are segmentally and phonotactically identical. This contrast is mysterious on the Surface Hypothesis, but explicable following syntactic and semantic explanations, such as those just reviewed.

How would the Surface Hypothesis work in Spanish for clitic agreement? To begin with, plural marking is also not problematic in Spanish DLD, as we have already seen. This means that the general plausibility of a hypothesis about low phonetic salience driving the morphosyntactic deficit in Spanish is low. More to the point, the core question of our project addresses gender agreement on clitics. Grinstead et al. [43] reported that noun–adjective agreement, including gender agreement, is not significantly different between DLD and control groups on an elicited production task. While not the identical version of gender agreement, as they were not looking at clitics, these are still the same segmental “o” and “a” vowels that mark gender on both morpheme types. This finding in fact seems consistent with Aguilar-Mediavilla et al. [59], who did not report significant differences between their DLD and control groups for agreement and who also, like Castilla and Pérez-Leroux [1], found no correlation between omission and agreement errors. In sum, though it is a promising account of inflectional morphology in actual deaf and hard-of-hearing children (e.g., [63]), who show across-the-board deficits, including both tense marking and plural marking in English, the Surface Hypothesis does not appear to be an adequate account of the DLD deficit.

1.5. Clitic Agreement Errors in Spanish DLD

Previous work on the production of direct-object clitics in Spanish-speaking children with DLD, summarized in Table 1, has been done with both monolingual Spanish-speaking children [3,57,64,65] and bilingual Spanish-speaking children, whose other language was English [44,45,66–68] or Catalan [69]. In addition to monolingual versus bilingual status, these studies also varied methodologically in that some used elicited production or cloze-type tasks, while other used either Frog Story/story-retell tasks or less-structured spontaneous production tasks. Also, children varied by age between preschool and early school age. Common to all of them was the finding that children diagnosed with DLD made errors of number and gender agreement, which in some cases was at a higher rate than typically developing control groups and in other cases was not.

Table 1. Summary of Findings from Previous Studies on Clitic-Antecedent Agreement Errors for Spanish-speaking Children with DLD.

Study	Age (Years Old)	Method	% Agreement Errors	Different from Controls?
		Monolingual Spanish		
Merino [64]	5–8	Elicited Production	25%	Yes
De la Mora [57]	5.3	Elicited Production–Prompted Response	17%	Yes
Morgan et al. [65]	5.3	Cloze Test	11%	Yes [a]
Jackson-Maldonado and Maldonado [3]	7.3	Frog Story–Story Retell	Mean number of errors = 1.58	No
		Bilingual Spanish-English		
Bedore and Leonard [44,66]	3;11–5;6	Elicited Production	38% [b]	Yes
Bedore and Leonard [45]	3;11–5;6	Spontaneous Production	8% [c]	Yes
Jacobson and Schwartz [67]	4;7	Elicited Production	63% feminine 12% masculine	Yes
Jacobson [68]	Lower grades— 7.2 years old Higher grades— 10.9 years old	Elicited Production	48% [d] 46% [e] 33% [f] 32% [g]	Yes Yes Yes No
		Bilingual Spanish-Catalan		
Bosch and Serra [69]	7;6	Spontaneous Production	18.25%	No

[a] Significance was for overall clitic differences including substitution and omission, not substitution, by itself. [b] Compiled from Bedore and Leonard [66], p. 915, Table 6. [c] Compiled from Bedore and Leonard [45], p. 216, Table 8. [d] Lower age group, with preverbal clitics, for gender. [e] Lower age group, with preverbal clitics, for number. [f] Higher age group, with preverbal clitics, for gender. [g] Higher age group, with preverbal clitics, for number.

Summarizing the findings from previous work on the topic of number and gender agreement in direct-object clitics in the language of child Spanish-speakers diagnosed with DLD, it seems fair to say that agreement errors are ubiquitous and persistent. More specifically, we see that they occur at least some of the time in children in the 5-year-old range who are monolinguals, and that there are significant differences between children with DLD from typically developing age controls. These specific contrasts are critical to us, as these are the populations we will test in this study.

We now ask whether online measures, specifically, Event-Related Potentials, might also contribute to what is known, by providing either convergent or divergent evidence of a deficit with a type of definite construction, direct-object clitics.

1.6. Event-Related Potentials and Agreement in Children with DLD

Event-Related Potentials (ERPs) provide a measure of the brain electrical activity temporally associated with an event, which can be sensory, motor, or cognitive. ERPs can provide information about language processing with highly precise temporal resolution and are classified according to their polarity (i.e., positive or negative deflections in the waveform), the time of their peak occurrence in milliseconds, and their topographical distribution across the scalp [70]. ERP studies of sentences processing have analyzed anaphoric relationships using agreement features (i.e., person, number, and gender) to understand syntactic and semantic dimensions of the processing of pronouns, including agreement [71]. ERP studies of younger and older adults on Spanish sentence processing have shown that morphosyntactic agreement mismatches usually show left anterior negativities (LAN) as compared to agreement matches, which occur between 300 and 500 ms, followed by a positive wave that emerges between 500 and 1000 ms (i.e., P600) after stimulus presentation [72,73]. Functionally, a LAN is taken to reflect automatic morphosyntactic parsing [73–75], while the P600 is thought to represent processes of syntactic revision, including reanalysis or repair (e.g., [76,77]). Adult-like LAN and P600 effects have been found in normal children as young as 2 years old [78,79], but other studies have shown only a delayed P600 effect in normal toddlers when syntactic violations are presented to participants [80].

Weber-Fox et al. [81] studied subject–verb agreement processing in teenagers with DLD (e.g., Every day, the children *pretends/pretend to be super-heroes.). They observed a right anterior negativity (RAN) in both control and DLD groups, and a reduced P600 in the DLD group as compared to controls. In contrast, in another study where syntactic errors of word category in German were analyzed, whereas typically developmental (TD) children showed a bilateral early starting anterior negativity (ELAN) and a posterior P600, children with DLD showed a comparable P600 but, unlike the TD children, there was only a late, clearly left-lateralized anterior negativity [82].

Different results have been observed in the comparison between TD children with respect to DLD children using grammatical and ungrammatical questions (Who did Joe see someone?/Who did Joe see?/syntactic error) [83], because whereas TD children displayed ELAN effect for the processing of questions containing a syntactic error, children with DLD did not. Epstein et al. [84] studied whether children with DLD had atypical processing of subject and object wh-questions. Children with DLD comprehended both question types poorly, and they found a smaller sustained positivity effect in the children with DLD, compared to TD children.

According to these findings, we should expect that the electrophysiological brain response of Spanish-speaking children with DLD would be different from that of typically developing children, when exposed to gender mismatches between direct-object clitics and their antecedents. In particular, we might expect ERP waveforms to reveal a sustained LAN effect, as has been observed in previous ERP studies, with children of other ages. Further, LAN amplitudes would likely be smaller in children with DLD than in TD children. This expectation is based on evidence that the syntactic deficit, characteristic of DLD children, is correlated with a smaller LAN effect for processing of morphosyntactic errors, like gender mismatch.

1.7. Summary and Research Questions

In summary, the two most prominent non-adult-like phenomena involving direct-object clitic pronouns in child languages, including Spanish, are the fact that typically developing children and children diagnosed with DLD omit pronominal and full DP direct objects and that, when they do use them, there can be a failure of number and gender agreement between the clitic pronoun and its antecedent. Gender agreement errors, the focus of our study, do not correlate with object omission, in the one large-sample behavioral study so far conducted. This suggests that the object of our study is an independent phenomenon. Two of the prominent grammatical theories of DLD (Interface Deficit and the CCH) predict that gender agreement errors should occur in children with DLD, though for different reasons, while a third, the UCC, does not. The four existing studies of monolingual Spanish-speaking children with DLD coincide in showing with behavioral measures (elicited and spontaneous production) that this agreement is indeed a problem. No electrophysiological measures have thus far been used to compare typically developing children and children diagnosed with DLD on this dimension of grammar. The work that has been done using ERPs on subject–verb agreement in children suggests that a smaller LAN or P600 in children with DLD versus Typically developing children should be expected. If this is indeed what we find, it would serve as converging evidence of the behavioral generalization that children with DLD struggle with this construction.

In light of what we have considered, it would seem possible to distinguish children diagnosed with DLD from typically developing children using ERP components corresponding to antecedent-clitic pronoun gender agreement mismatches, which leads us to the following research questions:

- Do child Spanish-speakers show the LAN-type effect we expect for gender agreement errors?
- If so, is this effect less pronounced for children diagnosed with DLD than it is for typically developing children of the same age?

2. Materials and Methods

2.1. Participants

Thirty-five monolingual Spanish-speaking children in public schools from Mexico City participated in our experiment. The mean age for the group was 73 months (6 years, 1 month) with a standard deviation of 14.06 months. All children were given pure tone hearing tests and passed at conventional levels. Further, none of them had experienced recent episodes of otitis media. All children also scored above 85 on a test of nonverbal intelligence, the Test of Nonverbal Intelligence (TONI-2; [85]). This was done to meet the former, stricter definition of specific language impairment, which all of our TD and DLD children meet, in addition to also meeting the DLD definition of Bishop et al. [86]. The children had neither social nor physical impairments to communication, oral structural problems, or frank neurological damage, which was determined at an initial diagnostic examination and parental consultation. All children were given the Battery for Language Assessment ([Batería del Lenguaje Objetiva y Criterial Screening] BLOC-S; [87]), which has 118 items, divided into four main modules that measure distinct domains of linguistic knowledge (morphology, syntax, semantics, and pragmatics). This battery has been used in other studies for diagnosis and treatment of monolingual Spanish-speaking children with DLD (e.g., [88]).

The children who had scores above the 85th percentile on all BLOC-S subtests were assigned to the typically developing group. The children who had scores on at least two of the BLOC-S subtests that were less than 1.25 standard deviations below the mean were assigned to the DLD group. Furthermore, all children in the DLD group had received an educational diagnosis of language impairment by the psychologist of their schools. The mean age of the 16 children in the DLD group was 73.73 months (SD = 11.46), while the mean age of the 19 children in the typically developing control group was 72.45 months (SD = 16.01). The groups were not significantly different from one another in age ($t(33) = 0.857$, $p = 0.398$).

At the beginning of the first testing session, after an appropriate explanation, informed consent was obtained from all participants according to Helsinki Declaration guidelines. The protocol was approved by the Ethics Committee from the Institute of Neurobiology at the National Autonomous University of Mexico (UNAM). Parents or legal guardians also provided written consent.

2.2. Procedures

2.2.1. Stimuli

Children were presented with audio-recorded sentences of the following declarative type. Each sentence contained a transitive verb, presented in the third person singular, preterit (past perfective):

8. El papá lo filmó.

    the father him filmed

    "The father filmed him."

These sentences were presented simultaneously with an image that either matched the gender of the 3rd person singular animate direct object of the verb, or did not match (e.g., lo vs. la—him vs. her). The sentence in 8, for example, would not match the image in Figure 1, as the gender of the child in the image is stereotypically feminine.

Figure 1. Image to Elicit Gender Agreement Mismatch with the Sentence "The father filmed him."

The direct object in all cases was either niña (girl), niño (boy), gata (cat-fem.), gato (cat-masc.), perra (dog-fem.) or perro (dog-masc.). The subjects varied among abuela (grandmother), abuelo (grandfather), gato (cat-masc.), hermana (sister), hermano (brother), maestra (teacher-fem.), maestro (teacher-masc.), mamá (mother), papá (father), perro (dog-masc.), señor (gentleman) or señora (lady). These subjects and objects were composed into sentences with 50 verbs selected from two corpora of child-directed speech in Mexican Spanish, to ensure that the vocabulary used would be age appropriate. The first comes from "La producción del lenguaje de niños mexicanos" [Language production of Mexican children]; [89] and Cómo usan los niños las palabras? [How do children use words?]; [90] and the second comes from "Spanish Screener for language impairment in children" (SSLIC; [91,92]).

Thus, each verb was paired with one of the 6 direct objects and one of the 12 subjects in such a way that 50 gender-congruent sentence–image pairs were created. Then each of these same 50 sentences was presented with the corresponding image, except that the gender of the direct object was switched, yielding 50 gender-incongruent sentence–image pairs, as in Figure 1. The target stimulus in all cases was the onset of the image-congruent or image-incongruent direct-object clitic. All sentences have the same number of words before and after the clitic (see Appendix A). There was no significant difference in duration between the congruent and incongruent sentences ($p > 0.05$). Each child listened to the same sentence twice, once in the agreement and once in the disagreement condition.

To ensure the time-lock between the clitic in the audio file and the ERP recording, the wave forms of all audio files were carefully inspected and marked at the onset of the clitic. Additionally, 25 image-congruent and image-incongruent sentences were used as filler items. Congruence in the filler sentences varied as a function of the theta role of the participants in the image (e.g., La niña la saludó. "The girl greeted her." with an image in which a grandmother is greeting a girl.) All sentences were spoken by a female native speaker of Mexican Spanish and were recorded on a digital-audio system, sampled at 20 kHz with a 16-bit resolution in stereo. The speaker rehearsed all the sentences prior to recording them to ensure that they were produced fluently. The average sound pressure level ranged from 63 to 67 dB sound pressure level (SPL).

2.2.2. Event-Related Potentials Recording

The EEG was recorded from 32 tin electrodes secured in an elastic cap (Electrocap, CompuMedics, Eaton, OH, USA) at the following locations (according to the international 10–20 system): Fp1, Fpz Fp2, F7, F3, Fz, F4, F8, FT7, FC3, FCz, FC4, FT8, T7, C3, Cz, C4, T8, TP7, CP3, CPz, CP4, TP8, P7, P3, Pz, P4, and P8. The electrooculogram (EOG) was also recorded from one supraorbital electrode and an electrode placed on a child's left cheek. The recordings were referenced against the left mastoid, and brain electrical activity over the right mastoid was also recorded. Offline, all electrodes were re-referenced to average mastoids. Electrode impedances were kept below 15 kOhms. The EEG signal was amplified with Neuroscan amplifiers (CompuMedics, NeuroScan Inc., Charlotte, NC, USA). A 200-Hz sampling rate was used to digitize the EEG, with a band-pass filter set from 0.1 to 100 Hz.

The child was seated in a comfortable chair, 70 cm from the computer screen in a sound-attenuated dimly lit recording chamber. All children were instructed to relax and maintain their gaze towards the center of the screen and to avoid blinking. Subjects passively listened to the stimuli while watching each scenario. The pictures appeared in the center of the screen, thus decreasing eye movement artifacts. Each spoken sentence was presented via headphones which were placed on children's heads. Each session began with a visual presentation of all the characters used in the pictures, which concluded once the child demonstrated that they were familiar with each one. The purpose of this step was to assure that children were familiar with the names and corresponding genders of the nouns with which they were presented, to support our measurement of their reactions to the gender mis-match. Each item began with the presentation of the visual stimuli for 800 ms and after 250 ms an auditory stimulus was presented for 550 ms. The inter-stimulus interval was 1500 ms.

The entire session, including fitting the cap, lasted 45 min to 1 h. Sentences were presented in a randomized order. Each participant received the following number of sentences per condition: 50 in the gender-matched condition, 50 in the gender-mismatched condition and 50 filler items.

2.3. Data Analysis

ERPs were computed off-line from 1100 ms epochs for each subject in the gender-matched and gender-mismatched conditions. Epochs were comprised of the 100 ms preceding and the 1000 ms following the presentation of the clitic. EEG epochs with electrical activity greater than $+/-150$ μV and amplifier blocking for 50 ms or more at any electrode site were considered artifacts and the whole segment was automatically rejected. EEG epochs with artifacts due to eye movements or excessive muscle activity were eliminated by visual inspection off-line before averaging.

Subjects with fewer than 50% artifact-free trials for each condition were excluded from the average. Five DLD subjects and 7 TD subjects were excluded at this step of the process, leaving 11 children with DLD (Mean age = 5;8, SD = 1.00) and 12 controls (Mean age = 5;9, SD = 0.71), whose data could be analyzed. The mean ages for the two groups were not significantly different ($p > 0.05$). Baseline correction was performed in relation to the 100 ms pre-stimulus time mentioned above.

Statistical Analysis

Statistical analyses were performed on mean amplitude values from two time windows, which could potentially have carried relevant ERP signatures for the gender match/mismatch comparison of interest: 250–450 ms and 500–850 ms. These time windows were determined according to previous agreement studies and from inspection of both individual and grand average waveforms. Repeated-measures ANOVAs were performed separately for each time window. Separate four-way ANOVAs were performed for each time window using Group (DLD vs. TD) as a between-subject factor and Agreement (match vs. mismatch), Anterior–posterior (Frontal [F7, F3, Fz, F4, F8], Frontal-central [FT7, FC3, FCz, FC4, FT8], Central [T7, C3, Cz, C4, T8], Central-parietal [TP7, CP3, CPz, CP4, TP8], Parietal [P7, P3, Pz, P4, and P8]), and Coronal (Left, Middle-left, Middle, Middle-right and Right) as within-subject factors. The Huynh—Feldt epsilon was applied to the degrees of freedom of those analyses with more than one degree of freedom in the numerator. Tukey's honest significant difference (HSD) post hoc tests were completed after the ANOVA.

3. Results

Grand average ERP waves elicited by the critical word (match vs. mismatch conditions) for both groups are shown in Figure 2. In the TD group, morphosyntactic gender violations elicited a negative shift starting at 250 ms (i.e., anterior negativity), and ending at around 500 ms. There was no such effect observed in children with DLD. A positive wave followed this LAN-like effect in the TD children.

3.1. ERP Time Window 250–500 ms

There was an important, significant Group by Agreement by Anterior-posterior interaction ($F(4, 84) = 3.92$, $p = 0.034$, epsilon $_{H\text{-}F} = 0.430$, $\eta^2_p = 0.157$). Post Hoc analyses showed in TD children a LAN effect (i.e., greater amplitude of negativity in the mismatch than with the match condition) in frontal ($MD_{HSD} = 3.22$, $p = 0.001$) and frontal-central areas ($MD_{HSD} = 2.36$, $p = 0.014$). In contrast, children with DLD displayed no such effect in frontal ($MD_{HSD} = 0.22$, $p = 0.81$) and frontal-central areas ($MD_{HSD} = 1.11$, $p = 0.24$).

ANOVA results also revealed significant Group by Anterior-posterior by Coronal interaction ($F (16, 336) = 3.89$ $p = 0.001$, epsilon = 0.378, $\eta^2_p = 0.156$). Post hoc comparisons indicate the occurrence of, for both agreement conditions, smaller amplitudes of negativities for TD children than there were for children with DLD in the right frontal area ($MD_{HSD} = 3.54$, $p = 0.017$), the right frontal-central area ($MD_{HSD} = 2.48$, $p = 0.03$) and the right central area ($MD_{HSD} = 1.98$, $p = 0.036$), but greater amplitudes in the middle central area ($MD_{HSD} = -2.91$, $p = 0.052$).

No significant main effect of Group ($F < 1$) and Group by other important factor interactions were observed: Group by Agreement interaction ($F < 1$), Group by Agreement by Coronal interaction ($F < 1$), Group by Coronal interaction ($F (4, 84) = 1.96$, $p = 0.13$, epsilon $_{H\text{-}F} = 0.700$, $\eta^2_p = 0.085$), Group by Anterior–posterior interaction ($F (4, 84) = 2.2$, $p = 0.14$, epsilon $_{H\text{-}F} = 0.426$, $\eta^2_p = 0.094$), and Group by Agreement by Anterior–posterior by Coronal interaction ($F (16, 336) = 1.18$, $p = 0.31$, epsilon $_{H\text{-}F} = 0.534$, $\eta^2_p = 0.053$).

3.2. ERP Time Window 500–850 ms

Though there was a significant Group by Agreement by Anterior–posterior interaction ($F(4, 84) = 4.95$, $p = 0.018$, epsilon $_{H\text{-}F} = 0.409$, $\eta^2_p = 0.191$), post hoc analyses did not show a clear effect of Agreement on this positive waveform (i.e., P600) at posterior regions. TD children showed a trend towards an effect of Agreement in the frontal area ($MD_{HSD} = 1.58$, $p = 0.068$) but not in the central-parietal area ($MD_{HSD} = -0.43$, $p = 0.63$) or in the parietal area ($MD_{HSD} = -1.01$, $p = 0.28$). In contrast, children with DLD displayed no effect of Agreement in any of these areas (frontal: $MD_{HSD} = -0.51$, $p = 0.95$); central-parietal ($MD_{HSD} = 0.60$, $p = 0.52$); (parietal: $MD_{HSD} = 0.31$, $p = 0.75$).

There was a significant Group by Coronal interaction ($F (4, 84) = 3$, $p = 0.036$, epsilon $_{H\text{-}F} = 0.760$, $\eta^2_p = 0.13$) and there was also a significant Group by Anterior–posterior by Coronal interaction ($F(16, 336) = 3.1$, $p = 0.01$, epsilon $_{H\text{-}F} = 0.335$, $\eta^2_p = 0.13$). For both match and mismatch conditions, TD children displayed larger positive waves than children with DLD in the right frontal area ($MD_{HSD} = 1.98$, $p = 0.08$) and in the right frontal-central area ($MD_{HSD} = 1.4$, $p = 0.08$). This effect was observed in the opposite direction (i.e., TD children displayed a smaller positive wave than children with DLD) in the middle central area ($MD_{HSD} = -2.28$, $p = 0.047$), middle central-parietal area ($MD_{HSD} = -2.62$, $p = 0.021$), and middle and middle-right parietal area ($MD_{HSD} = -1.99$, $p = 0.077$ and $MD_{HSD} = -1.7$, $p = 0.065$).

There were no significant main effects of Group (F < 1) and Group by other important factor interactions: Group by Agreement interaction (F < 1), Group by Agreement by Coronal interaction (F < 1), Group by Anterior–posterior interaction (F (4, 84) = 1.30, $p = 0.28$, epsilon $_{H\text{-}F} = 0.434$, $\eta^2_p = 0.058$), and Group by Agreement by Anterior–posterior by Coronal interaction (F < 1).

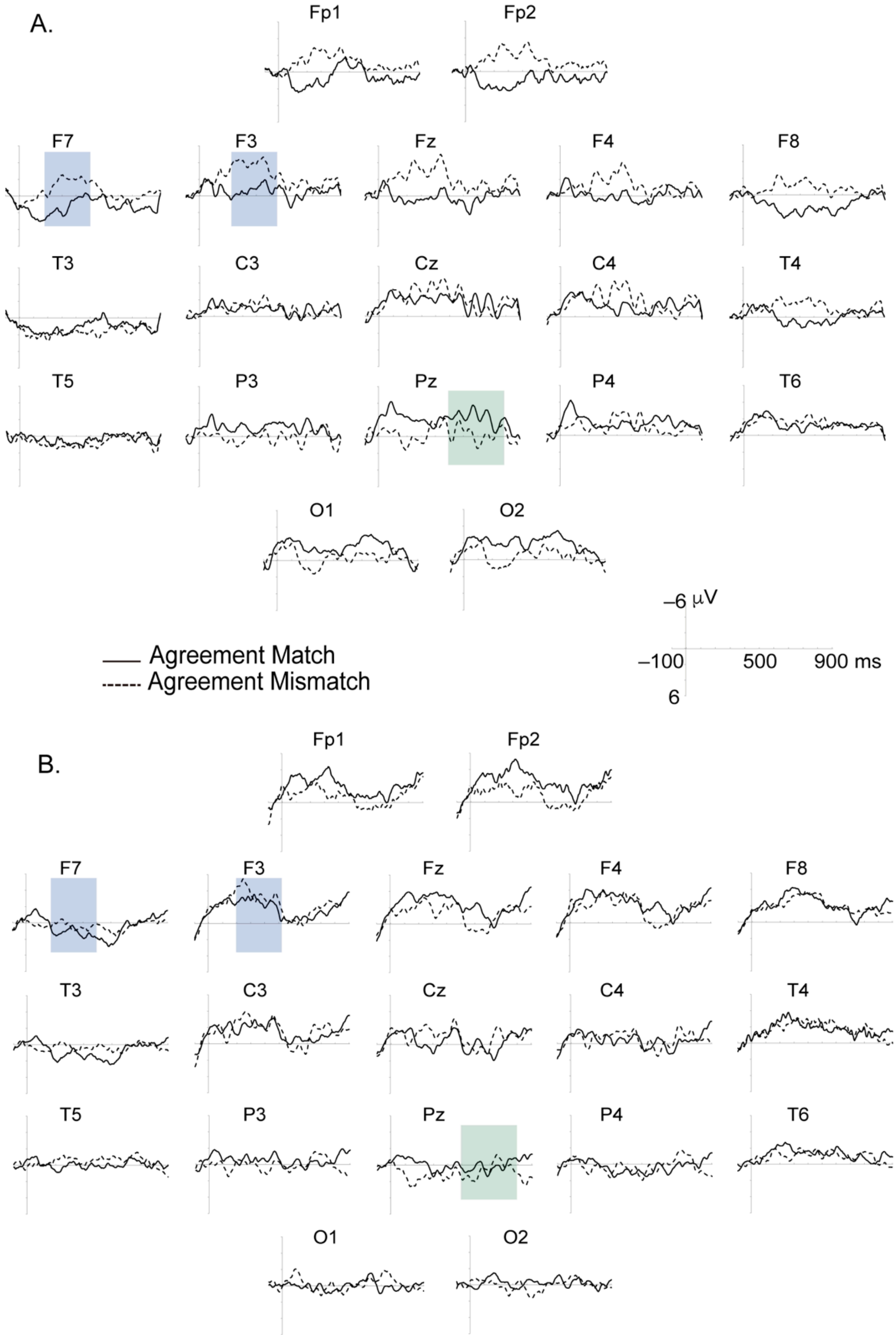


Figure 2. Grand average ERPs of (**A**), the TD children and (**B**), the children with DLD. The gender-mismatch of the clitic in the sentence/picture pair condition (dotted line) is plotted against the gender-match condition (solid line). The axis of the ordinates indicates the onset of the clitic. Negative voltage is plotted up. Purple and green shadow boxes show the time windows analyzed.

4. Discussion

To summarize, we have seen that typically developing children show ERP LAN effects that were greater when presented with agreement mismatches than with agreement matches. This was true in the frontal and frontal-central areas. This was not the case for children with DLD, however. Further, there was a near-significant P600 type effect for TD children, but not DLD children, in the frontal region. We believe that the non-significance of the P600 in the TD sample is due to our relatively smaller sample size and to the wide variance in our measurements.

Thus, with regard to our research questions, typically developing children showed significant differences, particularly with regard to the dimension of the ERP thought to correspond to morphosyntax (LAN), between agreement match and agreement mismatch. In contrast, this was not true of children with DLD. This suggests that they are not as sensitive to this type of ungrammaticality, unlike their same-aged, typically developing peers.

Our results are consistent with the large majority of behavioral studies of clitic agreement in Spanish-speaking children with DLD, both monolingual [57,64,65] and bilingual [44,45,66–68]. In this way, our study provides converging evidence of the DLD deficit in this domain of grammar, independently of the task demands inherent in the behavioral measures that have been used previously.

With respect to theoretical accounts of the DLD deficit, these results appear consistent with both the Interface Deficit as well as the Computational Complexity Hypothesis (and possibly the Unique Checking Constraint, with the modifications alluded to above).

Conceptually, there are a number of potential causes to which we could attribute our results, and the apparent consensus finding in the literature, that child Spanish-speakers with DLD fail to produce or detect adult-like gender agreement on direct-object clitic pronouns. First, it could be the case that children with DLD have weak, less well-developed lexical representations of the gender of the nouns with which adjectives agree. Such a claim would be supported by the fact that children with DLD have been consistently shown to have smaller vocabularies than typically developing age matches (e.g., [93–95]). If this were the only problem, however, we might not expect morphosyntactic problems with areas of grammar that do not depend on lexical development, such as null subjects and determiners, which are problematic for children with DLD (e.g., [48,50]), and which have across-the-board morphosyntactic properties, not typical of lexically dependent processes, as with clitics.

Another possible candidate cause is working memory. There is substantial evidence that children with DLD have fewer working memory resources to work with than do typically developing children (e.g., [96–99]). Working memory is obviously relevant to pronominal coreference where that coreference is anaphoric in nature. That is, it is one thing to have lexical representations of gender, but it is another to be able to hold these representations in memory so that morphosyntax can process them. One argument against such an explanation, however, comes from work by Noonan et al. [100], who showed that children with DLD may present with or without deficits in working memory. Specifically, they demonstrated that children diagnosed with DLD judge tense errors as ungrammatical less than typically developing controls. However, those who were also diagnosed with short-term memory deficits (diagnosed using a verbal working memory task, and two visuospatial working memory tasks) were significantly worse than those who lacked such a diagnosis on judging tense errors that occurred later in the sentence. From this finding, it seems likely that DLD and short-term memory measures are at least somewhat independent of one another. The lexicon, because tense is not lexically dependent, would not seem to play a substantial role here, either. Further, the results presented in the current report are of clitic-antecedent coreference via deixis, not anaphora. The stimuli in our experiment were visible during the auditory presentation of the sentences. In this way, reference was not established via anaphora, but rather by deictic processes, which one imagines require very little in the way of working memory capacity, though this relationship may not be very well understood.

Finally, there is the previously addressed question of whether DLD could consist of a general failure of morphosyntax. As alluded to earlier, there seems to be evidence that this is false, in the sense that child English-speakers [61] and child Spanish-speakers [43] mark plural on nouns as do typically developing controls. Similarly, child Spanish-speakers are not statistically different from age-matched in controls in noun–adjective agreement [43]. Thus, the pattern appears to be that only those dimensions of morphosyntax that are sensitive to discourse—the semantic class of definites—are problematic for children with DLD, while more local, discourse-independent syntactic relations such as plural marking on nouns and noun–adjective agreement are not problematic.

In sum, morphosyntactic constructions are difficult for children with DLD when the constructions depend critically on lexical development, as in the clitic pronouns in our current study, but also when the constructions occur independently of lexical features, as in definite determiners and null subjects. Further, children with DLD appear to have problems with morphosyntactic constructions when there is relatively little working memory load, as in our current study with deictic clitic pronouns, but also when the construction is presented with varying levels of memory load, as in tense marking in Noonan et al. [100]. Finally, it does not seem to be a general problem of morphosyntax itself, but rather the subset of morphosyntactic constructions, which fall in the semantic natural class of definites, that are most problematic.

If an Interface Deficit account is on the right track, then the problem centers on the inability of definites, including direct-object clitics, to link with antecedents in the Conversational Common Ground. On the basis of our current results, it would appear that the Conversational Common Ground is accessed not only via anaphoric processes that likely depend on the possibly domain-general, performance system of working memory, but also on the more direct, possibly domain-specific, linguistic mechanism of deixis. In this way, while there are likely multiple dimensions to the DLD deficit, one dimension characteristic of it may be a general ability to use definites, whether their presupposition of uniqueness (e.g., [101,102]) is satisfied via anaphoric processes or whether it is derived via deixis. In future work, we hope to add greater specificity and empirical substantiation to this speculative claim, with the objective always of giving providing greater understanding of the nature of this disorder.

Author Contributions: Conceptualization, P.R.-R., J.G., J.S.-P.; methodology, T.F., M.R.-C.; formal analysis, J.S.-P. and T.F.; investigation, P.R.-R.; data curation, P.R.-R., M.R.-C.; writing—original draft preparation, P.R.-R., T.F., M.R.-C., J.S.-P. and J.G.; writing—review and editing, J.G., J.S.-P. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partially supported by a grant from the Universidad Nacional Autónoma de México PAPIIT (IN207520) to TF and from the National Science Foundation (NSF-1551903) to JG.

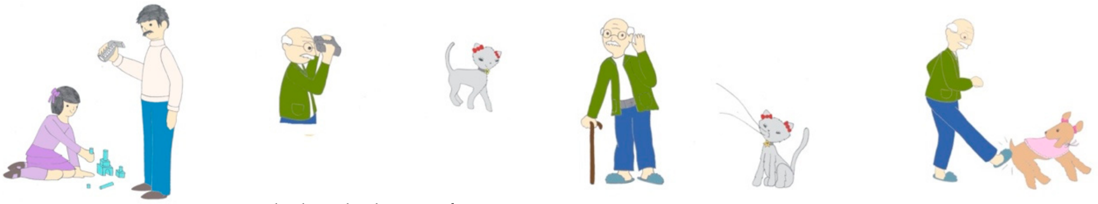
Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board (or Ethics Committee) of the Institute of Neurobiology at National Autonomous University of Mexico (UNAM) (INEU/SA/CB/146 07/01/2015).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study. Written informed consent has been obtained from parents or legal guardians of children to publish this paper.

Data Availability Statement: The data presented in this study are openly available in FigShare on at doi: 10.6084/m9.figshare.13637642 (accessed on 25 Janurary 2021).

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A



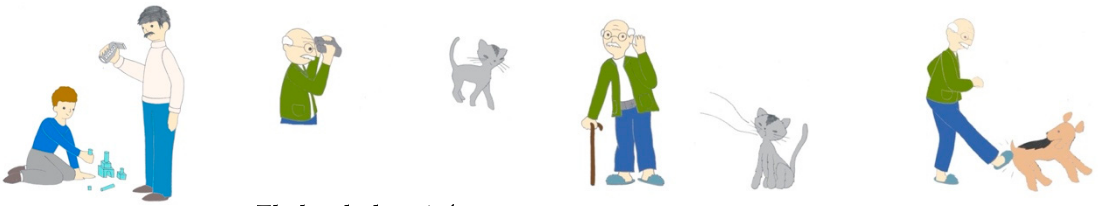
El papá lo filmó.
The father filmed him.

El abuelo lo miró.
The grandpa looked at him.

El abuelo lo oyó.
The grandpa heard him.

El abuelo lo pateó.
The grandpa kicked him.

Mismatch



El papá lo filmó.
The father filmed him.

El abuelo lo miró.
The grandpa looked at him.

El abuelo lo oyó.
The grandpa heard him.

El abuelo lo pateó.
The grandpa kicked him.

Match

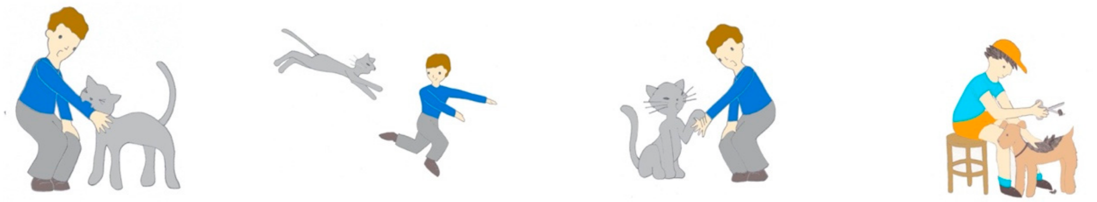
El gato lo mordió.
The cat bit him.

El gato lo persiguió.
The cat chased him.

El gato lo arañó.
The cat scratched him.

El hermano lo cortó.
The brother cut him.

Mismatch



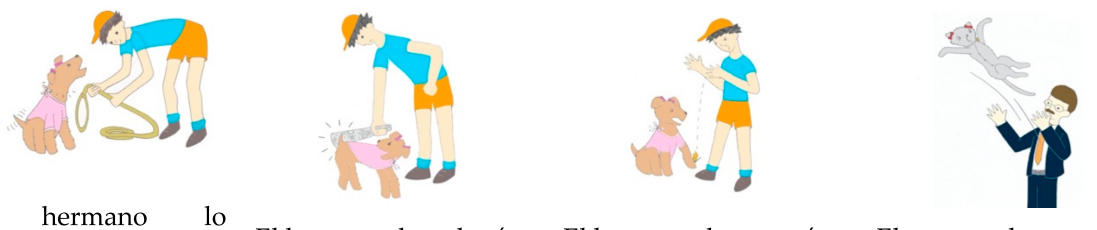
El gato lo mordió.
The cat bit him.

El gato lo persiguió.
The cat chased him.

El gato lo arañó.
The cat scratched him.

El hermano lo cortó.
The brother cut him.

Match



El hermano lo desamarró.
The brother untied him.

El hermano lo golpeó.
The brother hit him.

El hermano lo quemó.
The brother burned him.

El maestro lo aventó.
The teacher threw him.

Mismatch

Figure A1. *Cont.*

El hermano lo desamarró.
The brother untied him.

El hermano lo golpeó.
The brother hit him.

El hermano lo quemó.
The brother burned him.

El maestro lo aventó.
The teacher threw him.

Match

La hermana la pisó.
The sister stepped on her.

La maestra la apachurró.
The teacher smashed her.

La maestra la calló.
The teacher shut her up.

La maestra la despeinó.
The teacher messed with her hair.

Match

La hermana la pisó.
The sister stepped on her.

La mestra la apachurró.
The teacher smashed her.

La maestra la calló.
The teacher shut her up.

La maestra la despeinó.
The teacher messed with her hair.

Mismatch

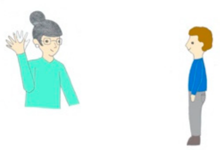
La maestra la saludó.
The teacher greeted her.

La mamá la amarró.
The mother tied her up.

La mamá la bañó.
The mother bathed her.

La mamá la cargó.
The mother carried her.

Match



La maestra la saludó.
The teacher greeted her.

La mamá la amarró.
The mother tied her up.

La mamá la bañó.
The mother bathed her.

La mamá la cargó.
The mother carried her.

Mismatch

Figure A1. *Cont.*

Figure A1. Examples of stimuli used.

References

1. Castilla, A.P.; Pérez-Leroux, A.T. Omissions and Substitutions in Spanish Object Clitics: Developmental Optionality as a Property of the Representational System. *Lang. Acquis.* **2010**, *17*, 2–25. [CrossRef]
2. Wexler, K.; Gavarro, A.; Torrens, V. Feature Checking and Object Clitic Omission in Child Catalan and Spanish. In *Romance Languages and Linguistic Theory 2002. Selected Papers from 'Going Romance', Groningen, 28–30 November 2002*; Bok-Bennema, R., Hollebrandse, B., Kampers-Manhe, B., Sleeman, P., Eds.; John Benjamins: Amsterdam, The Netherlands, 2004; pp. 253–268.
3. Jackson-Maldonado, D.; Maldonado, R. Grammaticality differences between Spanish-speaking children with specific language impairment and their typically developing peers. *Int. J. Lang. Commun. Disord.* **2017**, *52*, 750–765. [CrossRef]
4. Pérez-Leroux, A.T.; Castilla, A.P.; Brunner, J. General and specific effects of lexicon in grammar: Determiner and object pronoun omissions in child Spanish. *J. Speech Lang. Hear. Res.* **2012**, *55*, 313–327. [CrossRef]
5. Simon-Cereijido, G.; Gutiérrez-Clellen, V. Spontaneous language markers of Spanish language impairment. *Appl. Psycholinguist.* **2007**, *28*, 317–339. [CrossRef]
6. Ivanov, I. L1 Acquisition of Bulgarian Object Clitics: Unique Checking Constraint or Failure to Mark Referentiality. In *BUCLD 32 Proceedings*; Chan, H., Jacob, H., Kapia, E., Eds.; Cascadilla Press: Somerville, MA, USA, 2008; pp. 189–200.
7. Marinis, T. The Acquisition of Clitic Objects in Modern Greek: Single Clitics, Clitic Doubling, Clitic Left Dislocation. *ZAS Pap. Linguist.* **2000**, *15*, 1–23.
8. Pirvulescu, M.; Roberge, Y. Licit and illicit null objects in L1 French. In *Theoretical and Experimental Approaches to Romance Linguistics*; Gess, R., Rubin, E., Eds.; John Benjamins: Amsterdam, The Nethererlands, 2005; pp. 197–212.
9. MacWhinney, B. *The CHILDES Project*; Lawrence Erlbaum: Mahwah, NJ, USA, 2000.
10. Landa, M.A. Conditions on Null Objects in Basque Spanish and Their Relation to Leísmo and Clitic Doubling. Ph.D. Thesis, University of Southern California, Los Angeles, CA, USA, 1995.
11. Choi, J.K. [-Person] direct object drop: The genetic cause of a syntactic feature in Paraguayan Spanish. *Hispania* **2000**, *83*, 531–543. [CrossRef]
12. Yépez, M.V. Direct Object Clitics in Quiteño Spanish. Ph.D. Thesis, Cornell, Ithaca, NY, USA, 1986.
13. Schwenter, S.A. Null Objects across South America. In *Selected Proceedings of the 8th Hispanic Linguistics Symposium*; Face, T.L., Klee, C., Eds.; Cascadilla Proceedings Project: Somerville, MA, USA, 2006; pp. 23–36.
14. Costa, J.; Lobo, M. Clitic omission, null objects or both in the acquisition of European Portuguese? In *Romance Languages and Linguistic Theory*; Baauw, S., Drijkoningen, F., Pinto, M., Eds.; John Benjamins: Amsterdam, The Netherlands, 2007; pp. 59–71.
15. Wang, Q.; Lillo-Martin, D.; Best, C.; Levitt, A. Null subject vs. null object: Some evidence from the acquisition of Chinese and English. *Lang. Acquis.* **1992**, *2*, 221–254. [CrossRef]
16. Stalnaker, R. Pragmatic Presuppositions. In *Semantics and Philosophy*; Munitz, M., Unger, D., Eds.; New York University Press: New York, NY, USA, 1974; pp. 197–219.
17. Roberts, C. Demonstratives as definites. In *Information Sharing: Reference and Presupposition in Language Generation and Interpretation*; van Deemter, K., Kibble, R., Eds.; CSLI Press—Stanford: Palo Alto, CA, USA, 2002; pp. 89–196.
18. Roberts, C. Uniqueness in Definite Noun Phrases. *Linguist. Philos.* **2003**, *26*, 287–350. [CrossRef]
19. Wexler, K. Very Early Parameter Setting and the Unique Checking Constraint: A New Explanation of the Optional Infinitive Stage. *Lingua* **1998**, *106*, 23–79. [CrossRef]

20. Bittner, M. Time and Modality without Tense or Modals. In *Tense Across Languages*; Musan, R., Rathert, M., Eds.; Niemeyer: Tübingen, Germany, 2011; pp. 147–188.
21. Reichenbach, H. *Elements of Symbolic Logic*; MacMillan: New York, NY, USA, 1947.
22. Brown, R.W. *A First Language: The Early Stages*; Harvard University Press: Cambridge, MA, USA, 1973.
23. de Villiers, J.G.; de Villiers, P.A. A crosssectional study of the acquisition of grammatical morphemes. *J. Psycholinguist. Res.* **1973**, *2*, 267–278. [CrossRef] [PubMed]
24. Hernández-Pina, R. *Teorías Psicosociolingüísticas y su Aplicación a la Adquisición del Español Como Lengua Materna*; Siglo XXI: Madrid, España, 1984.
25. Wimmer, H.; Perner, J. Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition* **1983**, *13*, 103–128. [CrossRef]
26. Baron-Cohen, S. *Mindblindness: An. Essay on Autism and Theory of Mind*; MIT Press: Cambridge, MA, USA, 1995.
27. Avrutin, S. Psycholinguistic Investigations in the Theory of Reference. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 1994.
28. Schaeffer, J. *The Acquisition of Direct Object Scrambling and Clitic Placement*; John Benjamins: Amsterdam, The Netherlands, 2000.
29. Onishi, K.; Baillargeon, R. Do 15-month-old infants understand false beliefs? *Science* **2005**, *308*, 255–258. [CrossRef] [PubMed]
30. Woodward, A.L.; Sommerville, J.A.; Gerson, S.; Henderson, A.M.E.; Buresh, J. The emergence of intention attribution in infancy. In *The Psychology of Learning and Motivation: Vol 51*; Ross, B.H., Ed.; Elsevier Academic Press: San Diego, CA, USA, 2009; pp. 187–222.
31. Baker, N.D.; Greenfield, P.M. The Development of New and Old Information in Young Children's Early Language. *Lang. Sci.* **1988**, *10*, 3–34. [CrossRef]
32. Grinstead, J. Subjects and Interface Delay in Child Spanish and Catalan. *Language* **2004**, *80*, 40–72. [CrossRef]
33. Grinstead, J.; Spinner, P. The clausal left periphery in child Spanish and German. *Probus* **2009**, *21*, 51–82. [CrossRef]
34. Grinstead, J.; De la Mora, J.; Pratt, A.; Flores, B. Temporal Interface Delay and Root Nonfinite Verbs in Spanish-Speaking Children with Specific Language Impairment: Evidence from the Grammaticality Choice Task. In *Hispanic Child Languages: Typical and Impaired Development*; Grinstead, J., Ed.; John Benjamins: Amsterdam, The Netherlands, 2009; pp. 239–263.
35. Grinstead, J.; Baron, A.; Vega-Mendoza, M.; De la Mora, J.; Cantú-Sánchez, M.; Flores, B. Tense marking and spontaneous speech measures in Spanish specific language impairment: A discriminant function analysis. *J. Speech Lang. Hear. Res.* **2013**, *56*, 352–363. [CrossRef]
36. Chomsky, N. Questions of Form and Interpretation. *Linguist. Anal.* **1975**, *1*, 75–109.
37. Jackendoff, R. *Consciouness and the Computational Mind*; MIT Press: Cambridge, MA, USA, 1987.
38. Jackendoff, R. *The Architecture of the Language Faculty*; MIT Press: Cambridge, MA, USA, 1997.
39. Grinstead, J.; Lintz, P.; Vega-Mendoza, M.; De la Mora, J.; Cantú-Sánchez, M.; Flores-Avalos, B. Evidence of optional infinitive verbs in the spontaneous speech of Spanish-speaking children with SLI. *Lingua* **2014**, *140*, 52–66. [CrossRef]
40. Rice, M.L.; Wexler, K.; Hershberger, S. Tense over Time: The Longitudinal Course of Tense Acquisition in Children with Specific Language Impairment. *J. Speech Lang. Hear. Res.* **1998**, *41*, 1412–1431. [CrossRef] [PubMed]
41. Rice, M.L.; Wexler, K.; Redmond, S.M. Grammaticality Judgments of an Extended Optional Infinitive Grammar: Evidence from English-Speaking Children with Specific Language Impairment. *J. Speech Lang. Hear. Res.* **1999**, *42*, 943–961. [CrossRef]
42. Chomsky, N. *The Minimalist Program*; MIT Press: Cambridge, MA, USA, 1995.
43. Grinstead, J.; Cantu-Sanchez, M.; Flores-Avalos, B. Canonical and Epenthetic Plural Marking in Spanish-Speaking Children with Specific Language Impairment. *Lang. Acquis.* **2008**, *15*, 329–349. [CrossRef]
44. Bedore, L.; Leonard, L. "Grammatical morphology deficits in Spanish-speaking children with SLI": Erratum. *J. Speech Lang. Hear. Res.* **2002**, *45*, 1015. [CrossRef]
45. Bedore, L.; Leonard, L. Verb inflections and noun phrase morphology in the spontaneous speech of Spanish-speaking children with specific language impairment. *Appl. Psycholinguist.* **2005**, *26*, 195–225. [CrossRef]
46. Picallo, M.C. Gender and Number in Romance. *Lingue Ling.* **2008**, *7*, 47–66. [CrossRef]
47. Jakubowicz, C.; Nash, L. Functional Categories and Syntactic Operations in (Ab)normal Language Acquisition. *Brain Lang.* **2001**, *77*, 321–339. [CrossRef]
48. Anderson, R.T.; Souto, S.M. The use of articles by monolingual Puerto Rican Spanish-speaking children with specific language impairment. *Appl. Psycholinguist.* **2005**, *26*, 621–647. [CrossRef]
49. Restrepo, M.A.; Gutierrez-Clellen, V.F. Article Use in Spanish-Speaking Children with Specific Language Impairment. *J. Child Lang.* **2001**, *28*, 433–452. [CrossRef]
50. Grinstead, J.; Lintz, P.; Pratt, A.; Vega-Mendoza, M.; De la Mora, J.; Cantú-Sánchez, M.; Flores-Avalos, B. Null Subject Licensing in Monolingual Spanish SLI: A Discriminant Function Analysis. *Languages* **2018**, *3*, 17. [CrossRef]
51. Leonard, L.B.; Eyer, J.A.; Bedore, L.M.; Grela, B.G. Three Accounts of the Grammatical Morpheme Difficulties of English-Speaking Children with Specific Language Impairment. *J. Speech Lang. Hear. Res.* **1997**, *40*, 741–753. [CrossRef] [PubMed]
52. Rice, M.; Wexler, K. *Test. of Early Grammatical Impairment-the Examiner's Manual*; The Psychological Corporation: San Antonio, TX, USA, 2001.
53. Jakubowicz, C.; Roulet, L. Do French-Speaking Children with SLI Present a Selective Deficit on Tense? In *BUCLD Proceedings 28*; Brugos, A., Micciulla, L., Smith, C.E., Eds.; Cascadilla Press: Somerville, MA, USA, 2004; pp. 256–266.

54. Wexler, K.; Schaeffer, J.; Bol, G. 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* **2004**, *7*, 148–198. [CrossRef]
55. Leonard, L.B.; Dromi, E.; Adam, G.; Zadunaisky-Ehrlich, S. Tense and Finiteness in the Speech of Children with Specific Language Impairment Acquiring Hebrew. *Int. J. Lang. Commun. Disord.* **2000**, *35*, 319–335. [CrossRef] [PubMed]
56. De la Mora, J.; Paradis, J.; Grinstead, J.; Flores, B.; Cantú, M. Object Clitics in Spanish Speaking Children with and without SLI. In *Symposium for Research on Child Language Disorders*; University of Wisconsin: Madison, AL, USA, 2003.
57. De la Mora, J. Direct Object Clitics and Determiners in Spanish Specific Language Impairment. Master's Thesis, University of Alberta, Edmonton, AB, Canada, 2004.
58. Leonard, L. Language learnability and specific language impairment in children. *Appl. Psycholinguist.* **1989**, *10*, 179–202. [CrossRef]
59. Aguilar-Mediavilla, E.; Sanz-Torrent, M.; Serra-Raventós, M. Influence of phonology on morpho-syntax in Romance languages in children with Specific Language Impairment (SLI). *Int. J. Lang. Commun. Disord.* **2007**, *42*, 325–347. [CrossRef]
60. Oetting, J.B. Language-Impaired and Normally Developing Children's Acquisition of English Plural. Ph.D. Thesis, University of Kansas, Kansas, KS, USA, 1993.
61. Rice, M.L.; Oetting, J.B. Morphological Deficits of Children with SLI: Evaluation of Number Marking and Agreement. *J. Speech Lang. Hear. Res.* **1993**, *36*, 1249–1257. [CrossRef] [PubMed]
62. Rice, M.L.; Wexler, K. Toward Tense as a Clinical Marker of Specific Language Impairment in English-Speaking Children. *J. Speech Hear. Res.* **1996**, *39*, 1239–1257. [CrossRef] [PubMed]
63. Svirsky, M.A.; Stallings, L.M.; Lento, C.L.; Ying, E.; Leonard, L.B. Grammatical Morphologic Development in Pediatric Cochlear Implant Users May Be Affected by the Perceptual Prominence of the Relevant Markers. *Ann. Otol. Rhinol. Laryngol.* **2002**, *189*, 109–112. [CrossRef]
64. Merino, B. Language development in normal and language handicapped Spanish-speaking children. *Hisp. J. Behav. Sci.* **1983**, *5*, 379–400. [CrossRef]
65. Morgan, G.; Restrepo, M.A.; Auza, A. Varaibility in the grammatical profiles of Spanish-speaking children with specific language impairment. In *Hispanic Child Languages: Typical and Impaired Development*; Grinstead, J., Ed.; John Benjamins: Amsterdam, The Netherlands, 2009; pp. 283–302.
66. Bedore, L.; Leonard, L. Grammatical Morphology Deficits in Spanish-speaking children with specific language impairment. *J. Speech Lang. Hear. Res.* **2001**, *44*, 905–924. [CrossRef]
67. Jacobson, P.; Schwartz, R. Morphology in incipient bilingual Spanish-speaking preschool children with specific language impairment. *Appl. Psycholinguist.* **2002**, *23*, 23–41. [CrossRef]
68. Jacobson, P. The effects of language impairment on the use of direct object pronouns and verb inflections in heritage Spanish speakers: A look at attrition, incomplete acquisition and maintenance. *Biling-Lang. Cogn.* **2012**, *15*, 22–38. [CrossRef]
69. Bosch, L.; Serra, M. Grammatical Morphology Deficits of Spanish-Speaking Children with Specific Language Impairment. *Amst. Ser. Child Lang. Dev.* **1997**, *6*, 33–45.
70. Luck, S. *An introduction to the Event-Related Potential Technique*; MIT Press: Cambridge, MA, USA, 2014.
71. Silva-Pereyra, J.; Gutiérrez-Sigut, E.; Carreiras, M. An ERP study of coreference in Spanish: Semantic and grammatical gender cues. *Psychophysiology* **2012**, *49*, 1401–1411. [CrossRef] [PubMed]
72. Barber, H.; Carreiras, M. Grammatical Gender and Number Agreement in Spanish: An ERP Comparison. *J. Cogn. Neurosci.* **2005**, *17*, 137–153. [CrossRef] [PubMed]
73. Molinaro, N.; Barber, H.; Carreiras, M. Grammatical agreement processing in reading: ERP findings and future directions. *Cortex* **2001**, *47*, 908–930. [CrossRef]
74. Bornkessdel-Schlesewsky, I.; Schlesewsky, M. The extended argument dependency model: A neurocognitive approach to sentence comprehension across language. *Psychol. Rev.* **2006**, *113*, 787–821. [CrossRef]
75. Friederici, A.; Weissenborn, J. Mapping sentence form onto meaning: The syntax–semantic interface. *Brain Res.* **2007**, *1146*, 50–58. [CrossRef] [PubMed]
76. Friederici, A.; Hahne, A.; Saddy, D. Distinct neurophysiological patterns reflecting aspects of syntactic complexity and syntactic repair. *J. Psycholinguist. Res.* **2002**, *31*, 45–63. [CrossRef]
77. Kaan, E.; Swaab, T. Electrophysiological evidence for serial sentence processing: A comparison between non-preferred and ungrammatical continuations. *Brain Res. Cogn. Brain Res.* **2003**, *17*, 621–635. [CrossRef]
78. Hahne, A.; Eckstein, K.; Friederici, A. Brain signatures of syntactic and semantic processes during children's language development. *J. Cogn. Neurosci.* **2004**, *16*, 1302–1318. [CrossRef] [PubMed]
79. Oberecker, R.; Friedrich, M.; Friederici, A. Neural correlates of syntactic processing in two-year-olds. *J. Cogn. Neurosci.* **2005**, *17*, 1667–1678. [CrossRef] [PubMed]
80. Silva-Pereyra, J.; Rivera-Gaxiola, M.; Kuhl, P.K. An event-related brain potential study of sentence comprehension in preschoolers: Semantic and morphosyntactic processing. *Brain Res. Cogn. Brain Res.* **2005**, *23*, 247–258. [CrossRef] [PubMed]
81. Weber-Fox, C.; Leonard, L.; Wray, A.; Tomblin, J.B. Electrophysiological correlates of rapid auditory and linguistic processing in adolescents with specific language impairment. *Brain Lang.* **2010**, *115*, 162–181. [CrossRef]
82. Sabisch, B.; Hahne, A.; Glass, E.; von Suchodoletz, W.; Friederici, A. Children with specific language impairment: The role of prosodic processes in explaining difficulties in processing syntactic information. *Brain Res.* **2009**, *1261*, 37–44. [CrossRef]

83. Fonteneau, E.; van der Lely, H.K. Electrical brain responses in language-impaired children reveal grammar-specific deficits. *PLoS ONE* **2008**, *3*, e1832. [CrossRef]
84. Epstein, B.; Hestvik, A.; Shafer, V.; Schwartz, R. ERPs reveal atypical processing of subject versus object Wh-questions in children with specific language impairment. *Int. J. Lang. Commun. Disord.* **2013**, *48*, 351–365. [CrossRef]
85. Brown, L.; Sherbenou, R.J.; Johnsen, S.K. *Test. de Inteligencia No-Verbal TONI-2 [Test. of Nonverbal Intelligence 2]*; TEA Press: Madrid, España, 1995.
86. Bishop, D.V.; Snowling, M.J.; Thompson, P.A.; Greenhalgh, T.; CATALISE Consortium. Phase 2 of CATALISE: A multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *J. Child Psychol. Psychiatry* **2017**, *58*, 1068–1080. [CrossRef] [PubMed]
87. Puyuelo, M.; Renom, J.; Solanas, A.; Wiig, E. *Evaluación de Lenguaje BLOC Screening [Language Assessment BLOC]. Manual de Usuario*; Editorial Masson: Barcelona, España, 2002.
88. Puyuelo, M.; Salavera, C.; Wiig, E. Aportaciones a la evaluación del lenguaje mediante BLOC-C y BLOC-SR [Usefulness of the BLOC-C and BLOC-SR tests in language assessment]. *Rev. Logop. Foniatría Audiol.* **2013**, *33*, 83–96. [CrossRef]
89. Alva, E.; Hernández-Padilla, E. *La Producción del Lenguaje de Niños Mexicanos. Un Estudio Transversal de Niños de 5 a 12 años [Language Production of Mexican Children. A Cross-Sectional Study of Children Aged 5 to 12 Years]*; Editorial UNAM: Mexico City, Mexico, 2001.
90. Alva, E.; Arias-Trejo, N.; Mazón, N.; Mejía, I.; Hernández-Padilla, E.; Carrión, R.; Pérez, B. *¿Cómo Usan los Niños las Palabras? [How do Children Use Words?]*; Editorial UNAM: CDMX, México, 2001.
91. Restrepo, M.A.; Gorin, J.; Gray, S. Screening Spanish-speaking children for language impairment: Results from a scale development grant. In *Inaugural Bilingual Research Conference*; University of Houston: Houston, TX, USA, 2012.
92. Kapantzoglou, M.; Thompson, M.; Gray, S.; Restrepo, M.A. Assessing measurement invariance for Spanish sentence repetition and morphology elicitation tasks. *J. Speech Lang. Hear. Res.* **2016**, *59*, 254–266. [CrossRef]
93. Gray, S. The relationship between phonological memory, receptive vocabulary, and fast mapping in young children with specific language impairment. *J. Speech Lang. Hear. Res.* **2006**, *49*, 955–969. [CrossRef]
94. Nash, M.; Donaldson, M. Word Learning in Children with Vocabulary Deficits. *J. Speech Lang. Hear. Res.* **2005**, *48*, 439–458. [CrossRef]
95. Spaulding, T.J.; Hosmer, S.; Schechtman, C. Investigating the interchangeability and diagnostic utility of the PPVT-III and PPVT-IV for children with and without SLI. *Int. J. Speech Lang. Pathol.* **2013**, *15*, 453–462. [CrossRef] [PubMed]
96. Montgomery, J.W. Verbal working memory in sentence comprehension in children with specific language impairment. *J. Speech Lang. Hear. Res.* **2000**, *43*, 293–308. [CrossRef]
97. Conti-Ramsden, G. Processing and linguistic markers in young children with specific language impairment. *J. Speech Lang. Hear. Res.* **2003**, *46*, 1029–1037. [CrossRef]
98. Gathercole, S. Non-word repetition and word learning: The nature of the relationship. *Appl. Psycholinguist.* **2006**, *27*, 513–543. [CrossRef]
99. Vugs, B.; Knoors, H.; Cuperus, J.; Hendriks, M.; Verhoeven, L. Interactions between working memory and language in young children with specific language impairment (SLI). *Child Neuropsychol.* **2016**, *22*, 955–978. [CrossRef] [PubMed]
100. Noonan, N.; Redmond, S.M.; Archibald, L. Contributions of Children's Linguistic and Working Memory Proficiencies to Their Judgments of Grammaticality. *J. Speech Lang. Hear. Res.* **2014**, *57*, 979–989. [CrossRef] [PubMed]
101. Russell, B. On Denoting. *Mind* **1905**, *66*, 479–493. [CrossRef]
102. Heim, I. Articles and Definiteness. In *Semantics: An International Handbook of Contemporary Research*; Stechow, A.V., Wunderlich, D., Eds.; De Gruyter: Berlin, Germany, 1991.

children

MDPI

Article

Lexical and Grammatical Errors in Developmentally Language Disordered and Typically Developed Children: The Impact of Age and Discourse Genre

Aleksandr N. Kornev [1,*] and Ingrida Balčiūnienė [1,2]

[1] Department of Logopathology, Saint-Petersburg State Pediatric Medical University, 194100 Saint-Petersburg, Russia; ingrida.balciuniene@vdu.lt
[2] Department of Lithuanian Studies, Vytautas Magnus University, 44248 Kaunas, Lithuania
* Correspondence: k1949@yandex.ru

Citation: Kornev, A.N.; Balčiūnienė, I. Lexical and Grammatical Errors in Developmentally Language Disordered and Typically Developed Children: The Impact of Age and Discourse Genre. *Children* **2021**, *8*, 1114. https://doi.org/10.3390/children8121114

Academic Editors: Eva Aguilar Mediavilla, Miguel Pérez Pereira, Elisabet Serrat-Sellabona and Daniel Adrover-Roig

Received: 17 August 2021
Accepted: 19 November 2021
Published: 2 December 2021

Abstract: Persistent lexical and grammatical errors in children's speech are usually recognized as the main evidence of language delay or language disorder. These errors are usually treated as a sign of a deficit in language competence. On the other hand, some studies have revealed the same kinds of grammatical errors in children with developmental language disorder (DLD) and in typically developed (TD) children. Quite often, DLD children use grammatical markers properly, but sometimes they do this erroneously. It has been suggested that the main area of the limitations in DLD children is language performance but not language competence. From the perspective of the resource deficit model, the error rate in DLD children should be influenced by the cognitive demands of utterance and text production. We presume that different genres of discourse demand a different number of cognitive resources and, thus, should differently impact the error rate in children's speech production. To test our hypothesis, we carried out an error analysis of two corpora of child discourse. The first corpus contained longitudinal data of discourse (personal narratives, fictional stories, chats, and discussions) collected from 12 children at four age points (4 years 3 months., 4 years 8 months., 5 years 3 months., and 5 years 9 months. years). Another corpus contained discourse texts (fictional stories and discussions) collected in the framework of a cross-sectional study from 6-year-old TD and DLD children; the DLD children had language expression but not comprehension difficulties. A comparative analysis between different discourse genres evidenced that the genre of discourse and age of assessment impacted the error distribution in the DLD and TD children. Such variables as the lexical and morphological error rates were impacted the most significantly. The results of the two studies confirmed our hypothesis regarding the probabilistic nature of lexical and grammatical errors in both DLD and TD children and the relationship between a cognitive loading of the genre and the error rate.

Keywords: developmental language disorder; language errors; grammatical errors; lexical errors; derivational errors; preschool age

1. Background

Language development is considered an essential part of mental development. Language is one of the main components of verbal reasoning as well as communicative and social interaction skills. Following the elegant expressions of [1], from the first months of life, a child tries to 'become a native speaker' and 'to be a proficient speaker'. 'To become a native speaker' means to learn the mother tongue, i.e., to acquire numerous language units and multiple rules and combine them for producing utterances following appropriate standards of phonology, vocabulary, and grammar. During this time, a child accumulates and elaborates a rich complex of declarative and procedural knowledge. 'To become a proficient speaker' means to master many complex and flexible skills and strategies for producing cohesive and coherent discourse texts according to multiple social and cultural traditions a child faces within his/her community. Despite the differences among numeral theories of speech/language development [2–7], scientists generally agree that during the first several years of life, a child makes many errors in the phonological, lexical, and grammatical domains. Along the second and third years of life, some children tend to use many frozen phrases and, thus, make fewer errors (i.e., the so-called expressive style of language development; see [8,9]). On the contrary, other children prefer to construct their utterances according to language rules they have deduced and make many more errors (i.e., so-called referential style of language development) [8,9]. According to a comprehensive longitudinal study based on the natural observation of native Russian language acquisition [10], until the 7th year of life, typically developed (TD) children sometimes make grammatical or lexical errors. For example, by the age of 7 years, Russian speaking children usually have acquired noun morphology and select proper case forms (nominative, genitive, dative, accusative, instrumental or prepositional case) while speaking; however, Russian nouns fall under three declensions, and these are further divided into declension paradigms and their sub-types [11]. The extremely complex system of noun declension provokes children to sometimes select incorrect declension forms (although the case form is usually selected correctly) even at the 6th–7th year of life [9]. Besides incorrect declension forms, occasional omissions of functional words and production of neologisms may be also observed.

On the other hand, some errors made by children are obviously incidental and might be considered as slips of the tongue (SoT)—'unintended, nonhabitual deviation(s) from a speech plan' [12] (p. 284). The phenomena of speech errors and SoT are usually explained differently. Speech error is usually recognized as a consequence of the temporary incomplete language competence [13] or a result is the simplification processes in the programming of utterance inherent to children during language acquisition [14]. Other explanations highlight the developmentally caused cognitive resource immaturity, which, in young children, underlies the inability to sustain error-free speech production [15,16]. There is considerable evidence that problem-solving activity and implicit learning are essential parts of language development [17–19]. When a child is acquiring a native language, he/she tries to listen their caregivers and other communicative partners carefully to select the more informative grammatical, phonological, and semantical segments in the speech stream and try to attribute them to some extralinguistic events, objects, or features. Then, a child usually tries to repeat some piece of utterance and to control the concordance between the target word/phrase and his/her own production. Additionally, a child detects scaffolding from a child-directed speech containing positive or negative evidence [20,21]. This complex multimodal activity should be served by several cognitive abilities and skills, such as executive functions, selective auditory attention, working memory, self-monitoring, and strategy shifting [14,22]. Most of these cognitive resources are not completely mature until the 8th–11th year of life [23–25]. The younger the child is, the more pronounced these limitations are. Later on, the maturity of the cognitive resources grows gradually along the preschool and school years [14,26,27]. Though language competence and discourse skills develop rapidly, they remain incomplete until adolescence [28]. Thus, in other words, children execute a complex work of learning a native language in the conditions of quite limited cognitive resources [16]. In the majority of children populations, regular speech errors disappear before the 6th–7th year of life (although, in different languages, these milestones may differ) [29].

However, let us look at the other direction of speech and language development, i.e., discourse development. It is well known that discourse production is loaded by some additional cognitive activities that depend on discourse genres.

Following [14], the development of discourse skills and cohesion devices '[...] involves two different but related problems: (1) the reorganization of stored linguistic representations so that they can form a system, and (2) the creation of a control process which constrains connected discourse as it is produced in real time' [14] (p. 62). The control process managed by the self-monitoring subsystem plays a very important role in discourse development. As verbal communication strategies and linguistic devices, the discourse genres are determined by particular cultural traditions, and a child should learn them properly. Hence, employing problem-solving strategy [19,30,31], he/she acquires some procedural knowledge and elaborates a set of unique skills. For example, a child has to acquire procedural knowledge necessary for planning and producing narrative, descriptive and other genres' passages while building coherent discourse, for using proper discourse markers, etc. A set of verbal communication skills includes, for example, skills of self-monitoring during narration or conversation, turn-taking skills necessary in a dialogue, pragmatic skills (e.g., manipulation with different registers of communication), and many other verbal behavior skills [32,33]. During this period, the abilities of speech behavior self-monitoring are essential to reach the social-communicative standards. The development of self-monitoring proficiency depends on age [14,34–36] and the maturity of cognitive resources [34]. In a few psycholinguistic studies of cognitive demands of different discourse genres, some evidence was obtained that, for example, a conversation is less demanding, while an expository discourse is more cognitively demanding than a narrative one [37]. For example, in narrative production, a child has to arrange the story structure and to verbalize intentions and goals, emotions, and other mental states held by characters [38–40]. When producing a spontaneous (unprepared) narrative, one has to develop a structure of events and produce oral discourse almost in parallel. This activity demands high resources. The basic skills that a child must acquire are the following: plan a logically well-organized semantic design, transform it into a relevant propositional structure, generate a verbal and syntactically cohesive text, and narrate it fluently to a listener. This activity is cognitively more demanding than a conversation by means of, for example, a dialog. High cognitive loading in the conditions of limited mental resources may result in a trade-off effect [17,41,42]. Suppose the child feels this conflict between the task demands and his/her available resources. In that case, it is possible to explore several strategies: (a) to simplify the lexical or syntactic structure of the utterances or (b) to reduce or simplify the content the child intends to express in the discourse text [43]. For example, it has been found that, in DLD children, the number of noun tokens per word within a story correlated negatively with a story structure, episode completeness, and CL/CU quotient; in TD children, correlations between the given measures were not evidenced. Moreover, in DLD children, visual story complexity significantly increased the percentage of noun tokens (which should be recognized as a negative characteristic for narratives) and the total number of grammatical errors. The more complex a visual story (according to the number of protagonists, actions, and semantically relevant features) was, the higher the noun percentage (during narration, verbs are used quite frequently (since they express actions/events), while nouns are more typical for descriptive texts. On the so-called 'narrativity index', see [44].) and the number of grammatical errors were. Presumably, a child sometimes does not feel a conflict between the task demands and the limitations in his/her language resources and, thus, try to verbalize too sophisticated content by too complex language structures. In such a case, a risk of incidence of speech errors his/her discourse may rise.

It is common knowledge that in some children, lexical and grammatical errors stay rather resistant despite the normal intelligence, hearing, and/or the absence of gross neurological disorders. Children cannot master sufficient language skills without the remedial treatment provided by a speech/language pathologist. This subpopulation is usually recognized as developmentally language-disordered (DLD) children (formerly known as specifically language-impaired—SLI—children) [45–47]. Different models of the DLD have been suggested. In some of them, it has been proposed that children with the DLD are not sensitive to main language markers and cannot process some language units (inflections, suffixes, prefixes, etc.) properly [48]. In these models, the DLD is treated as deviance of language development and its leading cause is defined as the domain-specific impairment. 'For many years, this was seen as a consequence of a deficit either in perceptual processing or in underlying language representations, depending on one's theoretical persuasion' [49] (p. 5). However, in some studies, it has been found that the DLD children made the same errors as the younger TD children [50]. For example, some comparative studies of grammatical features between the DLD children and younger TD children with the same MLU rate did not reveal any significant distinctions [51].

Moreover, to review the basic lexical and grammatical errors in DLD children, most of them can also be found in TD children at the early stage of their language development [52]. Thus, DLD children are recognized as children with delayed language development [45]. Adherents of the given view believe that slow and improper language development in DLD children is not caused by the inability to learn language rules or low perception of particular linguistic features. Instead, it is supposed that the main limitations preventing these children from typical language development are limited cognitive resources domain-general models [52,53].

On the other hand, several studies in DLD children have revealed some weaknesses of the executive functions (EF), such as working memory deficit, low cognitive flexibility, selective inhibition deficit, and impulsivity [54,55]. The more complex the speech programming activity is, the more essential high EF resources are required [56,57]. During the utterance programming, different cognitive actions (lexemes and word form selection, serial order lexical arrangement, inflectional morphemes selection, etc.) compete for the cognitive resources [22,58].

As for speech/language errors in DLD children, it is usually implicated that their language drawbacks in discourse production are the same as in sentence production. However, some studies have revealed limitations in the discourse production only, which is not a consequence of the low language competence [59–61]. The majority of such limitations were revealed in story (re-)telling. DLD children produced less structured and cohesive texts with poor macrostructure and erroneous microstructure in comparison to their TD peers [43,62–64]. Among the different causes of the discourse drawbacks, scholars have highlighted the cognitive resource deficit found in DLD children [65]. It was found that many DLD children have weak EF [66] and a small volume of working memory [67]. In the DLD population, more severe and persistent EF deficiency prevents children from developing proficient language performance skills. Due to the deficit of EF, the error rate in DLD children discourse usually rises [45]. It was established that in discourse production, different genres have distinct procedural demands [68] and, thus, can provoke speech/language errors to different extents [69].

Several studies have revealed that the discourse genre significantly impacts the part-of-speech distribution both in the TD and in the DLD children [70]. Thus, it is reasonable to expect an impact of the discourse genre on the error rate in children's speech. It was hypothesized that children's errors while producing a discourse might have different domain-specific (linguistic limitations) and domain-general (cognitive resource deficit) mechanisms. The former should result in more regular errors, while the latter should result in more incident errors. It seems reasonable to expect speech error distribution to be a rather variable measure. Both in TD and DLD children, speech error number and distribution might be influenced by multiple variables, such as age, language competence, level of discourse skills, individual cognitive resource, genre of discourse, and/or register of communication. Thus, the quantitative statistical measures of the error distribution should be informative to understand the nature of errors.

The aim of the current research was to test the hypothesis that the cognitive loading of discourse may differently provoke lexical and grammatical errors in TD (at different stages of age) children and in DLD children (in comparison to TD peers). The prediction was that different biological age, as well as different registers of communication and the (sub-)type of discourse genres differently impact the distribution of lexical and grammatical errors. In study 1, we aimed at assessing the impact of age on the lexical and grammatical error distribution in different discourse genres in TD children. In study 2, we presumed to reveal differences between the impact of genre on the lexical and grammatical error distribution between the TD and DLD children.

2. Materials and Methods

According to the main aim of the current project, two data sets were composed. Study 1 enabled us to assess longitudinally a group of TD children and to elicit discourse texts of different genres from the same children at four different age stages (Tables 1 and 2). In study 2, two groups of children (DLD as the experimental and TD as the control group) were assessed cross-sectionally and the same discourse genres as in study 1 were elicited in slightly different semi-experimental conditions (Tables 3 and 4).

2.1. Data Collection in the Study 1

2.1.1. Participants

Participants were monolingual Russian-speaking children attending stated kindergartens in Saint-Petersburg, Russia (Table 1).

Table 1. Sample characteristics of study 1.

N	12			
Mean age	Wave 1	Wave2	Wave 3	Wave 4
	4 years 3 months	4 years 8 months	5 years 3 months	5 years 9 months
Language	Russian			
City of residence	Saint-Petersburg, Russia			
Language development	Typically developing			
Inclusion criteria	Normal non-verbal IQ			
Exclusion criteria	Hearing and/or visual disorders			
	Neurological disorders			
	Speech and/or language impairment			
	Non-verbal IQ on Raven's matrix below 84			

In all participants, the nonverbal IQ (according to the Raven's Colored Progressive Matrices Test) was at a normal range (M = 109.13, SD = 5.44).

Initially, 36 children were selected for the study from a large sample of 60 children previously screened by the speech-language pathologist and confirmed as as typically developed children without any speech/language disorder. The 36 children passed as many as possible of the designed assessment sessions within each of the waves; however, a few of the children were excluded from the study after the 1st, 2nd, or 3rd wave due to different reasons; moreover, not all the children participated in all the assessment sessions. Thus, finally, for the current study, we selected 12 children of the most similar age, who passed all the waves and all the assessment sessions within each of the waves.

2.1.2. Procedure

The data contained genres such as personal narrative, fictional story (telling and retelling mode), chat, and discussion, in each wave of the assessment (Table 2). The given genres were different from the perspective of form (monologue vs. dialogue) and register of communication (peer-directed vs. adult-directed speech). The size (the total number of words) of the data is presented in Table 2.

The conversational map methodology [38] was modified by the authors of the paper to elicit peer-directed chats and personal narratives. During individual sessions, a child was given a doll (whose name was 'the same as the child's name' and whose age was also 'the same as the child's age'), and then he/she was involved in a chat with another doll (the experimenter's doll). In the flow of conversation, the experimenter's doll told a personal narrative about some of his/her experience and then asked the child's doll whether she/he has had a similar experience. In response to the given prompts, the children usually told personal narratives about similar experiences (events that had happened with the child or his parents/grandparents/siblings/friends, etc.). While a child was telling his/her narrative, the experimenter tried to minimize her impact, i.e., she did not help the child, did not ask any questions, and did not make any corrections. The only allowed prompts were neutral remarks, such as 'Uh-huh', or a repetition of the exact previous words of the child. In each of the waves, three personal narratives were elicited from each child. The experimenter introduced the following topics in the prompt narratives: shopping in a supermarket, visiting a doctor, a journey to a zoo, swimming in a lake, and some others familiar for children of the given culture and lifestyle. Moreover, the children were allowed to tell personal narratives on any topic they offered.

Table 2. The data of study 1. The total number of words.

Genre	Register	Wave 1	Wave 2	Wave 3	Wave 4
Personal narratives	Peer-directed speech	1526	1869	3030	3164
Fictional stories (telling)	Peer-directed speech	544	484	706	743
Fictional stories (telling)	Adult-directed speech	576	541	557	868
Fictional stories (retelling)	Adult-directed speech	593	674	618	794
Chats	Peer-directed speech	1356	2406	3910	3216
Discussions	Peer-directed speech	1332	1670	1932	1859

To elicit adult-directed fictional stories, the children were asked to tell and retell a story according to different picture sequences. In each of the waves, we used two picture sequences consisting of six colored pictures each; the protagonists of the picture sequences were animals quite familiar for children of the given culture and lifestyle (e.g., cats, dogs, mice, crawls) but one of the stories was always more complex than another one. The pictures were designed by the authors of the paper and painted by a professional artist, Ms. Unė Kurtinaitytė. During the assessment, the experimenter first placed the pictures in the correct sequence in a single, horizontal row in front of a child. Then, the child was allowed to look at the pictures to get the gist of the story (the time was unlimited). Next, the child was asked to tell a story according to the pictures (for the telling mode) or to listen to the story read by the experimenter and then to retell it (for the retelling mode). During the telling/retelling process, the pictures were still on the table and, thus, a child had a possibility to look at them all the time of assessment. The order of the tasks was counterbalanced regarding story complexity (easier story vs. more complex story) and task mode (telling vs. retelling). Sessions of the 1st and 2nd tasks were separated by a few minutes of chat between the experimenter and a child.

To elicit peer-directed fictional stories and discussions, we used the same dolls for stimulating the personal narratives and chats. At the beginning of the assessment, the experimenter's doll demonstrated hand-made his/her 'favorite book with pictures inside' and offered the child to look at them. Then, the experimenter's doll 'accidentally' dropped the book, and the images spilled on the table (the images were not stuck initially to the pages of the book). The experimenter's doll asked the child's doll for help and they both tried to place the pictures to repair the picture sequence. During this activity, the experimenter's doll offered erroneous versions of the picture sequence, asked many provoking questions, and, thus, involved the child in a discussion.

2.2. Data Collection in the Study 2

2.2.1. Participants

Participants were 10 monolingual Russian-speaking children with DLD and 14 TD peers (see Table 3).

DLD children were clinically referred and received a two-year course of speech therapy (five sessions a week) at the kindergarten; nevertheless, various phonetic, lexical, and grammatical errors still occurred in their speech data. Before the experiment, all the children were assessed by a speech-language pathologist by means of Russian language assessment tools [50] in order to confirm the TD vs. DLD status and to exclude children with language comprehension disorders from the experiment. (Children with language comprehension disorders were excluded to escape the non-relevant variable impact and to highlight the discourse production features related to the expressive language limitations.) Thus, the experimental group in our study may be characterized as DLD children with language expression but not comprehension difficulties. Nonverbal IQ (according to Raven's Colored Progressive Matrices Test) was at a normal range in both groups (*M* = 113.25, *SD* = 3.15 in the DLD; *M* = 121.00, *SD* = 4.90 in the TD).

Table 3. Sample characteristics of study 2.

	Experimental Group (DLD)	**Control Group (TD)**
N	10	14
Mean age	6 years 5 months	
Language	Russian	
City of residence	Saint-Petersburg, Russia	
Language development	Primarily impaired	Typically developing
Inclusion criteria	Normal non-verbal IQ	Normal non-verbal IQ
	Clinically referred DLD	
Exclusion criteria	Language comprehension disorders	Speech and/or language impairment
	Hearing and/or visual disorders	
	Neurological disorders	
	Non-verbal IQ on Raven's matrix below 84	

2.2.2. Procedure

The data contained genres such as fictional story (telling and retelling) and discussion (Table 4). The given genres were different from the perspective of the form (monologue vs. dialogue) but belonged to the same register of communication (adult-directed speech). The size (the total number of words) of the data is presented in Table 4.

Table 4. The data of study 2. The total number of words.

Genre	Register	DLD	TD
Fictional stories (telling)	Adult-directed speech	520	1130
Fictional stories (retelling)	Adult-directed speech	314	1156
Discussions	Adult-directed speech	6249	8687

The children were asked to tell and retell a story according to different picture sequences to elicit fictional narratives. The order of the tasks was counterbalanced regarding story complexity (as in study 1, one of the stories was easier, while another one was more complex) and task mode (storytelling vs. retelling). The methodology has been previously presented in several publications [43,71].

To elicit discussions, the 'nonsense picture method' [72] was employed. An experimenter demonstrated a child a picture with an unrealistic scenario (e.g., a cow sitting on a tree, a pig flying in the sky, etc.) and asked the child to evaluate its plausibility. During the conversation, the experimenter asked the child as many as possible provoking questions and tried to involve the child in a discussion [72].

2.3. Development of the Corpus of Russian Children's Language

The data (audio-records of the sessions) were transcribed orthographically using CHAT tools [73]. Two experts double-checked the transcriptions independently and extended them by encoding for language errors and linguistic dysfluencies to perform automated analysis using CLAN tools [73]. The data analyzed comprised, in total, 45, 942 words without mazes.

2.4. The Analyzed Variables

Russian is a highly inflective and morphologically rich language with many grammatical categories. First, Russian nouns are specified as animate vs. inanimate. Furthermore, as mentioned above, Russian nouns fall into three declensions related to the gender form and may be further divided into different declension paradigms; moreover, nouns involve two number forms and six case forms. Each Russian verb is specified for perfective vs. imperfective aspect; the given forms of aspect are differentiated by prefixation, suffixation (or a combination of them), and root alternation (suppletion). Verbs are also specified for reflexive vs. non-reflexive form, they involve three tense forms (past, present, and future tense) and three mood forms (indicative, imperative, and conditional). There are five participles different in passive vs. active form and in perfective vs. imperfective aspect. Participles are declined as adjectives; the two indeclinable adverbial participles are often called gerunds. In Russian, nouns agree with their modifiers in gender, case, and number; verbs agree with nouns and pronouns in gender, person, and number. Word order in Russian is syntactically flexible and determined by pragmatics [11]. The most common syntactic structure 'subject-predicate-object may have up to 30 possible word order variants, including six with the subject omission [74]. In addition, subordination using participles and gerunds instead of relative and adverbial clauses is common in Russian.

A significant number of studies in Russian child language has been devoted to grammar acquisition [10,75,76], and, thus, the qualitative features of morphological errors in the early noun, verb, pronoun, and adjective production are well-studied; however, quantitative studies have been carried out much rarely, especially at the later stages of language acquisition. Similarly, the qualitative features of lexical and derivational errors have been widely analyzed [77,78], however, we still lack knowledge about their quantitative features. Therefore, this study analyzed both qualitative and quantitative features of the following types of language errors, as exemplified below.

Lexical errors were considered as cases in which the child referred to something by another name. In our corpus, lexical errors fell into (a) hyponyms replaced by hyperonyms (e.g., 'mouse's house' instead of 'mouse's hole'), (b) hyperonyms replaced by hyponyms (e.g., 'chicken' instead of 'baby-bird'), and (c) incorrect selections within words of the same semantic group (e.g., motion terms 'to go', 'to run', and 'to fly').

Morphological errors were considered as cases in which the child selected an inappropriate form for marking the case, declension paradigm, gender, number, person, tense, aspect, mood, or other morphological categories. Typical examples of morphological errors were as in the following:

(1) *Tam byla* [= byl] *poni.* 'There was:FEM [= was: MASC] a ponny: MASC.'

(2) *Koshka zabiralasj* [= zabralasj] *na derevo.* 'The cat was climbing [= climbed] up the tree.'

In the case (1), the child used a feminine form for the word 'was' instead of a masculine form, while the noun 'ponny' falls to the masculine gender. In the case (2), the child used incorrectly an imperfect aspect form instead of the proper perfect form.

Syntactic errors were divided into (a) the agreement errors, (b) the government errors, (c) the incorrect prepositional constructions, (d) the omissions of functional words, (e) the incorrect word order, and (f) other 'messy sentences'. Typical examples of the syntactic errors were as in the following:

(3) *Ona stala zalezatj k derevu* [= na derevo]. 'She started climbing to [= up] the tree.'(4) *Ptica ne mozhet letatj s vodoj* [= v vode]. 'A bird cannot fly with [= in] a water.'

In both cases (3,4), the child used incorrect preposition (although the government of the preposition on the noun case form was correct). The given syntactic errors (3,4) not only destructed the sentence but also involved inappropriate semantics.

Derivational errors fell into (a) the incorrect prefixation, (b) the incorrect suffixation, and (c) the incorrect compounding. Usually, derivational errors were occasional word-forms constructed by a child (e.g., a diminutive form of 'a pet' or 'a swing' that are possible theoretically but do not exist in Russian).

During statistical analysis, the quotient for each type of error was estimated by dividing the number of errors of the given type by the total number of words in the corpus. Thus, four different quotients were estimated: (1) the lexical error Q_L, (2) morphological error Q_M, (3) syntactic error Q_S, and (4) derivational error Q_D. The statistical difference between the measures was calculated by means of the ANOVA (in the case of normally distributed variables) or by the Mann–Whitney U criterion (in the case of non-normally distributed variables). The MANOVA (general linear model—GLM) was applied to estimate the determinants' impact on the dependent variables.

3. Results

In this section, we present separate results of study 1 and study 2.

3.1. Results of the Study 1

Although the participants had no speech/language disorders, they made a low number of errors, i.e., 0.02 errors per word (or 20 errors per 1000 words) with high individual variability (from 0 to 0.19). As for different types of errors, their distribution in the whole data was uneven (Figure 1).

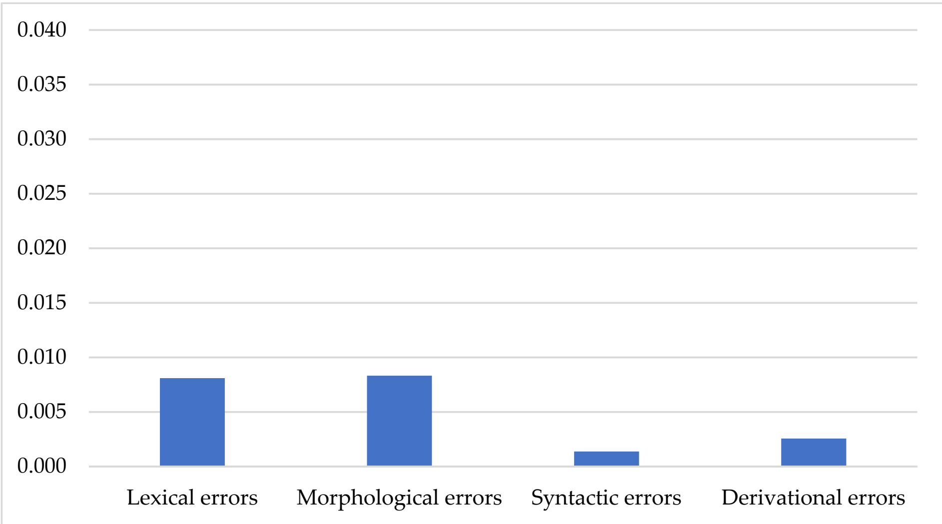


Figure 1. Error rate quotients per word in different types of errors.

Lexical (Q_L) and morphological (Q_M) error quotients were significantly highest among all types of errors, while syntactic (Q_S) error quotient was the lowest (Table 5).

Table 5. The error quotient among the waves.

	The Number of Errors Per Word Mean (Standard Deviation)			
	Lexical Errors	**Morphological Errors**	**Syntactic Errors**	**Derivational Errors**
Wave 1	0.0125 (0.0270)	0.0173 (0.0019)	0.0007 (0.0030)	0.0031 (0.0088)
Wave 2	0.0081 (0.0189)	0.0153 (0.0017)	0.0024 (0.0082)	0.0022 (0.0077)
Wave 3	0.0022 (0.0048)	0.0118 (0.0013)	0.0013 (0.0037)	0.0023 (0.0069)
Wave 4	0.0096 (0.0143)	0.0076 (0.0008)	0.0011 (0.0032)	0.0026 (0.0057)
Significance				
1–2	0.690	1.000	0.178	1.000
1–3	0.002	0.048	1.000	1.000
1–4		0.021	1.000	1.000

Along the waves, the mean of the total error quotient (Q_T) changed following the U curve, but the differences between the waves did not reach a significant level. The maximal error rates were evidenced in the 1st wave, while the minimal error rates were evidenced in the 3rd wave (Figure 2).

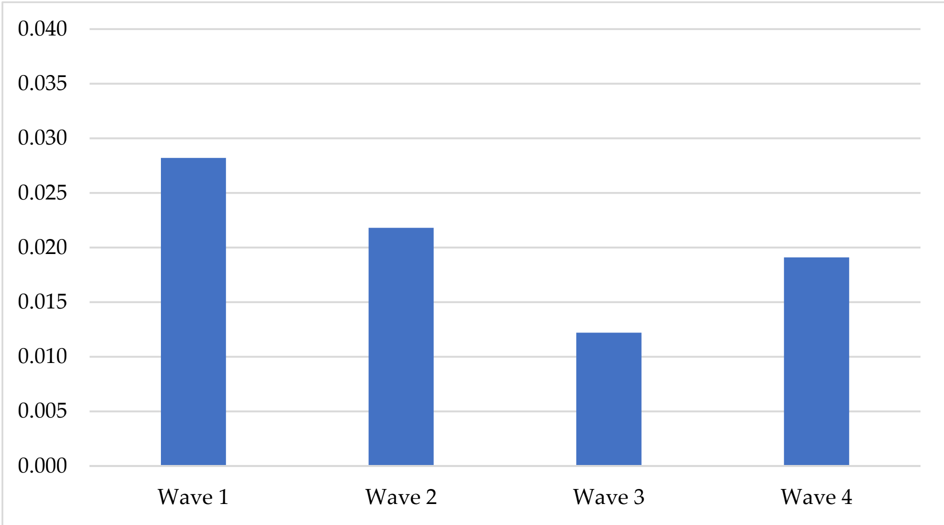


Figure 2. Total error rate per word in the waves of assessment.

Distribution of the quotient for all types of errors are presented in Figure 3.

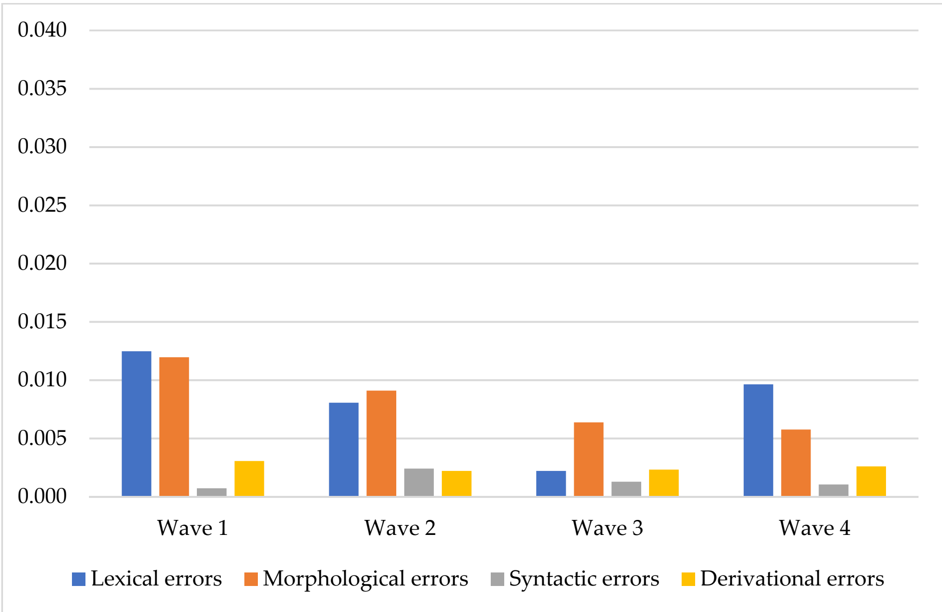


Figure 3. Distribution of different types of error quotients in the waves of assessment.

The total number of errors per word changed gradually from wave to wave but all types of errors had their specific patterns. The most dramatic change was revealed between the 2nd and the 3rd wave.

A more selective quantitative analysis of the rate of different types of errors in each wave revealed distinctions in the errors' 'behavior'. The lexical error rate was maximal in the 1st wave and minimal in the 3rd wave (U-shaped pattern), while the syntactic error rate, vice versa, was minimal in the 1st and in the 4th wave and maximal in the 3rd wave (inverted U-shaped pattern), and the morphological errors reduced gradually along the waves.

As for the percentage of different types of errors among all errors, it was rather stable in the first three waves but changed in the 4th wave (Figure 4).

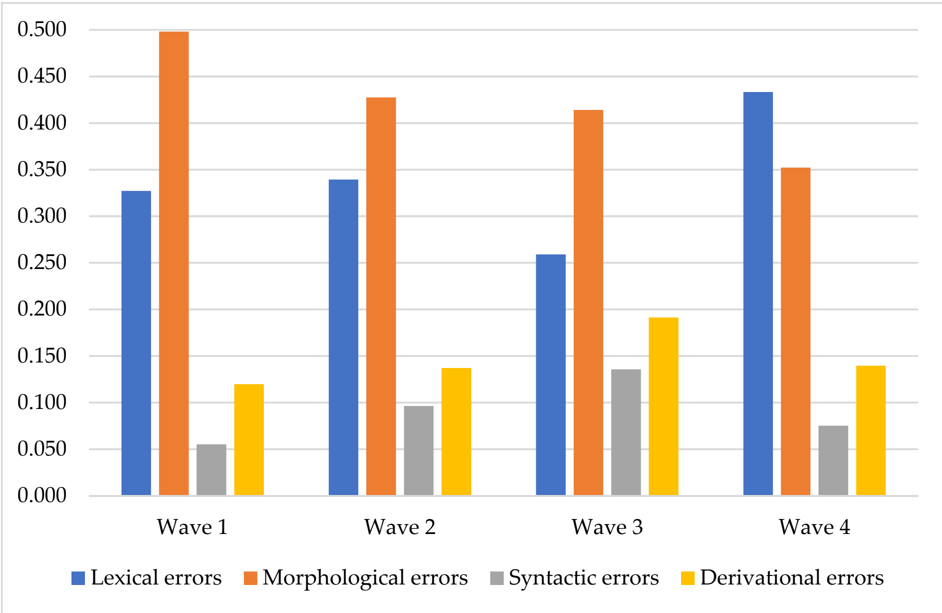


Figure 4. The percentage of different types of errors (among all errors) in the waves of assessment.

To estimate the impact of such determinants as Wave and Genre on the error distribution, the MANOVA analysis of the total data in all the waves was carried out (Table 6).

Table 6. The impact of group and genre determinants on error distribution.

Dependent Variables	**Independent Variables**			
	F	***p***	***η^2***	***Power***
		Wave		
Lexical errors per word	5.365	0.001	0.050	0.932
Morphological errors per word	3.793	0.011	0.036	0.813
		Genre		
Lexical errors per word	5.794	0.000	0.117	0.999
Morphological errors per word	2.086	0.045	0.046	0.797

Both Wave and Genre significantly impacted the Q_L and Q_M indexes with relatively low effect size.

Comparative between-genre ANOVA analysis with a post hoc multiple pairwise comparisons revealed that the Q_L quotient in the storytelling was higher than in the personal narrative ($p = 0.001$), retelling ($p = 0.022$), chat ($p = 0.000$), and discussion ($p = 0.000$). As can be seen in Figure 5, the register of communication (adult-directed story vs. peer-directed story) significantly influenced only the distribution of lexical errors in the storytelling.

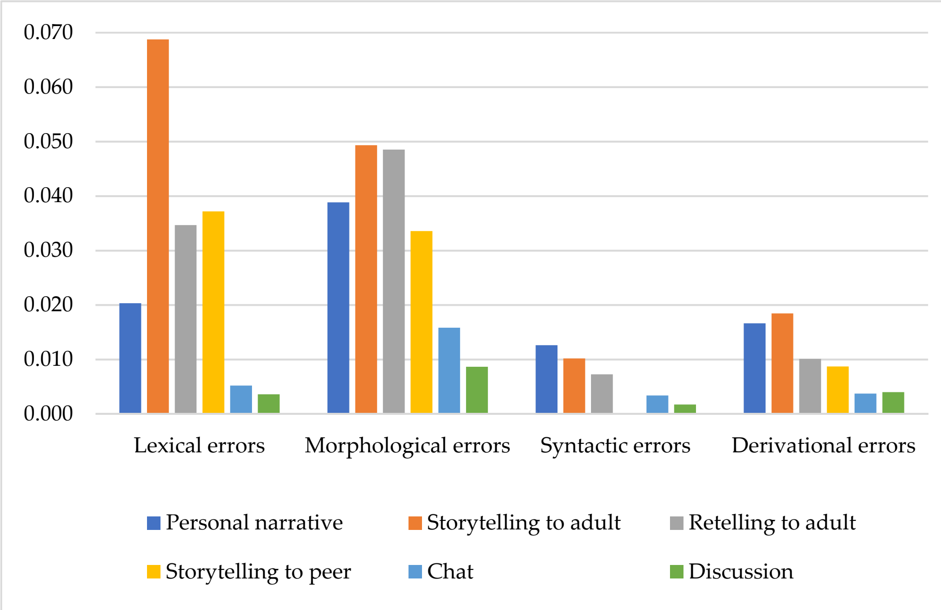


Figure 5. The impact of *genre* on error distribution.

3.2. Discussion on the Study 1

The aim of study 1 was to assess the impact of Age variable on the lexical and grammatical error distribution in different discourse genres in TD children. Distributional analysis of the total number of errors revealed that the lexical and morphological errors were dominant, while the syntactic and derivational ones were much rarer. This result is congruent with our knowledge about the patterns of development of different parts of language devices [11,78,79]. The basic syntactic skills become matured in the second and third years of life. Acquisition of the derivative words and mastering the derivational morphemes begin at about 2 years 6 moths and develop the most active during the 4th and 5th years, usually resulting in a high number of word innovations. On the other hand, the differentiation between the word meanings and the development of a polysemy continue along the school years until adolescence. Basic morphology devices are mastered by typically developed children at about the 6th or 7th year of life. In this regard, we can infer that earlier developed speech/language skills are more resistant to errors, while later developed ones, conversely, are more variable and more error prone. Our participants' total number of errors significantly reduced in the 3rd wave compared to the 1st one. Until recently, such data regarding the changes in error patterns during late preschool years were not available. A small amount of data with error samples have been obtained instead in special experimental assessments, but not in the spontaneous discourse; spontaneous data, if accepted, have not been supported by an age-related analysis [79,80]. To our knowledge, there have not been any previous attempts to analyze the entire range of error profiles in TD children's discourse.

It is difficult to compare our data with other data. Among the publications related to grammatical errors in children's speech [80,81], very few contain statistics of distribution. For example, [80] presented a collection of child errors, and in his corpus, lexical errors occupy 92% and morphological 8% of all errors. This is much less than in our data (35%). Furthermore, it should be noted that in different languages the weight of the formally similar types of errors may have a different mechanism and may be grounded in different strategies of the speaker.

Analysis of the age-related changes in four types of errors presented new data about the discourse development in children from 4 to 6. As shown in Figures 2 and 3, participants became more skillful in the discourse production during the two years of a longitudinal study. The total number of errors changed following the U-shaped curve. This trend is relevant to many other developmental studies representing the same pattern [2,82]. However, the rate of different types of error changed according to different patterns. The most clear-cut patterns of changes were presented in the lexical and morphological errors. The morphological errors gradually reduced and followed a descending linear curve. On the other hand, the lexical errors path was observed as a U-shaped curve—they radically reduced after the 2nd wave (mean age 4 years 8 months) and raised after the 3rd (mean age 5 years 3 months). One more interesting point in our data was the internal distribution of different types of errors. The pattern of this distribution was constant along the first three waves and changed in the 4th wave. This occurred because the rate of the morphological errors was raised, but the lexical errors were reduced in only the 4th wave. Concurrently, the remaining two types of errors remained constant. This could probably be explained by the fact that age from 5 to 6 is critical for the development of some cognitive resources and executive functions [26,83,84].

In our participants, individual variabilities of error distribution were probably caused by the two main circumstances: (a) partial immaturity of some discourse language skill subsystems and (b) different cognitive demands of the discourse genres. The last inference is supported by our finding of the distinct and significant impact of different genres on the error distribution. Moreover, in different types of error this impact manifested with distinct patterns (Figure 5).

To our knowledge, this study is the first to confirm the impact of age and genre on the error distribution in the child discourse. Moreover, statistical analysis revealed that the assessed four types of errors were differently sensitive to age and genre. It should be emphasized that the four types of errors represent the basic sub-systems of language development and distinct patterns of language maturity: (1) growth of lexical and conceptual diversity, (2) development of inflectional morphology devices, (3) development of syntactic structures, and (4) development of derivational morphology devices [79]. Only the lexical errors were similarly by the genre and only the morphological errors were sensitive to age. Some well-known developmental patterns can explain this. According to [10,78], children continue to master morphology devices for noun inflections which play a crucial role in Russian. However, this grammatical device seems to be not fully mastered and not yet automatized in most of our participants. As for the accuracy in the vocabulary use, children of this age range acquire many new infrequent words relatively fast, but the semantic differentiation develops later [79]. Children's word choice errors reflect incomplete knowledge of the meaning of the incorrectly used words [2]. The discourse programming demands more conceptually rich vocabulary than utterance production. In this regard, lexical errors are multiple and influenced by age and genre. The same data has been published previously by other authors [2,79].

3.3. Results of the Study 2

The aim of the study 2 was to reveal differences between the impact of genre on the lexical and grammatical error distribution between the TD and DLD children. The total number of errors per word in the DLD group was higher than in the TD group ($F = 6.114$; $p = 0.025$) (but only in the conversation). Comparative estimation of different types of errors revealed that the DLD children made significantly more lexical errors per word ($F = 4.530$; $p = 0.037$) and almost significantly more syntactic errors ($F = 3.719$; $p = 0.058$).

The MANOVA analysis of the impact of Group and Genre on the error rate in all participants (DLD + TD) revealed that both Group and Genre determinants predicted the lexical and morphological errors rates (Table 7).

Group significantly determined the morphological errors and almost significantly determined the lexical and syntactic errors. On the other hand, Genre significantly determined the lexical errors. The comparative between-group analysis of error distribution in the storytelling and retelling did not reveal significant distinctions between the groups. However, the same analysis of error distribution in the discussion revealed different results:

the DLD children made significantly more lexical errors (U = 11.5; p = 0.018) and syntactic errors (U = 13.5; p = 0.032) (Figure 6).

Table 7. The impact of Group and Genre determinants on error distribution.

Dependent Variables	Independent Variables			
	F	*p*	η^2	*Power*
		Group		
Lexical errors per word	3.849	0.054	0.058	0.489
Morphological errors per word	4.116	0.047	0.062	0.515
Syntactic errors per word	3.520	0.065	0.054	0.455
		Genre		
Lexical errors per word	3.919	0.025	0.112	0.686

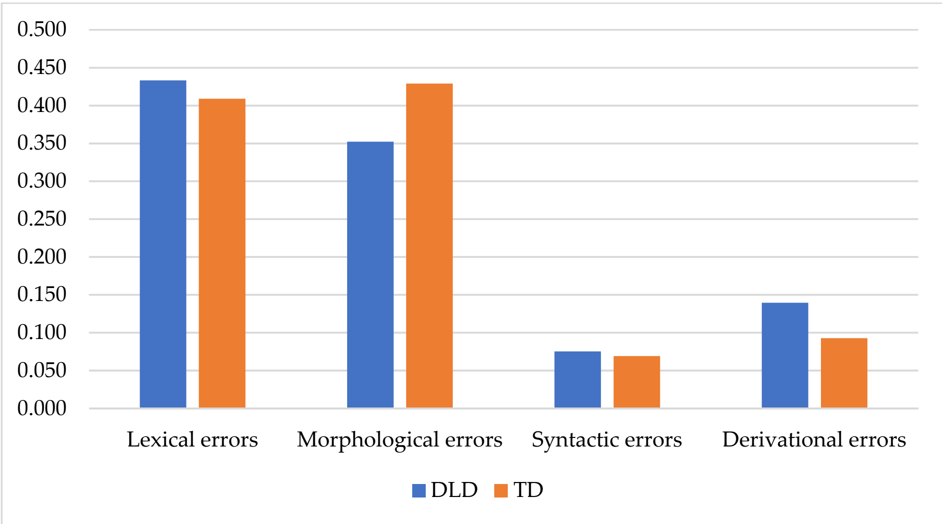


Figure 6. Error percentage distribution in TD and DLD children.

To estimate the Genre's impact on the different errors' rates within the same group, a comparative analysis of errors per word in the storytelling and conversational reasoning was carried out. In the DLD group, we did not find a difference between the genres. In the TD group, significant differences between the storytelling and conversation were in the morphological (U = 34.0; p = 0.000) and syntactic errors (U = 70.0; p = 0.034) (Figure 7).

Statistical comparison (Mann–Whitney test) of the error quotient rate distribution in storytelling, retelling, and discussion evidenced that Genre (storytelling vs. discussion) significantly discriminated the Q_M. (U = 119.0; p = 0.006), Q_S (U = 150.0; p = 0.013), and Q_D (U = 139.5; p = 0.006) (Figure 7).

Comparative analysis of the distribution of error quotient in different genres in the DLD group confirmed that Genre (storytelling vs. discussion) significantly discriminated only the syntactic errors (U = 22.5; p = 0.006). However, a comparison of the percentage of the types of errors among all errors in the DLD children in different genres revealed significant differences between the lexical (U = 82.0; p = 0.010) and syntactical errors (U = 16.5; p = 0.017).

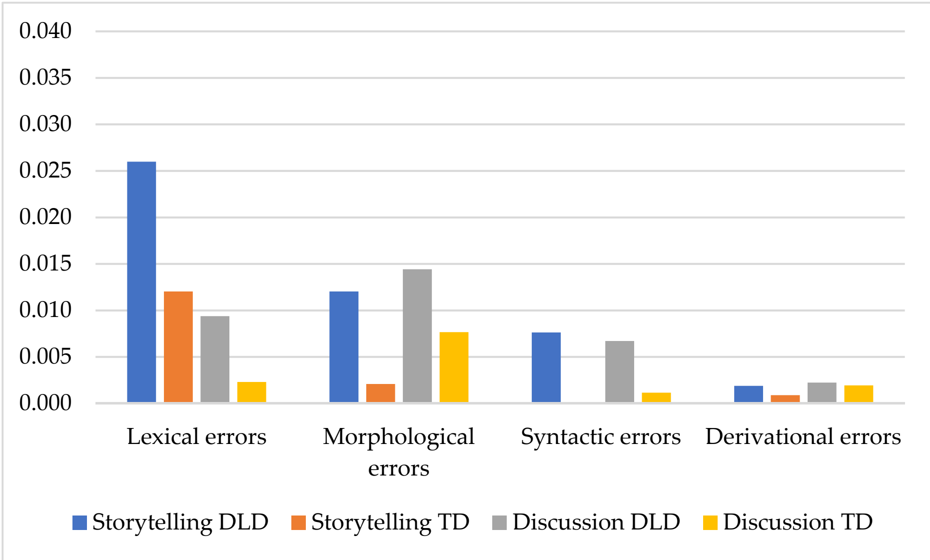


Figure 7. Error distribution in the storytelling and discussion. The number of errors per word.

3.4. Discussion on the Study 2

At the beginning of the discussion, we need to remind that all DLD participants were attending special kindergartens for DLD children and receiving remedial treatment course. At the beginning of the experiment, the DLD participants had been taking part in the remedial treatment for approximately two years. As a result, some speech/language drawbacks (especially, in morphological and syntactic domains) had probably been partially compensated before the beginning of the experiment. Nevertheless, the scores of language assessment (phonological and morphosyntactic tasks) used to select the children for the experiment were below the age-norms in all the DLD children. Moreover, it was noticed that in the semi-structured discourse elicitation conditions, many of the DLD participants were trying to escape linguistically demanding phrases. Comparative analysis of narrative microstructure in the same sample of the DLD and TD children (see [43]) revealed that the DLD children tended to use simple syntactical structures, their communication units were not complex (the mean number of clauses per communication unit was low), and the lexical diversity index (noun lemma/token ratio) was also quite low. Although the DLD and TD children did not demonstrate significant differences in the total number of errors in storytelling, discussions about the 'nonsense pictures' provoked the DLD children to make more errors than the TD children. This result concord to studies carried out in other languages [85] and might be explained by different nature of the given genres. Narrative (storytelling according to picture sequence in a self-pace mode) might be characterized as partially prepared and structured speech, while the discussion is much more spontaneous and much less structured genre. Moreover, in the current study, Genre impacted differently not only the total number but also a distribution of different types of errors. Namely, cognitively high demanding tasks (such as storytelling elicited following quite structured procedure) provoked the DLD children to make a lot of morphological errors; while spontaneous, playful discussions with an experimenter lead the DLD children to numerous lexical errors. This tendency concords with our previous experiments [71]. Thus, the results of study 2 evidenced a selective impact of the discourse genre on error distribution in the DLD children.

3.5. Results of Comparison between the Study 1 and Study 2

Comparison of the error rates between the TD children from the study 1 wave 4 (mean age 5 years 9 months), the TD children from study 2 (mean age 6 years 5 months), and their DLD peers (mean age 6 years 5 months) did not reveal any significant distinctions (Figure 8).

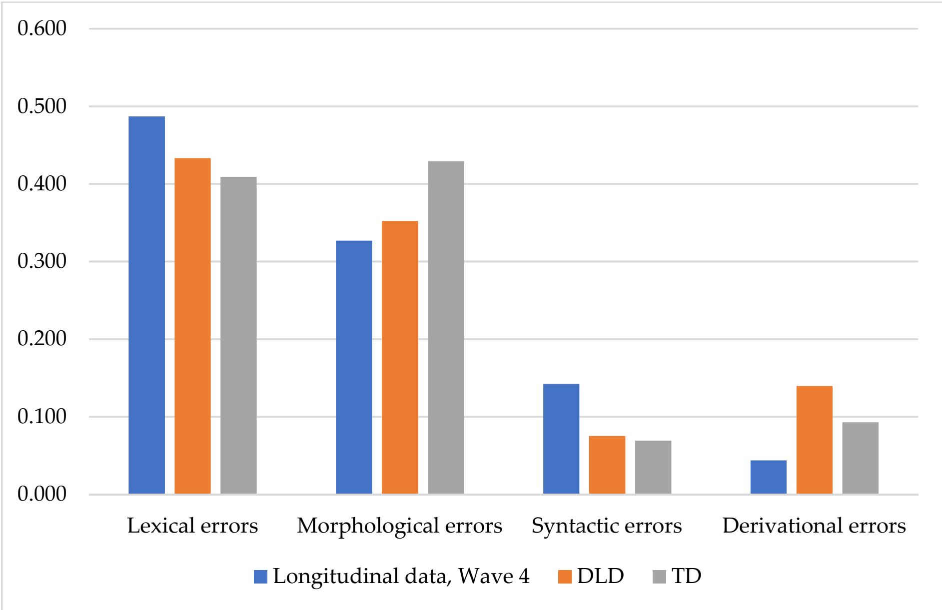


Figure 8. Distribution of the percentage of different error types within the groups from study 1 and study 2.

4. General Discussion on the Study 1 and Study 2

Our two studies evidenced that the correctness of speech production in the same children is not a constant measure. The high variability of error scores was significantly influenced by the genre of discourse and the age of participants. Thus, our prediction about the impact of the genre on the error rate was confirmed. This may be explained by the different cognitive loading of different genres. These data are congruent with our previous studies where the genre impact on the distribution of phonological structures (types of syllables with different complexity) [71] and on the part-of-speech profile [70] was evidenced.

It should be noted that the distribution of the percentage of different types of errors was very similar between the TD and DLD children. However, the latter made significantly more lexical errors.

According to the traditional attitudes of speech/language pathology, low language competence is the main distinction between language disorders and TD children. The development of language competence is usually equated to the score of correct responses in language battery tasks (the proper response is considered the response without errors). Thus, it is implicitly suggested that typically developed children do not make any errors. However, this suggestion is not true. Both adults and children make errors and slips of the tongue in their colloquial speech [2,86]. While there are plenty of psycholinguistic studies of errors in colloquial speech in adults, scarce data are available regarding colloquial speech and errors in children. From the psycholinguistic and psychological perspective, speech is some kind of behavior consisting of multiple skills. From this perspective, procedural knowledge acquisition in speech development is in close relation to mastering relevant skills.

In linguistic studies of children's language, speech errors are recognized as a sign of limited language competence. However, language competence is usually rather stable, and its limitations are expected to manifest approximately similar to different speech acts.

Like many other skills, each new speech skill passes through a stage of unstable, variable execution, and automation. The less mature and automated the skill is, the more variable it is, and the more mistakes are made [87,88]. The number of errors the subject makes is a manifestation of the skill's instability. From the perspective of this regularity, speech errors mark the weak chains in a particular speech/language subsystem. In the scope of this suggested model, higher rates of lexical and morphological errors, compared to syntactic and derivational errors, mean that the syntactic and derivational subsystems, in 4–5-year-old children, are more matured and automated and, therefore, children make fewer errors. For the same reason, the syntactic and derivational errors were less sensitive to age and genre.

The main limitations of our study were relatively small size of the sample and quite narrow range of cognitive assessments. In the framework of the ongoing investigation, we are replicating the longitudinal study (with an application of all methods and approaches employed in the given study 2) in DLD children. This will enable us to compare a process of discourse acquisition between the TD and DLD children from different perspectives, including error analysis.

5. Conclusions

Comparison of the data obtained in the longitudinal and cross-sectional studies evidenced that both in TD and DLD children, genre of discourse significantly impacted the distribution of lexical and grammatical errors. Different genres had a different impact on the pattern of the error distribution. On the other hand, different types of errors were influenced by the genre to different extents. Lexical errors were the most sensitive to the genre both in TD and DLD children, but the distribution of only morphological errors discriminated TD children from DLD peers. Moreover, the general patterns of error distribution in different genres of discourse are reasonably similar between TD and DLD children.

Our data confirmed that methods, tasks, and approaches to children's speech/language assessment may have an essential impact on the error rate. Thus professionals (e.g., speech/language pathologists) should consider the genre and communication register. According to our current data, individual variabilities of different types of grammatical devices are not equally sensitive to the demands of different genres.

Author Contributions: Conceptualization, A.N.K. and I.B.; methodology, A.N.K. and I.B.; software, A.N.K.; validation, A.N.K. and I.B.; formal analysis, A.N.K. and I.B.; investigation, A.N.K. and I.B.; resources, A.N.K. and I.B.; data curation, A.N.K. and I.B.; writing—original draft preparation, A.N.K. and I.B.; writing—review and editing, A.N.K. and I.B.; visualization, I.B.; supervision, A.N.K.; project administration, A.N.K.; funding acquisition, A.N.K. and I.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Russian Foundation for Basic Research (RFBR), grant number 19-29-14078.

Institutional Review Board Statement: The study was approved by the Ethics Committee of Saint-Petersburg State Pediatric Medical University (#4/2, 13 July 2011).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to original informed consent provisions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Berman, R.A.; Slobin, D.I. *Relating Events in Narrative. A Crosslinguistic Developmental Study*; Lawrence Erlbaum Associates, Inc.: Hillsdale, NJ, USA, 1994.
2. Bowerman, M. Starting to talk worse: Clues to language acquisition from children's late speech errors. In *U Shaped Behavioral Growth*; Strauss, S., Ed.; Academic Press: New York, NY, USA, 1982; pp. 101–145.
3. Platt, C.B.; MacWhinney, B. Error assimilation as a mechanism in language learning. *J. Child Lang.* **1983**, *10*, 401–414. [CrossRef]
4. Ellis, N.C. Emergentism, connectionism and language learning. *Lang. Learn.* **1998**, *48*, 631–664. [CrossRef]
5. Tomasello, M. The usage-based theory of language acquisition. In *The Cambridge Handbook of Child Language*; C.U.P.: Cambridge, UK, 2009; pp. 69–87.
6. Slobin, D.I. *The Crosslinguistic Study of Language Acquisition: Volume 4*; Psychology Press: New York, NY, USA, 2014.

7. Lieven, E.V.M. Usage-based approaches to language development: Where do we go from here? *Lang. Cogn.* **2016**, *8*, 346–368. [CrossRef]
8. Nelson, K. Structure and strategy in learning to talk. *Monogr. Soc. Res. Child Dev.* **1973**, *38*, 1–135. [CrossRef]
9. Bates, E.; Bretherton, I.; Snyder, L. *From First Words to Grammar: Individual Differences and Dissociable Mechanisms*; C.U.P.: Cambridge, UK, 1988.
10. Gvozdev, A.N. *Formirovanie u Rebenka Grammaticheskogo Stroia Russkogo Yazyka [The Development of Russian Grammar in Children]*; Akad. Pedag. Nauk RSFSR: Moscow, Russia, 1949.
11. Shvedova, N.Y. (Ed.) *Russkaya Grammatika [Russian Grammar]*; Nauka: Moscow, Russia, 1980.
12. Dell, G.S. A spreading-activation theory of retrieval in sentence production. *Psychol. Rev.* **1986**, *93*, 283–321. [CrossRef] [PubMed]
13. Shanker, S. The generativist-interactionist debate over specific language impairment: Psycholinguistics at a crossroads. *Am. J. Psychol.* **2002**, *115*, 415–450. [CrossRef]
14. Karmiloff-Smith, A. Language and cognitive processes from a developmental perspective. *Lang. Cogn. Process.* **1985**, *1*, 61–85. [CrossRef]
15. Kiparsky, P.; Menn, L. On the acquisition of phonology. In *Language Learning and Thought*; MacNamara, J., Ed.; Academic Press: New York, NY, USA, 1977; pp. 77–78.
16. Bates, E.; Dale, P.S.; Thal, D. Individual differences and their implications for theories of language development. In *The Handbook of Child Language*; Fletcher, P., MacWhinney, B., Eds.; Blackwell: Oxford, UK, 1995; pp. 96–151.
17. Bock, J.K. Toward a cognitive psychology of syntax: Information processing contributions to sentence formulation. *Psychol. Rev.* **1982**, *89*, 1. [CrossRef]
18. Higgs, T.V. Language acquisition and language learning: A plea for syncretism. *Mod. Lang. J.* **1985**, *69*, 8–14. [CrossRef]
19. Igualada, A.; Bosch, L.; Prieto, P. Language development at 18 months is related to multimodal communicative strategies at 12 months. *Infant Behav. Dev.* **2015**, *39*, 42–52. [CrossRef]
20. Strapp, C.M.; Bleakney, D.M.; Helmick, A.L.; Tonkovich, H.M. Developmental differences in the effects of negative and positive evidence. *First Lang.* **2008**, *28*, 35–53. [CrossRef]
21. Kilani-Schoch, M.; Balčiūnienė, I.; Korecky-Kröll, K.; Laaha, S.; Dressler, W.U. On the role of pragmatics in child-directed speech for the acquisition of verb morphology. *J. Pragmat.* **2009**, *41*, 219–239. [CrossRef]
22. Hsu, H.J.; Bishop, D.V. Sequence-specific procedural learning deficits in children with specific language impairment. *Dev. Sci.* **2014**, *17*, 352–365. [CrossRef] [PubMed]
23. Plude, D.J.; Enns, J.T.; Brodeur, D. The development of selective attention: A life-span overview. *Acta Psychol.* **1994**, *86*, 227–272. [CrossRef]
24. Booth, J.R.; Burman, D.D.; Meyer, J.R.; Lei, Z.; Trommer, B.L.; Davenport, N.D.; Mesulam, M.M. Neural development of selective attention and response inhibition. *Neuroimage* **2003**, *20*, 737–751. [CrossRef]
25. Korkman, M.; Lahti-Nuuttila, P.; Laasonen, M.; Kemp, S.L.; Holdnack, J. Neurocognitive development in 5-to 16-year-old North American children: A cross-sectional study. *Child Neuropsychol.* **2013**, *19*, 516–539. [CrossRef] [PubMed]
26. Nelson, K. *Language in Cognitive Development: The Emergence of the Mediated Mind*; C.U.P.: Cambridge, UK, 1998.
27. Deák, G.O. Interrelationship of language and cognitive development (Overview). In *Encyclopedia of Language Development*; Brooks, P., Kempe, V., Eds.; Sage: Thousand Oaks, CA, USA, 2014; pp. 284–291.
28. Scott, C.M.; Stokes, S.L. Measures of syntax in school-age children and adolescents. *Lang. Speech Heart Serv. Sch.* **1995**, *26*, 309–319. [CrossRef]
29. Lieven, E.V.M. Variation in a crosslinguistic context. In *The Cross-Linguistic Study of Language Acquisition. Vol. 5. Expanding the Contexts*; Slobin, D.I., Ed.; Psychology Press: New York, NY, USA, 1997; pp. 199–263.
30. Menn, L. Development of articulatory, phonetic, and phonological capabilities. In *Language Production. Vol. 2*; Butterworth, B., Ed.; Academic Press: London, UK, 1983; pp. 3–50.
31. Alderson-Day, B. Verbal problem-solving difficulties in autism spectrum disorders and atypical language development. *Autism Res.* **2014**, *7*, 720–730. [CrossRef] [PubMed]
32. Cooper, C.R.; Cooper, R.G. Skill in peer learning discourse: What develops? In *Discourse Development*; Kuczaj, S.A., Ed.; Springer: New York, NY, USA, 1984; pp. 77–97.
33. Felton, M.; Kuhn, D. The development of argumentative discourse skill. *Discourse Process* **2001**, *32*, 135–153.
34. Postma, A.; Kolk, H.; Povel, D.J. On the relation among speech errors, disfluencies, and self-repairs. *Lang. Speech* **1990**, *33*, 19–29. [CrossRef] [PubMed]
35. Shriberg, L.D.; Kwiatkowski, J. Self-monitoring and generalization in preschool speech-delayed children. *Lang. Speech Heart Serv. Sch.* **1990**, *21*, 157–170. [CrossRef]
36. Postma, A.; Oomen, C.C.E. Critical issues in speech monitoring. In *Phonological Encoding and Monitoring in Normal and Pathological Speech*; Hartsuiker, R.J., Bastiaanse, R., Postma, A., Wijnen, A., Eds.; Psychology Press: London, UK, 2005; pp. 169–178.
37. Berman, R.A. The psycholinguistics of developing text construction. *J. Child Lang.* **2008**, *35*, 735–771. [CrossRef] [PubMed]
38. Peterson, C.; McCabe, A. *Developmental Psycholinguistics. Three Ways of Looking at a Child's Narrative*; Plenum: New York, NY, USA, 1983.
39. Langacker, R.W. Discourse in cognitive grammar. *Cogn. Linguist.* **2001**, *12*, 143–188. [CrossRef]
40. Nicolopoulou, A. The elementary forms of narrative coherence in young children's storytelling. *Narrat. Inq.* **2008**, *18*, 299–325.

41. Crystal, D. Towards a 'bucket' theory of language disability: Taking account of interaction between linguistic levels. *Clin. Linguist. Phon.* **1987**, *1*, 7–22. [CrossRef]
42. Masterson, J.J.; Kamhi, A.G. Linguistic trade-offs in school-age children with and without language disorders. *J. Speech Lang. Heart Res.* **1992**, *35*, 1064–1075. [CrossRef]
43. Balčiūnienė, I.; Kornev, A.N. Doing new things with language: Narrative language in SLI preschoolers. *Est. Pap. Appl. Linguist.* **2016**, *12*, 25–42. [CrossRef]
44. Balčiūnienė, I.; Kornev, A.N. Evaluation of narrative skills in language-impaired children. Advantages of a dynamic approach. In *Atypical Language Development in Romance Languages*; Aguilar-Mediavilla, E.M., Buil-Legaz, L., López-Penadés, R., Sanchez-Azanza, V.A., Adrover-Roig, D.D., Eds.; John Benjamins Publishing Company: Amsterdam, The Netherlands, 2019; pp. 127–141.
45. Bishop, D.V. Grammatical errors in specific language impairment: Competence or performance limitations? *Appl. Psycholinguist.* **1994**, *15*, 507–550. [CrossRef]
46. Leonard, L.B. *Children with Specific Language Impairment*; MIT Press: Cambridge, MA, USA, 1998.
47. Bishop, D.V.; Snowling, M.J.; Thompson, P.A.; Greenhalgh, T.; CATALISE consortium. CATALISE: A multinational and multidisciplinary Delphi consensus study. Identifying language impairments in children. *PLoS ONE* **2016**, *11*, e0158753.
48. Van der Lely, H.K.J. Domain-specific cognitive systems: Insight from Grammatical-SLI. *Trends Cogn. Sci.* **2005**, *9*, 53–59. [CrossRef]
49. Bishop, D.V.M.; Nation, K.; Patterson, K. When words fail us: Insights into language processing from developmental and acquired disorders. *Phil. Trans. R. Soc. B* **2014**, *369*, 20120403. [CrossRef]
50. Kornev, A.N. *Osnovy Logopatologiyi Detskogo Vozrasta: Klinicheskiye i Psikhologicheskiye Aspekty [The Basics of Developmental Logopathology: Clinical and Psychological Issues]*; Rechj: Sankt-Peterburg, Russia, 2006.
51. Grimm, H.; Weinert, S. Is the syntax development of dysphasic children deviant and why? New findings to an old question. *J. Speech Lang. Heart Res.* **1990**, *33*, 220–228. [CrossRef]
52. Archibald, L.M. The reciprocal influences of working memory and linguistic knowledge on language performance: Considerations for the assessment of children with developmental language disorder. *Lang. Speech Heart Serv. Sch.* **2018**, *49*, 424–433. [CrossRef] [PubMed]
53. Smith, L.B. How domain-general processes may create domain-specific biases. In *Language Acquisition and Conceptual Development*; Bowerman, M., Levinson, S., Eds.; C.U.P.: Cambridge, UK, 2001; pp. 101–131.
54. Henry, L.A.; Messer, D.J.; Nash, G. Executive functioning in children with specific language impairment. *J. Child Psychol. Psychiatry* **2012**, *53*, 37–45. [CrossRef] [PubMed]
55. Montgomery, J.W.; Gillam, R.B.; Evans, J.L.; Schwartz, S.; Fargo, J.D. A comparison of the storage-only deficit and joint mechanism deficit hypotheses of the verbal working memory storage capacity limitation of children with developmental language disorder. *J. Speech Lang. Heart Res.* **2019**, *62*, 3808–3825. [CrossRef]
56. Karmiloff-Smith, A. Development itself is the key to understanding developmental disorders. *Trends Cogn. Sci.* **1998**, *2*, 389–398. [CrossRef]
57. Ullman, M.T.; Pierpont, E.I. Specific language impairment is not specific to language: The procedural deficit hypothesis. *Cortex* **2005**, *41*, 399–433. [CrossRef]
58. Vogels, J.; Krahmer, E.; Maes, A. How cognitive load influences speakers' choice of referring expressions. *Cogn. Sci.* **2015**, *39*, 1396–1418. [CrossRef]
59. Wagner, C.R.; Nettelbladt, U.; Sahlén, B.; Nilholm, C. Conversation versus narration in preschool children with language impairment. *Int. J. Lang. Commun. Disord.* **2000**, *35*, 83–93. [PubMed]
60. Miniscalco, C.; Hagberg, B.; Kadesjö, B.; Westerlund, M.; Gillberg, C. Narrative skills, cognitive profiles and neuropsychiatric disorders in 7–8-year-old children with late developing language. *Int. J. Lang. Commun. Disord.* **2007**, *42*, 665–681. [CrossRef] [PubMed]
61. Pearce, W.M.; James, D.G.; McCormack, P.F. A comparison of oral narratives in children with specific language and non-specific language impairment. *Clin. Linguist. Phon.* **2010**, *24*, 622–645. [CrossRef]
62. McCabe, A.; Bliss, L.; Barra, G.; Bennett, M. Comparison of personal versus fictional narratives of children with language impairment. *Am. J. Speech Lang. Pathol.* **2008**, *17*, 194–206. [CrossRef]
63. Blom, E.; Boerma, T. Why do children with language impairment have difficulties with narrative macrostructure? *Res. Dev. Disabil.* **2016**, *55*, 301–311. [CrossRef]
64. Kornev, A.N.; Balčiūnienė, I. Fictional narrative as a window to discourse development: A psycholinguistic approach. In *Events and Narratives in Language*; Badio, J., Ed.; Peter Lang Publishing: Frankfurt am Main, Germany, 2017; pp. 171–188.
65. Weismer, S.E. Capacity limitations in working memory: The impact on lexical and morphological learning by children with language impairment. *Top. Lang. Disord.* **1996**, *17*, 33–44. [CrossRef]
66. Yang, H.C.; Gray, S. Executive function in preschoolers with primary language impairment. *J. Speech Lang. Heart Res.* **2017**, *60*, 379–392. [CrossRef]
67. Weismer, S.E.; Plante, E.; Jones, M.; Tomblin, J.B. A functional magnetic resonance imaging investigation of verbal working memory in adolescents with specific language impairment. *J. Speech Lang. Heart Res.* **2005**, *48*, 405–425. [CrossRef]
68. Hadley, P.A. Language sampling protocols for eliciting text-level discourse. *Lang. Speech Heart Serv. Sch.* **1998**, *29*, 132–147. [CrossRef] [PubMed]

69. Liles, B.Z. Narrative discourse in children with language disorders and children with normal language: A critical review of the literature. *J. Speech Lang. Heart Res.* **1993**, *36*, 868–882. [CrossRef]
70. Balčiūnienė, I.; Kornev, A.N. Quantitative analysis of language competence vs. performance in Russian- and Lithuanian-speaking 6 year-olds. In *Frontiers in Artificial Intelligence and Applications. Human Language Technologies—The Baltic Perspective, Proceedings of the Ninth International Conference of Baltic HLT 2020, Kaunas, Lithuania, 22–23 September 2020*; Utka, A., Vaičenonienė, J., Kovalevskaitė, J., Kalinauskaitė, D., Eds.; IOS Press Book Series; IOS Press: Amsterdam, The Netherlands, 2020; pp. 225–232.
71. Kornev, A.N.; Balčiūnienė, I. Word structure in typically-developing and primarily language-impaired children: A usage based corpus analysis of Russian preschoolers. In *On Under-Reported Monolingual Child Phonology*; Babatsouli, E., Ed.; Multilingual Matters: Bristol, UK, 2020; pp. 201–228.
72. MacWhinney, B. *The CHILDES Project: Tools for Analyzing Talk*, 3rd ed.; Lawrence Erlbaum Associates: Hillsdale, MI, USA, 2000.
73. Stoll, S.; Abbot-Smith, K.; Lieven, E.V.M. Lexically restricted utterances in Russian, German, and English child-directed speech. *Cogn. Sci.* **2009**, *33*, 75–103. [CrossRef]
74. Gagarina, N.V. *Stanovlenie Grammaticheskikh Kategorii Russkogo Glagola v Detskoi Rechi [The Development of Grammatical Categories of the Russian Verb in Child Language]*; Nauka: Sankt-Peterburg, Russia, 2008.
75. Voeikova, M.D. *Rannie Etapy Usvoeniia Detmi Imennoj Morfologii Russkogo Yazyka [Early Phases of the Acquisition of Russian Noun Morphology by Children]*; Znak: Moskva, Russia, 2011.
76. Eliseeva, M.B. *Foneticheskoe i Leksicheskoe Razvitie Rebenka Rannego Vozrasta [Phonetic and Lexical Development of Young Children]*; RGPU im. A. I. Gertsena: Sankt-Peterburg, Russia, 2008.
77. Ceitlin, S.N. *Ocherki Po Slovoobrazovaniju i Formoobrazovaniju [Essays on Word and Form Derivation]*; Znak: Moskva, Russia, 2009.
78. Clark, E.V. *First Language Acquisition*; C.U.P.: Cambridge, UK, 2009.
79. Stemberger, J.P. Speech errors in early child language production. *J. Mem. Lang.* **1989**, *28*, 164–188. [CrossRef]
80. Jaeger, J.J. *Kids' Slips: What Young Children's Slips of the Tongue Reveal about Language Development*; Erlbaum: Mahwah, NJ, USA, 2005.
81. Stavy, R. (Ed.) *U-Shaped Behavioral Growth*; Elsevier: Amsterdam, The Netherlands, 2012.
82. Anderson, V. Assessing executive functions in children: Biological, psychological, and developmental considerations. *Pediatr. Rehabil.* **2001**, *4*, 119–136. [CrossRef] [PubMed]
83. Brocki, K.C.; Bohlin, G. Executive functions in children aged 6 to 13: A dimensional and developmental study. *Dev. Neuropsychol.* **2004**, *26*, 571–593. [CrossRef] [PubMed]
84. Nettelbladt, U.; Hansson, K.; Nilholm, C. Why ask questions? Contextual effects on grammatical structure in the language production of children with specific language impairment. *Child Lang. Teach. Ther.* **2001**, *17*, 89–106. [CrossRef]
85. Fromkin, V.A. *Speech Errors as Linguistic Evidence*; Mouton: The Hague, The Netherlands, 1973.
86. Sporns, O.; Edelman, G.M. Solving Bernstein's problem: A proposal for the development of coordinated movement by selection. *Child Dev.* **1993**, *64*, 960–981. [CrossRef] [PubMed]
87. Zaal, F.T.; Daigle, K.; Gottlieb, G.L.; Thelen, E. An unlearned principle for controlling natural movements. *J. Neurophysiol.* **1999**, *82*, 255–259. [CrossRef] [PubMed]
88. Petryński, W. Formation of the sensorimotor operation pattern from a system-theoretical perspective. *Int. J. Phys. Educ. Fit. Sports* **2018**, *7*, 30–42. [CrossRef]

children

MDPI

Article

Consistency of a Nonword Repetition Task to Discriminate Children with and without Developmental Language Disorder in Catalan–Spanish and European Portuguese Speaking Children

Nadia Ahufinger [1,*], Amy Berglund-Barraza [2], Anabela Cruz-Santos [3], Laura Ferinu [1], Llorenç Andreu [1], Mònica Sanz-Torrent [4] and Julia L. Evans [2]

[1] Estudis de Psicologia i Ciències de l'Educació, Universitat Oberta de Catalunya, 08018 Barcelona, Spain; lferinu@uoc.edu (L.F.); landreub@uoc.edu (L.A.)
[2] Department of Speech, Language, Hearing Sciences, University of Texas at Dallas, Richardson, TX 75080, USA; Amy.Berglund@utdallas.edu (A.B.-B.); jle130030@utdallas.edu (J.L.E.)
[3] Research Centre in Education, Institute of Education, University of Minho, 4710-057 Braga, Portugal; acs@ie.uminho.pt
[4] Facultat de Psicologia, Universitat de Barcelona, 08035 Barcelona, Spain; monicasanz@ub.edu
* Correspondence: nadiahufinger@uoc.edu; Tel.: +34-93-326-3473

Abstract: Nonword repetition has been proposed as a diagnostic marker of developmental language disorder (DLD); however, the inconsistency in the ability of nonword repetition tasks (NRT) to identify children with DLD raises significant questions regarding its feasibility as a clinical tool. Research suggests that some of the inconsistency across NRT may be due to differences in the nature of the nonword stimuli. In this study, we compared children's performance on NRT between two cohorts: the children in the Catalan–Spanish cohort (CS) were bilingual, and the children in the European Portuguese cohort (EP) were monolingual. NRT performance was assessed in both Spanish and Catalan for the bilingual children from Catalonia-Spain and in Portuguese for the monolingual children from Portugal. Results show that although the absolute performance differed across the two cohorts, with NRT performance being lower for the CS, in both Catalan and Spanish, as compared to the EP cohort in both, the cut-points for the likelihood ratios (LH) were similar across the three languages and mirror those previously reported in previous studies. However, the absolute LH ratio values for this study were higher than those reported in prior research due in part to differences in wordlikeness and frequency of the stimuli in the current study. Taken together, the findings from this study show that an NRT consisting of 3-, 4-, and 5-syllable nonwords, which varies in wordlikeness ratings, when presented in a random order accurately identifies and correctly differentiates children with DLD from TD controls the child is bilingual or monolingual.

Keywords: developmental language disorder (DLD); specific language impairment (SLI); nonword repetition; diagnostic markers of DLD/SLI; likelihood ratio; Catalan; European Portuguese

Citation: Ahufinger, N.; Berglund-Barraza, A.; Cruz-Santos, A.; Ferinu, L.; Andreu, L.; Sanz-Torrent, M.; Evans, J.L. Consistency of a Nonword Repetition Task to Discriminate Children with and without Developmental Language Disorder in Catalan–Spanish and European Portuguese Speaking Children. *Children* **2021**, *8*, 85. https://doi.org/10.3390/children8020085

Academic Editors: Eva Aguilar Mediavilla, Miguel Pérez Pereira, Elisabet Serrat-Sellabona and Daniel Adrover-Roig
Received: 15 December 2020
Accepted: 21 January 2021
Published: 26 January 2021

1. Introduction

The ability to repeat nonwords has been associated with early childhood vocabulary knowledge and development, grammar skills, and utterance lengths [1]. Additionally, differences of nonword repetition ability between clinical populations and their age-matched typically developing peers (TD) has been suggested as a useful tool to explore language differences among clinical populations (e.g., developmental language disorder (DLD) [2,3], dyslexia [4], Down's syndrome [5], etc.). In the case of DLD, nonword repetition tasks (NRT) have been investigated as a method to differentiate TD children from children with DLD, but the results from these studies have been inconclusive. The purpose of this study was to examine the diagnostic consistency of NRT to discriminate between monolingual and/or bilingual children with DLD from TD controls when developed in the native languages of children. More specifically, we compared the precision of language-specific NRT to identify DLD in a group of simultaneous bilingual Catalan–Spanish speaking children with and without DLD to that of a group of monolingual European Portuguese children with and without DLD.

Developmental language disorder (DLD), historically referred to as specific language impairment (SLI), is a neurodevelopmental disorder characterized by normal nonverbal intelligence paired with a persistent inability to master comprehension and production of language in the absence of intellectual or emotional disability, hearing loss, or other medical conditions or syndromes known to cause language disorders [6–8]. Historically, the focus of much of the research in DLD has been on determining the degree to which the deficits are specific to the language system or extend to nonlinguistic aspects of cognition. Some researchers historically used the term SLI to refer to children with deficits that were believed to be specific to the "language" system [9], while other researchers used a broader interpretation of the term to denote the presence of both language-based deficits as well as weaknesses in areas that go beyond language [6–8]. Consistent with the central tenants of a broad characterization of the disorder, the research community recently shifted to use the term DLD to refer to those children who fall into the broader definition of SLI [10–12]. In keeping with this trend, we use the term DLD in this manuscript to refer to this more broadly defined group of children with language-based deficits. Adding support to these domain-general theories of DLD, nonword repetition performance has been consistently shown to be lower in children with DLD than their TD peers [13]. The ability to repeat nonwords relies on numerous underlying factors, such as phonological working memory [3,14,15], auditory processing [16,17], and speech production [18]. Understanding the nature of the underlying cause of poor nonword repetition DLD is seen as a necessary missing piece to understand the underlying cause of language deficits in these children.

Arguments for a phonological working memory basis for the nonword repetition deficits in DLD are founded on Baddeley's theory of working memory, and more specifically, on the phonological loop portion of his theory [19,20]. The phonological loop's storage component is particularly implicated in DLD [21,22]. The phonological store is a short-term memory trace of the incoming phonological information that fades over a matter of seconds. It is argued that individuals with DLD have smaller storage capacity in the phonological loop than their TD peers. By using a fast-mapping paradigm, Alt [23] found that children with DLD had difficulty with fast-mapping words that were greater than two syllables long, which she attributed to initial encoding difficulties and decreased storage capacity of the phonological loop. Further evidence for the limited storage capacity comes from studies on nonword repetition in children with DLD, where children with DLD perform relatively similar to their TD peers on words with two syllables, and they perform progressively worse than their peers as the nonwords become longer [24]. These researchers have argued that as the nonword length increases, the amount of storage required to maintain these nonwords in working memory also increases. Thus, the decrease in performance on nonwords of increasing length in DLD is attributed to a lack of adequate phonological storage.

Syllable length is not the only factor shown to influence nonword repetition ability in both DLD and TD populations. Syllabic phonological frequency as well as wordlikeness have both been shown to impact nonword repetition [25,26], with higher phonological frequency and/or high wordlikeness resulting in higher nonword repetition accuracy. These syllabic and word-level influences on nonword repetition performance are unsurprising from Baddeley's working memory framework. In the model, working memory interacts with long-term, extant knowledge to the effect that a higher frequency nonword is hypothesized to have stronger long-term associative links at the phonological level (in the case of high phonological syllable frequency), or higher word-association links (for high wordlikeness), resulting in greater accuracy for these nonwords due to their decreased demands on memory. Frequency effects such as this have been observed at the word level, with higher frequency words identified quicker during lexical decision tasks as compared to low-frequency words [27–30].

One aspect of nonword repetition tasks that makes them highly useful for research and diagnostic purposes is that unlike other measures such as sentence comprehension and receptive and expressive vocabulary, NRT does not rely on the underlying structure of the language the task is being administered in, and as such, NRT provides a measure of language ability that is independent of environmental factors such as race and maternal education level [24,31], but one that is not biased by the morphosyntactic structure of the language [32–34]. Due to this, in recent years, a multitude of studies have investigated the efficacy of employing NRT to differentiate children with DLD from their TD peers across a wide range of languages including Spanish [32], Norwegian [35], Swedish [36,37], Vietnamese [38], Gulf Arabic [39], Italian [40], Icelandic [41], and Brazilian Portuguese [42]. However, it is important to note that although nonword repetition performance is not influenced by the syntax of a language, it is influenced by the lexicon. Specifically, NRT are sensitive to the underlying phonological, phonetic, syllable frequency, and the degree to which the nonword overlaps with real worlds in the lexicon. Taken together, this suggests that clearly monolingual children should be tested on NRT derived from their native language but also that bilingual children should be tested on NRT derived from all of their native language(s) [40,43,44].

The common deficits in nonword repetition performance in DLD and its cross-linguistic potential have motivated a great deal of research examining the degree to which NRT stimuli can be used to discriminate children with DLD from TD. However, there is large variability in these studies, both with respect to the degree to which performance is poorer for DLD as compared to TD controls *and* in the positive likelihood values reported across studies. Positive likelihood ratios (LH) give the percentage possibility of an individual being a part of a group given a score on a task or test [45], and the larger the LH value, the greater the likelihood the individual is a part of the group. Studies calculating LH values for nonword repetition accuracy report a range of values including small to moderate (LH = 2.78 at or below 70% accuracy [46]; LH = 6.67 at or below 50% accuracy [47]) to moderate-high (LH = 11 at or below 50% accuracy [48]).

A potential factor contributing to the variability across different NRT tasks to discriminate children with DLD from TD children may be differences in the NRT themselves. For example, Ellis Weismer et al. [46] reported data from the Dollaghan and Campbell [49] NRT, which is a version of NRT that consists of a set of 16 nonwords, four at each of four-syllable lengths (1-, 2-, 3-, 4-) that were controlled to all have low wordlikeness ratings. Further, the nonwords are presented in a sequential order beginning with the one-syllable nonwords and progressing to the four-syllable nonwords. In contrast, Girbau and Schwartz [32] examined NRT performance in Spanish–English bilingual children with and without DLD using a NRT task that consisted of 20 nonwords, four at each of five syllable lengths (1-, 2-, 3-, 4-, 5-).

To examine the consistency of NRT as a measure to identify and discriminate children with DLD from children with normal language across languages, in the current study, we compared children's performance on a NRT task in two ways. First, because there were no NRT in Catalan or European Portuguese, we used the same method to develop two NRT to compare the performance of a group of bilingual Catalan–Spanish speaking children with and without DLD to that of a group of monolingual European Portuguese speaking children. Since the Catalan–Spanish cohort was bilingual, we also compared their performance on a previously developed and published Spanish NRT task.

First, we ask if bilingual versus monolingual status matters when using an NRT task developed from scratch that matches a child's native language in its ability to differentiate children with DLD from their TD peers who are either simultaneous bilingual (e.g., Catalan–Spanish) or monolingual (e.g., European Portuguese). In line with the current research, we hypothesize that nonword repetition ability will have a high likelihood of predicting which children have DLD and which do not in both language populations. Second, we ask if the syllable length of the nonwords improves NRT's ability to identify children with DLD over total percent correct. Specifically, we examined whether the likelihood ratio values can accurately identify and differentiate DLD from children having typical language in an individual child. In keeping with the current literature, we hypothesize that because children with DLD may perform similar to their TD peers on shorter nonwords (two syllables), but because their accuracy may decline beyond two syllables, longer nonwords may have better diagnostic accuracy as compared to two-syllable nonwords. Third, we asked to what extent does the degree to which the nonword overlaps with real words in structure (i.e., wordlikeness) influence diagnostic accuracy. Since vocabulary has been shown to influence nonword repetition performance, and previous work consistently shows that performance is poorer for children with DLD on nonwords having low as compared to high wordlikeness, we predict that the wordlikeness of nonwords also will increase the diagnostic accuracy of NRT. Our working hypothesis is that a similar pattern of performance on NRT created in two different languages, using the same method to create the nonword lists, should result in similar LH ratios across the two linguistically different samples, and it would provide valuable data regarding the feasibility of NRT as a diagnostic marker of DLD.

2. Materials and Methods

2.1. Participants

2.1.1. Catalan–Spanish (CS)

The Catalan–Spanish (CS) cohort consisted of a total of 72 children: 36 children with developmental language disorder (DLD-CS) aged 5;4–15;5 and 36 typically developing children (TD-CS) with normal language aged 5;7–16;2 (A portion of the children of the Catalan–Spanish cohort were included in Ahufinger et al. [50]). There were a total of 20 girls (28%) and 52 boys (72%). The children with DLD-CS were pairwise matched to TD-CS controls based on age (+/−3 months) and sex at the time of the study. All the children were native simultaneous bilingual speakers of Catalan and Spanish that were exposed to both languages from birth. (Catalan is one of the 5 co-official languages that coexist in Spain. It is only spoken in a region called *Països Catalans*/Catalan countries formed by Catalonia, Valencia, Balearic Islands, and Northern Catalonia). According to the parental survey, all families of the present study reported that their children speak Catalan and Spanish. All children that participated in the present study were from different areas of Barcelona and the surrounding area (metropolitan area of Barcelona) where Central Catalan is spoken. In 2018, 52.7% of Barcelona citizens claimed that their initial language was Spanish, and 31.5% was Catalan [51]. In the school system, Catalan is the primary language of instruction. Later, in primary school, children receive 2–3 h of Spanish classes per week, while the rest of the subjects are taught in Catalan [52,53]. According to Alarcón and Garzón [54], children in Barcelona are equally proficient in both Spanish and Catalan, although the use of Spanish is more popular. For further information about Catalan and Spanish bilingualism and DLD, see Sanz-Torrent, Badia, and Serra [55] and Sanz-Torrent et al. [56]. As such, the Catalonian children in this study are considered simultaneously bilingual. The children with DLD-CS were selected with the collaboration of different institutions, organizations, and schools as Catalan Center of Resources for Hearing-Impaired People (CREDA), members of the Catalan service for school counseling and guidance (EAP) and Catalan Association of Specific Language Impairment (ATELCA). All the families who agreed to participate in the study were asked to sign an informed consent form. A final report containing the results of all the tests administered to the children was given to the family as a token of gratitude for their commitment and contribution to the study.

Catalan–Spanish Inclusion/Exclusion Criteria

The inclusion/exclusion criteria were defined following the DLD diagnostic criteria recommended by a Spanish expert committee that reached a consensus in 2015 [57]. All the children identified were assessed by two trained researchers to confirm the diagnostic criteria. The diagnostic inclusion/exclusion criteria for children with DLD-CS in the present study were as follows: (a) Catalan–Spanish Bilingual; (b) a nonverbal intellectual quotient

(NVIQ) > 75 (Kaufmann Brief Intelligence Test Matrices section, K-BIT [58]); (c) a score of –1.25 SD or greater on one of the three scales of the Clinical Evaluation of Language Fundamentals—Fourth Edition, Spanish (CELF-4; [59]): core language, expressive language, and receptive language; (d) normal hearing at 500, 1000, 2000, and 4000 Hz at 20 dB based on the American Speech-Language Hearing Association (ASHA) 1997 guidelines for hearing screening [60]; (e) normal or corrected-to-normal vision; (f) normal oral and speech motor abilities; and (g) absence of other medical or neurological conditions. The inclusion/exclusion criteria for the group of TD children were as follows: (a) Catalan–Spanish Bilingual; (b) a NVIQ > 85 (K-BIT Mat [58]); (c) standardized language scores within the normal range in the core language, expressive language, and receptive language CELF-4 scales; and (d) absence of a prior history of speech or psychological therapy (see Table 1). All children were classified as either TD or DLD based on standard clinical practice in Catalonia, which is based on the Spanish version of the standardized test where all stimuli are presented in Spanish, but if children answered correctly in Catalan, they were given credit for their answer. There currently are no normative language tests for the Catalan language. From 79 children with language difficulties that were initially screened, 43 were excluded for failure to meet the inclusion/exclusion criteria.

Table 1. Age and standardized scores for language and cognitive assessment measures for Catalan–Spanish children with developmental language disorder (DLD-CS) and typically developing (TD-CS) children.

	DLD-CS (n = 36)			**TD-CS (n = 36)**			**Comparison**	
Variable	**Mean**	**SD**	**Range**	**Mean**	**SD**	**Range**	**t(70)**	**p**
Age in months	107.72	28.76	64–185	111.58	28.68	67–194	−0.57	$p = 0.57$
K-BIT mat (NVIQ) [a]	98.67	12.10	76–119	102.53	8.82	88–124	−1.54	$p = 0.13$
CELF-CLS [b]	73.78	10.98	45–89	109.83	6.57	99–130	−16.91	$p < 0.01$
CELF-ELS [c]	73.86	8.69	52–87	109.39	7.73	89–128	−18.33	$p < 0.01$
CELF-RLS [d]	78.92	9.74	59–97	105.44	5.89	93–118	−13.98	$p < 0.01$

Note. For each variable, age-scaled scores have a mean of 100 and an SD of 15 (except age in months). [a] K-BIT mat = Kaufman Brief Intelligence; Matrices subtest, Spanish version [58]. NVIQ: Non-verbal intelligence. [b] CELF-4 CLS = Spanish Clinical Evaluation of Language Fundamentals, 4th edition: Core Language score [59]. [c] CELF-4 ELS = Spanish Clinical Evaluation of Language Fundamentals, Fourth Edition: Expressive Language score [59]. [d] CELF-4 RLS = Spanish Clinical Evaluation of Language Fundamentals, Fourth Edition: Receptive Language score [59].

2.1.2. European Portuguese (EP)

The European Portuguese cohort consisted of a total of 150 children; 75 children with developmental language disorder (DLD-EP) aged 7;0–11;11, and 75 European Portuguese speaking typically developing (TD-EP) children with normal language aged 7;0–11;11. There were a total of 42 girls (28%) and 108 boys (72%). The children with DLD-EP were pairwise matched to typical controls based on age (+/−6 months) and sex from the same classroom and school grade.

Prior to participant recruitment, permission to perform this study in the state-funded schools in the North of Portugal was requested from the Board of Education—North Division (DREN, Direcção Regional de Educação do Norte). First, the experimenter visited each of the 17 schools and presented the aims of the study to the school administrator, who gave his formal consent for the experimenter to contact each classroom teacher in order to gather the required information. Then, classroom regular and special education teachers were asked to identify children in the 7-, 8-, 9-, 10-, and 11-year-old age groups who had language problems (i.e., oral comprehension, oral production, reading problems, language delay, etc.) that might be potential DLD participants. All this identification process was based on existing assessments (e.g., portfolios, checklists, inventories) at the school at the time of the study. Once a list of children with potential DLD-EP was obtained, a written consent form for soliciting participation in this study was forwarded to the parents of the children in the target age ranges. Once the participants with DLD-EP in a given school were identified and parental consent was returned to the experimenter; then, typical developed controls were identified using the same procedures.

European Portuguese Inclusion/Exclusion Criteria

The inclusion/exclusion criteria were defined following the Portuguese National criteria. The diagnostic inclusion/exclusion criteria for children with DLD-EP in the present study were (a) Monolingual European Portuguese speaking—the only spoken language in the home setting; (b) a nonverbal intellectual quotient (NVIQ) > 80 assessed by school psychologists with the Wechsler test standardized for the European Portuguese population (WISC-III; [61]); (c) language impairment; (d) normal hearing; (e) normal or corrected-to-normal vision; (f) normal oral and speech motor abilities; and (g) absence of other medical or neurological conditions. Criteria a, c, d, e, f, and g were based on existing classroom observations, interviews, informal developmental checklists, portfolios, school reports, medical/clinical reports, and formal academic assessments provided by parents and teachers. At the time of data collection, there were no standardized language assessment measures in European Portuguese. The inclusion/exclusion criteria for the group of TD children were as follows: (a) Monolingual European Portuguese speaking—the only spoken language in the home setting; (b) a nonverbal intellectual quotient (NVIQ) > 80 assessed by school psychologists with WISC-III [61]; (c) normal language abilities; and (d) absence of a prior history of speech or psychological therapy. Criteria a, c, and d were based on interviews, portfolios, school reports, medical/clinical reports, and formal academic assessments provided by parents and teachers. The children all attended state-funded primary and intermediate grades (2nd to 5th grade) in schools located in a range of urban, suburban, and rural areas in two districts in the North of Portugal. A total of 179 children were initially identified for the study; 29 were subsequently excluded for failure to meet the inclusion/exclusion criteria and/or low NVIQ.

2.2. Nonword Repetition Task

2.2.1. Catalan Nonword Repetition Task (NRT-Cat)

Since there was no NRT task in Catalan (NRT-Cat), we developed one for the purpose of the current study. Based on a Graf Estes, Evans, and Else-Quest [3] meta-analysis showing that the Children's Test of Nonword Repetition (CNRep; [62]) has larger effect sizes in detecting DLD, we designed the task to mirror that of the CNRep ([62]) with respect to the total number of nonwords and distribution of high- and low wordlikeness of the nonwords. Additionally, the NRT-Cat was developed taking into account the original criteria established by Gathercole et al. [62]: "minimal articulatory output demands", "dominant syllable stress patterns", and "phoneme sequences within each non-word were all phonotactically and prosodically legal" (p. 106) for the English language.

Syllable Length

To develop the NRT-Cat, we first generated lists of words selected from a large corpus of children's songs and books as well as from spontaneous language samples. Verbs, diphthongs, and words with orthographic stress were not excluded from the list. We generated a list of 100 Catalan nonwords with different syllable length (from 2 to 6 syllables; each syllable length had 20 nonwords) taking into account characteristics of Catalan language such as the number of syllables, complexity, and prosody. The syllable stress pattern for Catalan was analyzed for development of the nonwords. It was taken into account that for the Catalan language, almost all of the words that finish in a consonant have the stressed pattern in the last syllable (i.e., per<u>fil</u>/capit<u>al</u>); multi-syllabic words that finish with vowels usually have the stressed pattern in the penultimate syllable (i.e., di<u>vo</u>rci/<u>ca</u>psa). All the words in Catalan that have the stressed syllable in the antepenultimate position have orthographic accent (e.g., <u>mú</u>sica), but these words were not included.

All of the Catalan nonwords began with consonants that occur regularly at the word initial position in Catalan (see Table A1); nonwords could finish with a consonant or vowel; nonwords contained Catalan regular stress patterns (non-orthographic stress patterns as described in the previous paragraph), diphthongs, and orthographic stress patterns were not included in the nonwords. Similar to the CNRep's original criteria [62], the phoneme sequences in each nonword conform to the phonotactic rules of Catalan. Additionally, similar to the methodological criteria outlined in Gathercole et al. [62], consonants that occur naturally in Catalan and double consonants clusters naturally occurring in Catalan were included to build the Catalan nonwords (e.g., *Galmet*, *Becra*, *Marlut*, *Palimatrenc*, *tropa*, etc.).

Wordlikeness

The degree to which a nonword "sounds like" a real word in a language has been shown to influence nonword repetition ability in children with and without DLD [3]. Moreover, it has been argued that "phonotactic knowledge of the possible/probable sequences of sounds within a language might be derived directly from the mental lexicon, depending on their similarity to known words" ([63] p. 568). We used the same procedure as Gathercole et al. [62] to determine the wordlikeness of each of the nonwords. A native female speaker of the Central Catalan dialect was recorded to create a digital recording of the list of 100 nonwords in a randomized order in 5 sets of 20 nonwords. Each set had 10 single words (CVC) and 10 consonant cluster words (CCV/CCVC). These words were presented to a group of 20 native Catalan-speaking adults (male and female). They were asked to rate how "wordlike" each of the nonwords auditory presented was according to the similarity they have with Catalan words. They had to do it trying to answer the following question: Would it be possible that this word is in the Catalan language? We used a scale that ranged from 1 ("It has very few possibilities") to 5 ("It has excellent possibilities"). The adults rated each nonword immediately upon hearing it on a scoring sheet. Then, the data were analyzed to calculate means of the 100 nonwords. Words were listed in rank order for each syllable length, with nonwords receiving the highest wordlikeness ratings ranked at the top and nonwords with the lowest wordlikeness ranked at the bottom. Next, the four nonwords from the bottom of the ranking that received the lowest mean wordlikeness ratings were chosen for each of 2-, 3-, 4-, 5-, and 6-syllable lengths, creating a list of 20 low wordlikeness nonwords. Likewise, the four nonwords at the top of the ranking that received the highest mean wordlikeness ratings for each of the 2- 3-, 4-, 5-, and 6-syllable lengths were selected, creating a list of 20 high wordlikeness nonwords. For the final list for each syllable length, four of the nonwords contained single consonants and four contained one or more consonant clusters. All the words had the same number of phonemes at each of the syllable lengths. A complete list of the 40 nonwords by high and low Wordlikeness for 2-, 3-, 4-, 5-, and 6-syllable lengths is shown in Table 2. Their phonetic transcriptions in International Phonetic Transcription (IPA) are shown in Table A1.

Table 2. The nonwords for the Catalan version of the nonword repetition task (NRT-Cat) by syllable length and wordlikeness.

2-Syllable	3-Syllable	4-Syllable	5-Syllable	6-Syllable
		High Wordlikeness Nonwords		
marlut	grimpola	palimatrenc	balatenesta	pomalimenasa
calim	llendira	xipileta	jolistarista	desperdiculandre
padot	vifatull	calisota	xepunetura	parmicaginosa
galmet	dubalet	clomitura	parantalaresc	situcalinomi
		Low Wordlikeness Nonwords		
pidop	satompa	becuradoc	lamerquitorma	diraculmestici
rolma	xolopi	daltrosqueti	lifortamasuc	purmesidocata
becra	calmepi	castretuma	lemulticada	casitedilafa
dapa	dalumi	milusota	zicuparamal	garelisupota

2.2.2. Spanish Nonword Repetition Task (NRT-Span)

The version of the Spanish nonword repetition task used in this study was a prior published version developed by Aguado [64]), which consists of a set of 80 nonwords that range from 2 to 5 syllables in length. The nonwords were also classified according to their overall frequency calculated at the syllable level (high frequency, low frequency) based on the database of Alameda and Cuetos [65]. The final NRT stimuli were divided into two lists (high frequency or low frequency) with each list having four groups of ten nonwords ranging from 2 to 5 syllables in length (10 words at each length in each list). Both nonword lists are controlled to have an equal number of syllables of nonwords, syllable structure, stress pattern, and order in which the syllables with their different structures were placed. Two phoneme consonant clusters were included. The task is based on children listening to a series of nonwords one by one, temporarily retaining the phonological information of these and then producing them verbally.

2.2.3. European Portuguese Nonword Repetition Task (NRT-EPort)
Syllable Length

Similar to Catalan, because there was no European Portuguese NRT, we developed one for the purpose of the current study using the same steps in developing the stimuli that we used for the Catalan NRT. For the NRT-EPort, we first generated lists of words selected from a large corpus of European Portuguese derived from interview-based corpora [66,67] and children's songs, children's books, grammar books, and European Portuguese language development studies [68–73]. Then, these real words were classified according to the number of syllables, articulatory complexity, and prosody. Verbs, diphthongs, and words with orthographic stress were not excluded from the list. From the remaining set of words, a set of 100 nonwords were generated from these real words that differed in syllable length (from 2 to 6 syllables) and contained naturally occurring European Portuguese stress patterns [67,69]. Syllable stress patterns in European Portuguese were analyzed for development of the nonwords. The stress is always on the second to last syllable regardless of syllable length. Some examples are 2-syllable words (bola/ball); 3-syllable words (panela/pot); 4-syllable words (fortaleza/fortress); 5-syllable words (psicologia/psychology); 6-syllable words (corajosamente/bravely).

The result was a set of 2-, 3-, 4-, 5-, and 6-syllable nonwords (20 per syllable length). All nonwords begin with consonants, and the consonants /r/, /l/, /ʃ/, /m/, and the vowels /ɔ/, /u/, /ɐ̃/, /ɐ/, /w/, /i/ never occurred in the final position of the nonwords. Similar to the CNRep's original criteria [62], two phoneme consonant clusters that occur within European Portuguese were included (e.g., *branco*, *plasma*, *grito*, *tropai*, etc.).

Wordlikeness

The same procedure as the Catalan version of the NRT was followed to derived wordlikeness rating for the nonwords in European Portuguese NRT. For each syllable length, four of the nonwords contained single consonants, and four contained one or more consonant clusters. All the words had the same number of phonemes for each of the syllable lengths. The nonwords had no plurals or Portuguese complex morpheme endings (e.g., mente). A complete list of the 40 European Portuguese nonwords, by high and low wordlikeness for 2-, 3-, 4-, five -, and 6-syllable lengths, is shown in Table 3. Their phonetic transcriptions in International Phonetic Transcription (IPA) are shown in Table A2.

2.3. Administration and Scoring of NRT-Cat, NRT-Spa, and NRT-EPort

For the Catalan and European Portuguese versions of the NRT, there was a fixed randomized order across all the nonwords such that the listener could not predict the syllable length and wordlikeness. For the Spanish version of the task, the 40 nonwords of each list (high nonword frequency and low nonword frequency) were presented sequentially in accordance to syllable length (i.e., all 2-syllable nonwords followed by all 3-syllable nonwords, etc.). Half of the sample was administered first in the high-frequency list and then the low-frequency list, and with the other half, the order was inverted.

Table 3. The nonwords for the European Portuguese version of the nonword repetition task (NRT-EPort) by syllable length and wordlikeness.

2-Syllable	3-Syllable	4-Syllable	5-Syllable	6-Syllable
		High Wordlikeness Nonwords		
naca	lofena	covilado	melanifito	turamisalato
fopa	banita	fenerade	bonifadade	detagapalico
trana	praleta	trapilado	craletonina	prinalvenioso
prota	bramato	cravastado	versatranista	volturacidade
		Low Wordlikeness Nonwords		
cafo	mafopa	lemanado	nocafozano	rolinicistato
tuma	dopeta	dilomopa	lodanapito	fatatuviricho
grapa	gremata	dragamato	defermicato	satopogatico
trila	tramafa	trafeleste	promoflicada	cremoforosada

To control for possible presentation effects, a digital recording of each of the NRTs was created by an adult female native speaker for each language. The female adult said each nonword aloud at the rate of one nonword every three seconds. The child's task was to repeat each nonword immediately. For each NRT task, children heard two practice items (e.g., *ticopo* and *mastruca* for Catalan and Spanish, and *banata* and *mencolate* for European Portuguese), and these were repeated until the child understood the task completely. Each child heard the following instructions as in other studies [62]:

> *"I would like you to play a game with me! When I switch on the disc in a minute (point to the recorder), you will hear a funny made-up word, a word that does not exist. I would like you to repeat the funny word back to me as soon as you have heard it. Did you understand? So, if the made-up word you heard was banata, you should say banata back to me. Let's try that now, ok?".*
>
> Catalan: *"M'agradaria molt que juguéssim junts/es! Ara quan encengui l'altaveu escoltaràs unes paraules divertides que estan inventades, són paraules que no existeixen. El que has de fer és repetir la cada paraula divertida just després d'escoltar-la. Ho has entès? Per exemple, si la paraula inventada que escoltes fos plàtan, tu has de dir plàtan. Anem a intentar-ho, d'acord?"*
>
> Spanish: *"¡Me gustaría mucho que jugáramos juntos/as! Ahora cuando encienda el altavoz escucharás unas palabras divertidas que están inventadas, son palabras que no existen. Lo que tienes que hacer es repetir cada palabra divertida justo después de escucharla. ¿Lo has entendido? Por ejemplo, si la palabra inventada que escuchas fuera plátano, tú debes decir plátano. Vamos a intentarlo, ¿de acuerdo?"*
>
> Portuguese: *"Vamos jogar um jogo! Quando eu ligar o gravador vais ouvir uma palavra inventada, que não existe, mas muito engraçada. Quero que repitas a palavra tal como a ouves, está bem? Vamos experimentar agora. Quando ouvires a palavra banata vais dizer banata tal e qual como a ouviste, está bem?"*

All the children in both of the language cohorts were able to perform the task following the practice items.

2.4. Scoring of the NRT-Cat, NRT-Spa, and NRT-EPort

For each NRT task, an adult experimenter scored the child's repetition attempt as it was produced, using a simple binary scoring procedure in which a correct repetition was scored as 1 (judged by the experimenter to be phonologically accurate), and an incorrect attempt was scored as 0 if the experimenter judged that the child had produced a sound that differed from the target nonword by one or more phonemes. Children's responses were also recorded and subsequently rescored from the recording a second time to ensure the accuracy of scoring. Correct and incorrect responses were annotated on the answer sheet by the experimenter. The total number of nonwords spoken correctly was calculated for each child.

2.5. Reliability: Scoring of the NRT-Cat, NRT-Spa, and NRT-EPort

For the Catalan NRT (NRT-Cat), a total of 28% (n = 20) of the children's responses were randomly identified from the sample to test reliability (DLD-CS = 10; TD-CS = 10). This reanalysis was performed by two trained independent Catalan–Spanish researchers. The scoring–rescoring agreement was 92.75% for the NRT-Cat Test for children with DLD-CS and was 97.5% for the TD-CS controls.

For the Spanish NRT (NRT-Spa), a total of 28.5% (n = 20) of the children's responses were randomly identified from the sample to test the reliability (DLD-CS = 10; TD-CS = 10). This reanalysis was performed by two trained independent Catalan–Spanish researchers. Scoring–rescoring agreement was 94.6% for the NRT-Cat Test for children with DLD-CS and 96.9% for the TD-CS controls.

For the European Portuguese NRT (NRT-Port), a total of 20% (n = 30) of the children's responses were randomly identified from the sample to test reliability (DLD-EP = 18; TD-EP = 12). This reanalysis was performed by two trained independent European Portuguese graduate adults. The scoring–rescoring agreement was 87.5% for the NRT-EPort Test for children with DLD-EP, and it was 92.5% for the TD-EP controls.

3. Results

3.1. Group Differences on the NRT

3.1.1. Syllable Length

The means and standard deviations on the NRT (NRT-Cat) for total percentage of words produced correctly (TPWC) for each syllable length are presented in Table 4 for the Catalan–Spanish and European Portuguese cohorts.

Table 4. Means and standard deviations on the NRT task for percentage of words correct (PWC) overall and for each syllable length for the Catalan and Spanish versions of the NRT for the Catalan–Spanish (CS) and European Portuguese (EP) children with developmental language disorder (DLD) and typically developing (TD) controls.

			Nonword Length			
Group	**2-syll PWC**	**3-syll PWC**	**4-syll PWC**	**5-syll PWC**	**6-syll PWC**	**Total PWC**
				CS-Catalan		
DLD [a] (n = 36)	70.8 (21.7)	68.4 (21.2)	56.9 (23.9)	31.5 (25.6)	12.1 (18.0)	48.0 (18.0)
TD [b] (n = 36)	97.9 (5.5)	93.4 (10.1)	92.0 (10.4)	84.0 (12.1)	73.6 (24.9)	88.0 (8.0)
				CS-Spanish		
DLD [a] (n = 34)	82.2 (14.3)	68.5 (21.7)	45.0 (21.8)	25.2 (21.3)	-	54.5 (16.9)
TD [b] (n = 36)	95.2 (5.5)	92.3 (7.5)	86.6 (11.2)	80.9 (15.2)	-	89.0 (7.3)
				European Portuguese		
DLD [a] (n = 75)	77.1 (15.5)	75.5 (13.5)	76.6 (16.2)	77.1 (20.7)	48.8 (17.2)	71.0 (12.2)
TD [b] (n = 75)	95.6 (6.9)	92.6 (9.2)	96.3 (7.3)	95.3 (7.6)	87.0 (13.2)	93.4 (5.5)

[a] Developmental Language Disorder. [b] Typically Developing. Note. The Spanish NRT version does not include nonwords of 6-syllable length.

Catalan–Spanish Cohort

Looking first at the Catalan version of the NRT, a univariate analysis of variance, controlling for age (ANCOVA) was first conducted comparing the overall nonword repetition performance for the Catalan nonwords for the DLD-CS and TD-CS groups. The analysis revealed a significant main effect of group $F(1, 71) = 150.86$, $p < 0.001$, partial $\eta^2 = 0.68$, power = 1.0, where the DLD-CS group's performance was significantly poorer than that of the TD-CS controls. A follow-up 2 × 5 Repeated Measures ANCOVA was conducted with group (DLD-CS, TD-CS) × syllable length (2-, 3-, 4-, 5-, 6-syllable) as the between and within subject variables, and age as the covariate, to examine differences in the groups' performance for each syllable length. Mauchly's test indicated that the assumption of sphericity was violated; therefore, degrees of freedom were corrected using Greenhouse–Geisser estimates of sphericity [74]. The analysis revealed a significant main effect of group $F(1, 69) = 152.78$, $p < 0.001$, partial $\eta^2 = 0.68$, power = 1.00; and significant main effect for syllable length, $F(3.36, 231.89) = 13.29$, $p < 001$, partial $\eta^2 = 0.16$, power = 1.00, and a significant length × group interaction $F(3.36, 231.89) = 23.65$, $p < 0.001$, partial $\eta^2 = 0.25$, power = 1.00. Follow-up analyses of covariance (ANCOVA) were conducted for each of the syllable length with age as a covariate to examine group differences at each of the six syllable lengths. The analysis revealed that the TD-CS performance was significantly better than that of the children with DLD-CS for all syllable lengths: 2-syllable $F(1, 71) = 51.32$, $p < 0.001$, partial $\eta^2 = 0.42$, power = 1.0; 3-syllable $F(1, 71) = 40.13$, $p < 0.001$, partial $\eta^2 = 0.36$, power = 1.0; 4-syllable, $F(1, 71) = 65.30$, $p < 0.001$, partial $\eta^2 = 0.48$, power = 1.0; 5-syllable, $F(1, 71) = 121.47$, $p < 0.001$, partial $\eta^2 = 0.63$, power = 1.0; 6-syllable, $F(1, 71) = 150.06$, $p < 0.001$, partial $\eta^2 = 0.68$, power = 1.0. The DLD-CS group performed worse than the TD-CS group at repeating each nonword across the syllable lengths (2-, 3-, 4-, 5-, 6- syllable). This finding illustrates that a higher percentage of correct nonwords are achieved at the lowest lengths, decreasing to a lower percentage of correct nonwords as the syllable length also increases (from 2 syllables to 6 length) for both groups.

For the Spanish version of the NRT (NRT-Spa), a univariate analysis of variance, controlling for age (ANCOVA) was first conducted comparing overall nonword repetition performance for the Spanish nonwords for the DLD-CS and TD-CS groups (Two of the DLD-CS participants did not complete the Spanish NRT). The analysis revealed a significant main effect of group $F(1, 69) = 130.8$, $p < 0.001$, partial $\eta^2 = 0.66$, power = 1.0, where the DLD-CS group's performance in repeating the Spanish nonwords was significantly poorer than that of the TD-CS controls. A follow-up 2 × 5 Repeated Measures ANCOVA was conducted with group (DLD, TD) × syllable length (2-, 3-, 4-, 5-syllable) as the between and within variables, and age as the covariate, to examine differences in the groups' performance for each syllable length. Mauchly's test indicated that the assumption of sphericity was violated; therefore, degrees of freedom were corrected using Greenhouse–Geisser estimates of sphericity [74]. The analysis revealed a significant main effect of group $F(1, 67) = 127.4$, $p < 0.001$, partial $\eta^2 = 0.65$, power = 1.00; and significant main effect for syllable length, $F(2.47, 166.0) = 11.2$, $p < 0.001$, partial $\eta^2 = 0.14$, power = 0.99, and a significant length × group interaction $F(2.47, 166.0) = 58.4$, $p < 0.001$, partial $\eta^2 = 0.46$, power = 1.00. Follow-up analyses of covariance (ANCOVA) were conducted for each of the syllable length, with age as a covariate to examine group differences at each of the six syllable lengths for the NRT-Spa. The analysis revealed that the TD-CS performance was significantly better than that of the children with DLD-CS for all syllable lengths: 2-syllable $F(1,67) = 28.9$, $p < 0.001$, partial $\eta^2 = 0.30$, power = 1.0; 3-syllable $F(1, 67) = 40.93$, $p < 0.001$, partial $\eta^2 = 0.37$, power = 1.0; 4-syllable, $F(1, 67) = 110.9$, $p < 0.001$, partial $\eta^2 = 0.62$, power = 1.0; 5-syllable, $F(1, 67) = 166.3$, $p < 0.001$, partial $\eta^2 = 0.71$, power = 1.0. The DLD-CS group performed worse than the TD-CS group at repeating each Spanish nonword across the syllable lengths (2-, 3-, 4-, 5-syllable). This finding illustrates that a higher percentage of correct nonwords is achieved at the lowest lengths, decreasing to a lower percentage of correct nonwords as the syllable length also increases (from 2-syllable to 5-syllable length) for both groups.

Finally, we compared the performance on the NRT-Cat and NRT-Spa tasks for the DLD-CS to determine if the performance for the DLD-CS cohort differed depending upon which language their phonological working memory was assessed. A Repeated Measures ANCOVA was conducted comparing total percentage correct for the Catalan and Spanish versions of the task (NRT-Cat and NRT-Spa) as the within variables, and age as the covariate, to examine differences in the DLD groups' performance for the NRT performance in Catalan as compared to Spanish. Mauchly's test indicated that the assumption of sphericity was not violated; therefore, degrees of freedom were not corrected using Greenhouse–Geisser estimates of sphericity [74]. The analysis revealed that there was not a significant main effect of language $F(1, 32) = 0.54$, $p = 0.46$, partial $\eta^2 = 0.01$, power = 0.11, indicating that the DLD-CS group's performance in repeating the Spanish nonwords did not differ from their performance repeating the Catalan nonwords. We next compared the performance of the TD-CS group on the NRT-Spa and NRT-Cat tasks. For the TD-CS, Mauchly's test indicated that the assumption of sphericity was not violated, and a Repeated Measures ANCOVA comparing total percentage correct for the NRT-Cat and NRT-Spa for the TD-CS controls revealed no difference in the TD-CS ability to repeat the Catalan and Spanish nonwords $F(1, 34) = 0.29$, $p = 0.59$, partial $\eta^2 = 0.00$, power = 0.08.

European Portuguese

The means and standard deviations for nonword repetition for percentage of words correct (PWC) for the total task and each syllable length are presented in Table 4 for the DLD-EP and TD-EP groups. Univariate analysis of variance, controlling for age (ANCOVA), was used to assess group differences in total percentage of words correct (TPWC) for the DLD-EP and TD-EP groups. The analysis revealed a significant effect of group $F(1, 149) = 232.00$, $p < 0.001$, partial $\eta^2 = 0.61$, power = 1.0, where the DLD-EP group performed significantly worse in repeating the nonwords as compared to the TD-EP controls. A 2 $\times$ 5 Repeated Measures ANCOVA was conducted with group (DLD, TD) $\times$ syllable length (2-, 3-, 4-, 5-, 6-syllable) as the between and within variables, and age as the covariate, to test for group differences in nonword repetition at each syllable length. Mauchly's test indicated that the assumption of sphericity had been violated; therefore, degrees of freedom were corrected using Greenhouse–Geisser estimates of sphericity [74]. The analysis revealed a significant main effect for group $F(1, 147) = 232.00$, $p < 0.001$, partial $\eta^2 = 0.61$, power = 1.0, a significant effect of syllable length, $F(3.74, 550.01) = 5.72$, $p < 001$, partial $\eta^2 = 0.03$, power = 0.97, and a significant length $\times$ group interaction F (3.74, 550.01) = 25.81, $p < 0.001$, partial $\eta^2 = 0.14$, power = 1.00. Follow-up analyses of covariance (ANCOVA) were conducted for each of the syllable lengths, with age as a covariate. Results indicated that the TD-EP's nonword repetition was better at all lengths as compared to the DLD-EP children: 2-syllable $F(1, 149) = 89.51$, $p < 0.001$, partial $\eta^2 = 0.37$, power = 1.0; 3-syllable $F(1, 149) = 86.04$, $p < 0.001$, partial $\eta^2 = 0.36$, power = 1.0; 4-syllable, $F(1, 149) = 94.30$, $p < 0.001$, partial $\eta^2 = 0.39$, power = 1.0; 5-syllable, $F(1, 149) = 58.11$, $p < 0.001$, partial $\eta^2 = 0.28$, power = 1.0; 6-syllable, $F(1, 149) = 243.51$, $p < 0.001$, partial $\eta^2 = 0.61$, power = 1.0. The results indicated that nonword repetition accuracy was poorer for both the DLD-EP as compared to TD-EP controls both for total words correct and at each syllable length.

3.1.2. Wordlikeness/Nonword Frequency

The means and standard deviations for nonword repetition for total percentage of words produced correctly (TPWC) for nonwords having high and low wordlikeness ratings are presented in Table 5 for the Catalan–Spanish and European Portuguese cohorts.

Table 5. Means and standard deviations on the NRT for percentage of words correct (PWC) for high and low wordlikeness for the Catalan and Spanish versions of the NRT for the Catalan–Spanish (CS) and European Portuguese (EP) children with developmental language disorder (DLD) and typically developing (TD) controls.

Group	High Wordlikeness	Low Wordlikeness
CS-Catalan		
DLD [a] (*n* = 36)	52.0 (18.8)	44.0 (19.0)
TD [b] (*n* = 36)	87.6 (9.3)	88.6 (9.4)
CS-Spanish		
DLD [a] (*n* = 34)	62.3 (17.2)	47.9 (19.5)
TD [b] (*n* = 36)	92.6 (6.5)	85.4 (10.2)
European Portuguese		
DLD [a] (*n* = 75)	75.1 (11.9)	67.0 (15.1)
TD [b] (*n* = 75)	94.8 (5.4)	91.9 (7.6)

[a] Developmental Language Disorder. [b] Typically Developing.

Catalan–Spanish Cohort

Looking first at the NRT-Cat, for the DLD-CS, the total percentage of correct nonword repetition of the Catalan nonwords having high wordlikeness ratings was 52.0% (18.8), and for nonwords having low wordlikeness ratings, it was 44.03% (19.0). For the TD controls, the total percentage of correct nonword repetition for nonwords having high wordlikeness ratings was 87.6% (9.3), and for nonwords having low wordlikeness ratings, it was 88.6% (9.4). A 2 $\times$ 2 Repeated Measures ANCOVA was conducted with group (DLD, TD) $\times$ wordlikeness (High, Low) as the between and within variables, and age as the covariate, to test for group differences based on wordlikeness. Mauchly's test indicated that the assumption of sphericity had not been violated; therefore, degrees of freedom were not corrected using Greenhouse–Geisser estimates of sphericity [74]. The analysis revealed a significant main effect for group $F(1, 69) = 149.6$, $p < 0.001$, partial $\eta^2 = 0.68$, power = 1.0, a significant wordlikeness $\times$ group interaction $F(1, 69) = 12.1$, $p < 0.001$, partial $\eta^2 = 0.15$, power = 0.93, but no main effect of wordlikeness, $F(1, 69) = 1.22$, $p = 0.273$, partial $\eta^2 = 0.01$, power = 0.19. The results indicated that the nonword repetition performance based on total words correct was greater for the TD-CS controls as compared to the children with DLD-CS regardless of the wordlikeness ratings of the nonwords. Furthermore, the nonword repetition performance for the children with DLD-CS was significantly poorer as compared to that of the TD-CS controls for nonwords having low wordlikeness as compared to nonwords having high wordlikeness ratings.

Looking next at the NRT-Spa, a 2 $\times$ 2 Repeated Measures ANCOVA was conducted with group (DLD-CS, TD-CS) $\times$ nonword frequency (High, Low) as the between and within variables, and age as the covariate, to test for group differences based on wordlikeness. Mauchly's test indicated that the assumption of sphericity had not been violated; therefore, degrees of freedom were not corrected using Greenhouse–Geisser estimates of sphericity [74]. The analysis revealed a significant main effect for group $F(1, 67) = 130.84$, $p < 0.001$, partial $\eta^2 = 0.54$, power = 1.0, a significant nonword frequency $\times$ group interaction $F(1, 67) = 7.3$, $p < 0.01$, partial $\eta^2 = 0.09$, power = 0.76, and a significant main effect of nonword frequency, $F(1, 69) = 10.29$, $p < 0.01$, partial $\eta^2 = 0.13$, power = 0.88. The results indicated that for the NRT-Spa, nonword repetition performance based on total words correct was greater for the TD-CS controls as compared to the children with DLD-CS regardless of the nonword frequency ratings of the nonwords. In contrast to the NRT-Cat, nonword repetition performance for the Spanish nonwords for the children with DLD-CS was no different from that of the TD-CS controls for nonwords having low nonword frequency as compared to nonwords having high nonword frequency ratings.

European Portuguese

For the DLD-EP, the total percentage of correct nonword repetition for nonwords having high wordlikeness ratings was 75.1% (11.99), and for nonwords having low wordlikeness ratings, it was 67.0% (15.1). For the TD controls, the total percentage of correct nonword repetition for nonwords having high wordlikeness ratings was 94.8% (5.4), and for nonwords having low wordlikeness ratings, it was 91.3% (7.6). A 2 × 2 Repeated Measures ANCOVA was conducted with group (DLD, TD) × wordlikeness (High, Low) as the between and within variables, and age as the covariate, to test for group differences based on wordlikeness. Mauchly's test indicated that the assumption of sphericity had not been violated; therefore, degrees of freedom were not corrected using Greenhouse–Geisser estimates of sphericity [74]. The analysis revealed a significant main effect for group $F(1, 147) = 232.00$, $p < 0.001$, partial $\eta^2 = 0.61$, power = 1.0, a significant wordlikeness × group interaction $F(1, 147) = 10.14$, $p < 0.001$, partial $\eta^2 = 0.06$, power = 0.88, but no effect of wordlikeness, $F(1, 147) = 3.04$, $p = 0.083$, partial $\eta^2 = 0.02$, power = 0.40. Similar to the Catalan–Spanish cohort, the results indicated that nonword repetition accuracy based on total words correct was greater for the TD-EP controls as compared to the children with DLD-EP regardless of the wordlikeness ratings of the nonwords. Furthermore, the nonword repetition accuracy for the children with DLD-EP was significantly poorer as compared to that of the TD-EP controls for nonwords having low wordlikeness as compared to nonwords having high wordlikeness ratings.

Differences between the Catalan–Spanish and European Portuguese Speaking Cohorts

To examine potential differences in performance for the Catalan–Spanish and European Portuguese speaking children, we first directly compared the performance for the DLD-CS children with that of the DLD-EP children first on the NRT-Cat and then NRT-Spa. A univariate analysis of variance was conducted comparing the overall nonword repetition performance for the DLD-CS in NRT-Cat and DLD-EP groups. The analysis revealed a significant main effect of group $F(1, 110) = 62.5$, $p < 0.001$, partial $\eta^2 = 0.36$, power = 1.0, where the DLD-CS group's performance was significantly poorer than that of the EP-DLD group. To determine if the same pattern held for the CS cohort in Spanish, a second univariate analysis of variance was conducted comparing the overall nonword repetition performance on the NRT-Spa for the DLD-CS and DLD-EP group. The analysis again revealed a significant main effect of group $F(1, 108) = 30.3$, $p < 0.001$, partial $\eta^2 = 0.22$, power = 1.0, where the DLD-CS group's performance on the NRT-Spa was also significantly poorer than that of the DLD-EP group.

We next compared the performance for the TD-CS children with that of the TD-EP children first on the NRT-Cat and then NRT-Spa. A univariate analysis of variance was conducted comparing overall nonword repetition performance for the TD-CS in NRT-Cat and TD-EP group. The analysis revealed a significant main effect of group $F(1, 110) = 16.0$, $p < 0.001$, partial $\eta^2 = 0.12$, power = 0.98, where the TD-CS group's performance was significantly poorer than that of the TD-EP group. To determine if the same pattern held in for the CS cohort in Spanish, a second univariate analysis of variance was conducted comparing overall nonword repetition performance on the NRT-Spa for the TD-CS and TD-EP groups. The analysis again revealed a significant main effect of group $F(1, 110) = 12.0$, $p < 0.01$, partial $\eta^2 = 0.10$, power = 0.93, where the TD-CS group's performance on the NRT-Spa was also significantly poorer than that of the TD-EPgroup.

In sum, for both the DLD and TD children in the Catalan–Spanish-speaking cohort, nonword repetition ability was poorer as compared to that of the European Portuguese-speaking cohort both when compared in Catalan and Spanish.

3.2. Use of NRT to Rule In/Rule Out Developmental Language Disorder

The above analysis replicates prior work showing that poor performance for children with DLD on NRT having low wordlikeness (i.e., [49]) is well documented (i.e., [46]). However, one question is related to the *diagnostic accuracy* of children's performance on

NRT that vary BOTH by syllable length AND wordlikeness to identify the presence of DLD in Catalan–Spanish and European Portuguese-speaking children.

3.2.1. Total Percentage of Words Correct (TOT PWC)

In this study, we calculated the likelihood ratios (LH) for children's performance on the NRT to assess its diagnostic accuracy. Ultimately, the value of a diagnostic test will depend upon its ability to alter a pre-test probability of a target condition into a post-test probability that will influence a clinical management decision. The positive LH ratio is the ratio of the proportion of patients who have the target condition and test positive to the proportion of patients without the target condition who also test positive. In conveying the meaning of diagnostic accuracy to clinicians, research suggests that LH ratios are more interpretable to clinicians and enable more appropriate interpretation of tests as compared to sensitivity and specificity measures. Specifically, general practitioners when asked to estimate the probability of a disease in a given patient give the most appropriate estimation of test performance when provided LHs as compared to sensitivity/specificity values as compared to LH ratios [75].

Likelihood ratio (LH) analyses were conducted using the presence/absence of DLD based on the gold standard classification for the diagnosis of DLD in each cohort to determine whether Catalan–Spanish and European Portuguese-speaking children's ability to repeat nonwords serves as a screening tool to detect and diagnose children with DLD in each cohort [76,77]. To determine the LH ratio for a positive result based on total percentage of words correct at test on the NRT (TOT-PWC), the true positive rate (proportion of children with DLD with a total NRT (TOT-PWC) at or below *x-determined* cutoff) was divided by the false positive rate (proportion of TD children with total NRT at or below *x-determined* cutoff) for each cohort. We used Haynes et al. [77] criteria to classify a positive test (i.e., accurately ruling in the disorder), which includes the following: (1) "High" as defined as LH ratio of 20 or higher having a probability of 95% or greater that the disorder is present, (2) "Intermediate High" defined as an LH ratio between 1 and 20, and (3) "Indeterminate" defined as an LH close to 1.0. To calculate the cutoff scores to maximize the ability to "rule in" DLD, we calculated the number and proportion of children in the DLD and TD groups whose scores were at or fell below (test positive) a given NRT (TOT PWC) value.

The number and proportion of children with a positive test result, and the LH ratios and prevalence for Total Words Correct for the Catalan–Spanish and European Portuguese-speaking children are shown in Table 6. As can be seen in Table 6, the LH ratio for a positive test result for the Catalan–Spanish-speaking children with DLD-CS compared to TD-CS controls for the Catalan version of the NRT was 33.00 (0.917/0.028), indicating that a child having a score of TOT-PWC of 70% or lower was 33 times more likely to be a child with DLD-CS as compared to a child with normal language. For the European Portuguese-speaking children with DLD-EP compared to TD-EP controls, the LH ratio for a positive test result was 35.00 (0.467/0.013), indicating that a TOT-PWC of 70% or lower was 35 times more likely to come from a child with DLD-EP as compared to one with normal language. These findings show that for the NRT-Cat and NRT-EPort, the best cut-point was the same, which was 70%. For the Spanish version of the NRT, the LH ratio for a positive test result was 33.88 (0.941/0.028), indicating that a TOT-PWC of 75% or lower was 33 times more likely to come from a child with DLD-CS as compared to a child with normal language.

Table 6. The number and proportion of children with a positive test result (ruling in disorder), Prevalence and Likelihood Ratio (LH) for each of the cutoff values based on total word percent correct (TOT-PWC) on the NRT for the Catalan–Spanish (CS) and European Portuguese (EP) children with developmental language disorder (DLD) and typically developing (TD) controls.

TOT-PWC	DLD (True Positive)		TD (False Positive)		Prevalence	LH
Cutoff	No.	Prop.	No.	Prop.		
			CS-Catalan			
≤65.0	30	0.833	1	0.028	0.50	30.00
≤70.0	**33**	**0.917**	**1**	**0.028**	**0.50**	**33.00**
≤75.0	34	0.944	3	0.083	0.50	11.33
≤80.0	35	0.972	7	0.194	0.50	5.00
≤85.0	35	0.972	10	0.278	0.50	3.50
≤90.0	36	1.00	21	0.583	0.50	1.71
≤95.0	36	1.00	32	0.889	0.50	1.13
≤100	36	1.00	36	1.00	0.50	1.00
			CS-Spanish			
≤65.0	25	0.735	0	0.000	0.48	Inf.
≤70.0	27	0.794	1	0.028	0.48	28.58
≤75.0	**32**	**0.941**	**1**	**0.028**	0.48	**33.88**
≤80.0	32	0.941	6	0.166	0.48	5.65
≤85.0	34	1.00	12	0.333	0.48	3.00
≤90.0	34	1.00	18	0.500	0.48	2.00
≤95.0	34	1.00	29	0.805	0.48	1.24
≤100	34	1.00	36	1.00	0.48	1.00
			European Portuguese			
≤65.0	22	0.293	0	0.000	0.50	Inf.
≤70.0	**35**	**0.467**	**1**	**0.013**	**0.50**	**35.00**
≤75.0	46	0.613	2	0.027	0.50	23.00
≤80.0	59	0.787	2	0.027	0.50	29.50
≤85.0	72	0.960	7	0.093	0.50	10.29
≤90.0	74	0.987	18	0.240	0.50	4.11
≤95.0	75	1.00	51	0.680	0.50	1.47
≤100	75	1.00	75	1.00	0.50	1.00

Note. Values ≤ 65.0 cutoffs are not presented because LH = Inf. The cutoffs for best LH are highlighted in bold.

3.2.2. 2-Syllable Percentage of Words Correct (TOT PWC)

As can be seen in Table 7, the LH ratio for a positive test result for the Catalan–Spanish-speaking children with DLD-CS compared to TD-CS controls for the NRT-Cat was 22.00 (0.611/0.028), indicating that a TOT-PWC of 75% or lower for 2-syllable words was 22 times more likely to come from a child with DLD-CS as compared to a child with normal language. Similarly to NRT-Cat, the LH ratio for a positive test result for the European Portuguese-speaking children with DLD-EP compared to TD-EP controls for the NRT-EPort was 13.00 (0.520/0.040), indicating that a TOT-PWC of 75% or lower was 13 times more likely to come from a child with DLD-EP as compared to one with normal language. These findings show that for the NRT-Cat and NRT-EPort, the best cut-point was the same, which was 75%. For the NRT-Spa, the LH ratio for a positive test result was 7.41 (0.411/0.055), indicating that a TOT-PWC of 80% or lower for 2-syllable words was seven times more likely to come from a child with DLD-CS as compared to a child with normal language.

Table 7. The number and proportion of children with a positive test result (ruling in disorder), prevalence and likelihood ratio (LH) for each of the cutoff value based on total word percentage correct (TOT-PWC) for 2-syllable length nonwords for the Catalan–Spanish (CS) and European Portuguese (EP) children with developmental language disorder (DLD) and typically developing (TD) controls.

2-syll TOT-PWC	DLD (True Positive)		TD (False Positive)		Prevalence	LH
Cutoff	No.	Prop.	No.	Prop.		
			CS-Catalan			
≤75.0	**22**	**0.611**	**1**	**0.028**	**0.50**	**22.00**
≤87.5	31	0.861	6	0.167	0.50	5.16
≤100	36	1.00	36	1.00	0.50	1.00
			CS-Spanish			
≤75.0	8	0.235	0	0.000	0.48	Inf.
≤80.0	**14**	**0.411**	**2**	**0.055**	0.48	**7.41**
≤85.0	17	0.500	3	0.083	0.48	6.00
≤90.0	27	0.794	9	0.250	0.48	3.18
≤95.0	32	0.941	20	0.555	0.48	1.69
≤100	34	1.00	36	1.00	0.48	1.00
			European Portuguese			
≤75.0	**39**	**0.520**	**3**	**0.040**	**0.50**	**13.00**
≤87.5	67	0.893	23	0.307	0.50	2.91
≤100	75	1.00	75	1.00	0.50	1.00

Note. The possible cutoffs for Spanish NRT are different from those for Catalan and European Portuguese NRT because for the Spanish version of the task, there are 20 nonwords for each syllable length and for the Catalan and European Portuguese versions, there are eight nonwords for each syllable length. Values ≤ 75.0 cutoffs are not presented because LH = Inf. The cutoffs for best LH are highlighted in bold.

3.2.3. 3-Syllable Percentage of Words Correct (TOT PWC)

As can be seen in Table 8, the LH ratio for a positive test result for the Catalan–Spanish-speaking children with DLD-CS compared to TD-CS controls for the NRT-Cat was 17.00 (0.472/0.028), indicating that a TOT-PWC of 62.5% or lower for 3-syllable words was 17 times more likely to come from a child with DLD-CS as compared to a child with normal language. Similarly to NRT-Cat, the LH ratio for a positive test result for the NRT-EPort for the European Portuguese-speaking children with DLD-EP compared to TD-EP controls was 22.00 (0.293/0.013), indicating that a TOT-PWC of 62.5% or lower was 22 times more likely to come from a child with DLD-EP as compared to a with normal language. These findings show that for the NRT-Cat and NRT-EPort, the best cut-point was the same, which was 62.5%. For the NRT-Spa, the LH ratio for a positive test result was 20.12 (0.559/0.028), indicating that a TOT-PWC of 70% or lower for 3- syllable words was 20 times more likely to come from a child with DLD-CS as compared to a child with normal language.

Table 8. The number and proportion of children with a positive test result (ruling in disorder), prevalence and likelihood ratio (LH) for each of the cutoff value based on total word percentage correct (TOT-PWC) for 3-syllable nonwords for the Catalan–Spanish (CS) and European Portuguese (EP) children with developmental language disorder (DLD) and typically developing (TD) controls.

3-syll TOT-PWC	DLD (True Positive)		TD (False Positive)		Prevalence	LH
Cutoff	No.	Prop.	No.	Prop.		
			CS-Catalan			
≤50.0	5	0.139	1	0.028	0.50	5.00
≤62.5	**17**	**0.472**	**1**	**0.028**	**0.50**	**17.00**
<75.0	25	0.694	2	0.056	0.50	12.50
≤87.5	35	0.972	16	0.444	0.50	2.19
≤100	36	1.00	36	1.00	0.50	1.00
			CS-Spanish			
≤55.0	9	0.264	0	0.000	0.48	Inf.
≤60.0	11	0.324	0	0.000	0.48	Inf.
≤65.0	12	0.353	0	0.000	0.48	Inf.
≤70.0	**19**	**0.559**	**1**	**0.028**	0.48	**20.12**
≤75.0	21	0.618	2	0.056	0.48	11.12
≤80.0	23	0.676	3	0.083	0.48	8.12
≤85.0	27	0.794	8	0.222	0.48	3.57
≤90.0	32	0.941	16	0.444	0.48	2.12
≤95.0	33	0.971	25	0.694	0.48	1.40
≤100	34	1.00	36	1.00	0.48	1.00
			European Portuguese			
≤50.0	5	0.067	0	0.000	0.50	Inf.
≤62.5	**22**	**0.293**	**1**	**0.013**	**0.50**	**22.00**
<75.0	47	0.627	9	0.120	0.50	5.22
≤87.5	71	0.947	34	0.453	0.50	2.09
≤100	75	1.00	75	1.00	0.50	1.00

Note. The possible cut-points for Spanish NRT are different from Catalan and European Portuguese NRT because for the Spanish version of the task, there are 20 nonwords for each syllable length, and for Catalan and European Portuguese, there are eight nonwords for each syllable length. Values ≤ 50.0/55.0 cutoffs are not presented, because LH = Inf. The cutoffs for best LH are highlighted in bold.

3.2.4. 4-Syllable Percentage of Words Correct (TOT PWC)

As can be seen in Table 9, the LH ratio for a positive test result for the Catalan–Spanish-speaking children with DLD-CS compared to TD-CS controls for the NRT-Cat was 23.00 (0.639/0.028), indicating that a TOT-PWC of 62.5% or lower for 4-syllable words was 23 times more likely to come from a child with DLD-CS as compared to one with normal language. Similarly to NRT-Cat, the LH ratio for a positive test result for the NRT-EPort for the European Portuguese-speaking children with DLD-EP compared to TD-EP controls was 20.00 (0.267/0.013), indicating that a TOT-PWC of 62.5% or lower was 20 times more likely to come from a child with DLD-EP as compared to one with normal language. These findings show that for the NRT-Cat and NRT-EPort, the best cut-point was 62.5%. These findings show that for the NRT-Cat and NRT-EPort, the best cut-point was the same, which was 70%. For the NRT-Spa, the LH ratio for a positive test result was 27.53 (0.765/0.028), indicating that a TOT-PWC of 60% or lower for 4-syllable words was 27 times more likely to come from a child with DLD-CS as compared to a child with normal language.

Table 9. The number and proportion of children with a positive test result (ruling in disorder), prevalence and likelihood ratio (LH) for each of the cutoff values based on total word percentage correct (TOT-PWC) for 4-syllable length nonwords for the Catalan–Spanish (CS) and European Portuguese (EP) children with developmental language disorder (DLD) and typically developing (TD) controls.

4-syll TOT-PWC	DLD (True Positive)		TD (False Positive)		Prevalence	LH
Cutoff	No.	Prop.	No.	Prop.		
			CS-Catalan			
≤50.0	17	0.472	0	0.028	0.50	Inf.
≤62.5	**23**	**0.639**	**1**	**0.028**	**0.50**	**23.00**
<75.0	31	0.861	6	0.017	0.50	5.17
≤87.5	34	0.944	16	0.444	0.50	2.13
≤100	36	1.00	36	1.00	0.50	1.00
			CS-Spanish			
≤55.0	23	0.676	1	0.028	0.48	24.35
≤60.0	**26**	**0.765**	**1**	**0.028**	**0.48**	**27.53**
≤65.0	27	0.794	3	0.083	0.48	9.53
≤70.0	30	0.882	4	0.111	0.48	7.94
≤75.0	33	0.971	7	0.194	0.48	4.99
≤80.0	34	1.00	11	0.306	0.48	3.27
≤85.0	34	1.00	15	0.417	0.48	2.40
≤90.0	34	1.00	25	0.694	0.48	1.44
≤95.0	34	1.00	29	0.806	0.48	1.24
≤100	34	1.00	36	1.00	0.48	1.00
			European Portuguese			
≤50.0	8	0.107	0	0.000	0.50	Inf.
≤62.5	**20**	**0.267**	**1**	**0.013**	**0.50**	**20.00**
<75.0	46	0.613	3	0.040	0.50	15.33
≤87.5	63	0.840	18	0.240	0.50	3.50
≤100	75	1.00	75	1.00	0.50	1.00

Note. The possible cut-points for Spanish NRT are different from Catalan and European Portuguese NRT, because for the Spanish version of the task, there are 20 nonwords for each syllable length, and for Catalan and European Portuguese, there are eight nonwords for each syllable length. Values ≤ 50.0/55.0 cutoffs are not presented, because LH = Inf. The cutoffs for the best LH are highlighted in bold.

3.2.5. 5-Syllable Percentage of Words Correct (TOT PWC)

As can be seen in Table 10, the LH ratio for a positive test result for the Catalan–Spanish-speaking children with DLD-CS compared to TD-CS controls for the NRT-Cat was 29.00 (0.806/0.028), indicating that a TOT-PWC of 50% or lower for 5-syllable words was 29 times more likely to come from a child with DLD-CS as compared to a child with normal language. Similarly to NRT-Cat, the LH ratio for a positive test result for the NRT-EPort for the European Portuguese-speaking children with DLD-EP compared to TD-EP controls was 22.00 (0.293/0.013), indicating that a TOT-PWC of 62.5% or lower was 22 times more likely to come from a child with DLD-EP as compared to a child with normal language. Unlike the previous syllable length, the best cut-point differed. For the NRT-Spa, the LH ratio for a positive test result was 16.94 (0.941/0.056), indicating that a TOT-PWC of 55% or lower for 5-syllable words was 16 times more likely to come from a child with DLD-CS as compared to a child with normal language.

Table 10. The number and proportion of children with a positive test result (ruling in disorder), prevalence and likelihood ratio (LH) for each of the cutoff value based on total word percentage correct (TOT-PWC) for 5-syllable length nonwords for the Catalan–Spanish (CS) and European Portuguese (EP) children with developmental language disorder (DLD) and typically developing (TD) controls.

5-syll TOT-PWC	DLD (True Positive)		TD (False Positive)		Prevalence	LH
Cutoff	No.	Prop.	No.	Prop.		
			CS-Catalan			
≤50.0	**29**	**0.806**	**1**	**0.028**	**0.50**	**29.00**
≤62.5	31	0.861	5	0.139	0.50	6.20
<75.0	35	0.972	10	0.278	0.50	3.50
≤87.5	35	0.972	29	0.805	0.50	1.20
≤100	36	1.000	36	1.000	0.50	1.00
			CS-Spanish			
≤45.0	27	0.794	2	0.056	0.48	14.29
≤55.0	**32**	**0.941**	**2**	**0.056**	**0.48**	**16.94**
≤60.0	33	0.971	5	0.139	0.48	6.99
≤65.0	33	0.971	7	0.194	0.48	4.99
≤70.0	33	0.971	9	0.250	0.48	3.88
≤75.0	33	0.971	12	0.333	0.48	2.91
≤80.0	34	1.00	15	0.417	0.48	2.40
≤85.0	34	1.00	24	0.667	0.48	1.50
≤90.0	34	1.00	26	0.722	0.48	1.38
≤95.0	34	1.00	32	0.889	0.48	1.13
≤100	34	1.00	36	1.00	0.48	1.00
			European Portuguese			
≤50.0	14	0.187	0	0.000	0.50	Inf.
≤62.5	**22**	**0.293**	**1**	**0.013**	**0.50**	**22.00**
<75.0	35	0.467	3	0.040	0.50	11.67
≤87.5	58	0.773	24	.320	0.50	2.42
≤100	75	1.00	75	1.00	0.50	1.00

Note. The possible cut-points for Spanish NRT are different from Catalan and European Portuguese NRT because for the Spanish version of the task, there are 20 nonwords for each syllable length, and for Catalan and European Portuguese, there are eight nonwords for each syllable length. Values ≤ 45.0/50.0 cutoffs are not presented, because LH = Inf. The cutoffs for the best LH are highlighted in bold.

3.2.6. 6-Syllable Percentage of Words Correct (TOT PWC)

As can be seen in Table 11, the LH ratio for a positive test result for the Catalan–Spanish-speaking children with DLD-CS compared to TD-CS controls for the NRT-Cat was 10.67 (0.889/0.083), indicating that a TOT-PWC of 25% or lower for 6-syllable words was 10 times more likely to come from a child with DLD-CS as compared to a child with normal language. In contrast, the LH ratio for a positive test result for the NRT-EPort for the *European Portuguese*-speaking children with DLD-EP compared to TD-EP controls was 18.33 (0.733/0.040), indicating that a TOT-PWC of 50% or lower was 18 times more likely to come from a child with DLD-EP as compared to a child with normal language. These findings show that for the NRT-Cat and NRT-EPort, the best cut-point differed.

Table 11. The number and proportion of children with a positive test result (ruling in disorder), prevalence and likelihood ratio (LH) for each of the cutoff value based on total word percentage correct (TOT-PWC) for 6-syllable length nonwords for the Catalan–Spanish (CS) and European Portuguese (EP) children with developmental language disorder (DLD) and typically developing (TD) controls.

6-syll TOT-PWC	DLD (True Positive)		TD (False Positive)		Prevalence	LH
Cutoff	No.	Prop.	No.	Prop.		
			CS-Catalan			
≤25.0	**32**	**0.889**	**3**	**0.083**	**0.50**	**10.67**
≤37.5	33	0.917	4	0.111	0.50	8.25
≤50.0	34	0.944	6	0.167	0.50	5.67
≤62.5	35	0.972	15	0.417	0.50	2.33
<75.0	36	1.00	20	0.556	0.50	1.80
≤87.5	36	1.00	26	0.722	0.50	1.38
≤100	36	1.00	36	1.00	0.50	1.00
			European Portuguese			
≤25.0	8	0.107	0	0.000	0.50	Inf.
≤37.5	30	0.400	0	0.000	0.50	Inf.
≤50.0	**55**	**0.733**	**3**	**0.040**	**0.50**	**18.33**
≤62.5	64	0.853	8	0.107	0.50	8.00
<75.0	73	0.973	18	0.240	0.50	4.06
≤87.5	74	0.987	49	0.653	0.50	1.51
≤100	75	1.00	75	1.00	0.50	1.00

Note. The Spanish NRT version does not include nonwords of 6-syllable length. Values ≤ 25.0 cutoffs are not presented, because LH = Inf. The cutoffs for best LH are highlighted in bold.

3.2.7. Summary of LH Ratios for Total Words Correct and by Syllable Length

There were similarities and differences in the cutoff points for total words percentage correct both overall and by syllable length for the two language cohorts and NRT. The absolute cutoff points having the best LH ratios for overall words produced correctly were similar for the NRT-Cat and NRT-EPort for ToT performance and for 2-, 3-, and 4-syllable lengths but differed for 5-syllable and 6-syllable lengths. However, the cut-points differed from the NRT-Spa. For the total performance for all three languages, a cut-point resulted in high LHs (i.e., greater than 20) based on Haynes et al. [77]. For 2-syllable length, NRT-Cat LHs were high and NRT-Span and NRT-EPort LHs were intermediate (i.e., 1–20). For 3-syllable lengths, NRT-Span and NRT-EPort LHs were high and that of NRT-Cat was intermediate but was close to the high cutoff (i.e., 17). For 4-syllable length, for all three languages, the LHs were all high. For 5-syllable length, NRT-Cat and NRT-EPort LHs were high and that of NRT-Span was intermediate but close to the high cutoff (i.e., 16.9). For 6-syllable length, the LHs were intermediate, with NRT-EPort LHs being close to high cutoff (i.e., 18.33).

3.3. Wordlikeness/Nonword Frequency

3.3.1. High Wordlikeness/High Nonword Frequency (TOT PWC)

As can be seen in Table 12, for nonwords having high wordlikeness, the LH ratio for a positive test result for the Catalan–Spanish speaking children with DLD-CS compared to TD-CS controls for the NRT-Cat was 29.00 (0.806/0.028), indicating that a TOT-PWC of 65% or lower for nonwords having high wordlikeness regardless of syllable length was 29 times more likely to come from a child with DLD-CS as compared to a child having normal language. For nonwords having high wordlikeness, the LH ratio for a positive test result for the NRT-EPort for the European Portuguese-speaking children with DLD-EP compared to TD-EP controls was 39.00 (0.520/0.013), indicating that a TOT-PWC of 75% or lower was 39 times more likely to come from a child with DLD-EP as compared to a child having normal language. These findings show that for the NRT-Cat and NRT-EPort, the best cut-point differed. For the Spanish NRT for high-frequency nonwords, the LH ratio for a positive test result was 16.41 (0.912/0.056), indicating that a TOT-PWC of 80% or lower for nonwords having high nonword frequency regardless of syllable length was 16 times more likely to come from a child with DLD-CS as compared to a child with normal language.

Table 12. The number and proportion of children with a positive test result (ruling in disorder), prevalence, and likelihood ratio (LH) for each of the cutoff values based on total word percentage correct (TOT-PWC) for HIGH wordlikeness (HWL) nonwords for Catalan and Portuguese NRT and HIGH nonword frequency (HNF) nonwords for Spanish NRT for the Catalan–Spanish (CS) and European Portuguese (EP) children with developmental language disorder (DLD) and typically developing (TD) controls.

HWL/NF TOT-PWC	DLD (True Positive)		TD (False Positive)		Prevalence	LH
Cutoff	No.	Prop.	No.	Prop.		
			CS-Catalan			
≤60.0	26	0.722	1	0.028	0.50	26.00
≤65.0	**29**	**0.806**	**1**	**0.028**	**0.50**	**29.00**
≤70.0	31	0.861	3	0.083	0.50	10.33
≤75.0	33	0.917	5	0.139	0.50	6.60
≤80.0	34	0.944	9	0.250	0.50	3.78
≤85.0	35	0.972	14	0.389	0.50	2.50
≤90.0	36	1.00	24	0.667	0.50	1.50
≤95.0	36	1.00	32	0.889	0.50	1.13
≤100	36	1.00	36	1.00	0.50	1.00
			CS-Spanish			
≤60.0	14	0.412	0	0.000	0.48	Inf.
≤65.0	17	0.500	0	0.000	0.48	Inf.
≤70.0	22	0.647	0	0.000	0.48	Inf.
≤75.0	27	0.794	2	0.056	0.48	14.29
≤80.0	**31**	**0.912**	**2**	**0.056**	0.48	**16.41**
≤85.0	33	0.971	3	0.083	0.48	11.65
≤90.0	34	1.00	10	0.278	0.48	3.60
≤95.0	34	1.00	26	0.722	0.48	1.38
≤100	34	1.00	36	1.00	0.48	1.00
			European Portuguese			
≤60.0	11	0.147	0	0.000	0.50	Inf.
≤65.0	17	0.227	0	0.000	0.50	Inf.
≤70.0	22	0.293	0	0.000	0.50	Inf.
≤75.0	**39**	**0.520**	**1**	**0.013**	**0.50**	**39.00**
≤80.0	56	0.747	3	0.040	0.50	18.67
≤85.0	67	0.893	7	0.093	0.50	9.57
≤90.0	72	0.960	18	0.240	0.50	4.00
≤95.0	75	1.00	48	0.640	0.50	1.56
≤100	75	.227	75	1.00	0.50	1.00

Note. Values ≤ 60.0 cutoffs are not presented because LH = Inf. The cutoffs for best LH are highlighted in bold.

3.3.2. Low Wordlikeness/Low Nonword Frequency (TOT PWC)

As can be seen in Table 13, for nonwords having low wordlikeness, the LH ratio for a positive test result for the Catalan–Spanish-speaking children with DLD-CS compared to TD-CS controls for the NRT-Cat was 31.00 (0.861/0.028), indicating that a TOT-PWC of 65% or lower for nonwords having low wordlikeness regardless of syllable length was 31 times more likely to come from a child with DLD-CS as compared to a child with normal language. In contrast, for nonwords having low wordlikeness, the LH ratio for a positive test result for the NRT-EPort for the European Portuguese children with DLD-EP compared to TD-EP controls was 35.00 (0.467/0.013), indicating that a TOT-PWC of 65% or lower was 35 times more likely to come from a child with DLD-EP as compared to a child with normal language. These findings show that for the NRT-Cat and NRT-EPort, the best cut-point was the same, which was 65%. For the NRT-Spa, the LH ratio for a positive test result was 14.82 (0.824/0.056), indicating that a TOT-PWC of 65% or lower for nonwords having low nonword frequency regardless of syllable length was 14 times more likely to come from a child with DLD-CS as compared to a child with normal language.

Table 13. The number and proportion of children with a positive test result (ruling in disorder), prevalence and likelihood ratio (LH) for each of the cutoff value based on total word percentage correct (TOT-PWC) for LOW wordlikeness (LWL) nonwords for Catalan and Portuguese NRT (NRT-Cat and NRT-Port) and LOW nonword frequency (HNF) nonwords for Spanish NRT (NRT-Spa) for the Catalan–Spanish (CS) and European Portuguese (EP) children with developmental language disorder (DLD) and typically developing (TD) controls.

LWL/LNF TOT-PWC	DLD (True Positive)		TD (False Positive)		Prevalence	LH
Cutoff	No.	Prop.	No.	Prop.		
			CS-Catalan			
≤55.0	28	0.778	1	0.028	0.50	28.00
≤60.0	29	0.806	1	0.028	0.50	29.00
≤65.0	**31**	**0.861**	**1**	**0.028**	**0.50**	**31.00**
≤70.0	34	0.944	2	0.056	0.50	17.00
≤75.0	34	0.944	4	0.111	0.50	8.50
≤80.0	35	0.972	7	0.194	0.50	5.00
≤85.0	36	1.00	14	0.389	0.50	2.57
≤90.0	36	1.00	19	0.528	0.50	1.89
≤95.0	36	1.00	33	0.917	0.50	1.09
≤100	36	1.00	36	1.00	0.50	1.00
			CS-Spanish			
≤55.0	21	0.618	0	0.000	0.48	Inf.
≤60.0	24	0.706	0	0.000	0.48	Inf.
≤65.0	**28**	**0.824**	**2**	**0.056**	0.48	**14.82**
≤70.0	30	0.882	6	0.166	0.48	5.29
≤75.0	32	0.941	7	0.194	0.48	4.84
≤80.0	33	0.971	10	0.278	0.48	3.49
≤85.0	34	1.00	17	0.472	0.48	2.12
≤90.0	34	1.00	24	0.667	0.48	1.50
≤95.0	34	1.00	30	0.833	0.48	1.20
≤100	34	1.00	36	1.00	0.48	1.00
			European Portuguese			
≤55.0	20	0.267	1	0.013	0.50	20.00
≤60.0	27	0.360	1	0.013	0.50	27.00
≤65.0	**35**	**0.467**	**1**	**0.013**	**0.50**	**35.00**
≤70.0	42	0.560	2	0.027	0.50	21.00
≤75.0	54	0.720	4	0.053	0.50	13.50
≤80.0	64	0.853	6	0.080	0.50	10.67
≤85.0	72	0.960	14	0.187	0.50	5.14
≤90.0	73	0.973	31	0.413	0.50	2.35
≤95.0	75	1.00	61	0.813	0.50	1.23
≤100	75	1.00	75	1.00	0.50	1.00

Note. Values ≤ 55.0 cutoffs are not presented because LH = Inf. The cutoffs for best LH are highlighted in bold.

3.3.3. Summary of LH Ratios for Wordlikeness/Nonword Frequency

For nonwords having high wordlikeness ratings, although the absolute cut-points differed for the Catalan and European Portuguese NRTs, the LH for both cohorts fell in the high range, according to Haynes et al. [77]. For the Spanish NRT, the LH for high nonword frequency was in the intermediate range. For nonwords having low wordlikeness for the Catalan and European Portuguese, the NRTs were the same and the LH for both cohorts again fell in the high range according to Haynes et al. [77]. For low-frequency nonwords, although the absolute cut-points were the *same* as those of the NRT-Cat and NRT-EPort, for the NRT-Spa, the LH fell in the intermediate high range according to Haynes et al. [77]. Taken together, the findings from the wordlikeness versus nonword frequency comparison indicates that wordlikeness may be more sensitive to DLD diagnosis regardless of the language used as compared to nonword frequency with the LHs for wordlikeness being more than double those of word frequency.

4. Summary of Results and Discussion

In this study, we investigated the diagnostic accuracy of using a nonword repetition task to discriminate DLD for TD across three different languages. For two of the three languages (Catalan and European Portuguese), NRT lists were created and administered using the same method, and for Spanish NRT, we used an existing list from Aguado [64]. Although there were absolute accuracy differences across the Catalan–Spanish and European Portuguese-speaking children, a similar pattern of performance accuracy was observed for both syllable length and wordlikeness (NRT-Cat and NRT-Port)/nonword frequency (NRT-Spa) across the two cohorts. Specifically, the children with DLD performed worse than their TD peers across all syllable lengths, with performance worsening as the syllable length increased.

For the LH analysis, a strikingly similar pattern was observed across the two cohorts in the pattern of cut-points for the LH ratios both for overall performance, syllable length, wordlikeness, and nonword frequency. Importantly, the similarities were highest for the Catalan and European Portuguese NRT. Specifically, the LHs for the two versions mirror each other both for syllable length and wordlikeness, where very high LH values were observed for both high and low wordlikness. This pattern was not evident in the Spanish NRT based on the frequency of the nonwords.

Children in both cohort groups with DLD repeated nonwords with high wordlikeness with greater accuracy than those with low wordlikeness, which is a pattern that was not evident in the TD controls. This finding was in line with past research, which has found similar patterns with children with DLD performing better with high wordlikeness and/or high phonotactic frequency nonwords, whereas this performance difference is not as extreme in the TD controls [25,78]. As has been mentioned previously, nonword repetition ability correlates with language proficiency, but it also correlates with the amount of language exposure in both typical and clinical populations [43]. With greater language exposure and proficiency comes greater long-term linguistic representations. This means that although a nonword may have low phonotactic probability or low wordlikeness, a child with larger exposure and proficiency has been exposed to more instances of words that conform to the low phonotactic/wordlikeness nonwords, and as such, they are more common for these children. On the other hand, children with DLD have limited linguistic representations within their extant long-term memory, and as such, the difference between low wordlikeness/frequency nonwords and high wordlikeness/frequency nonwords is much starker than for their TD peers, resulting in an advantageous interaction for high wordlikeness/frequency that is not found in the TD populations.

In this study, although performance for both groups of children with DLD was lower than that of their respective TD peers, overall, the bilingual Catalan–Spanish-speaking children had worse performance than the monolingual children who speak European Portuguese. This finding adds to previous literature showing that bilingual children perform less accurately during NRT than their monolingual peers [43,79]. These previous studies investigated sequentially bilingual children and found that these children were less accurate when repeating nonwords in the language with which they had less exposure and less proficiency than their monolingual peers. An additional study compared the NRT accuracy in four groups of children (monolingual English TD and DLD, and bilingual Spanish–English TD and DLD), and it found that the bilingual Spanish–English children performed worse on English nonwords than the respective monolingual groups, but the bilingual children with DLD outperformed the monolingual English DLD group on Spanish nonwords, with all three groups outperformed by the bilingual TD children [34].

It is important for the purposes of the current study that despite the overall difference in performance accuracy between the two language groups (Catalan–Spanish, European Portuguese) and language status (bilingual vs. monolingual), the LH values remained constant and nearly identical across the groups. In the case of the Catalan and European Portuguese languages, a possible reason for this stability may be the identical method used to create the nonword lists for both language cohorts. However, although the LH values found for the Spanish NRT lists were at times lower, they were also highly consistent with the other language lists. This consistency in LH values strongly suggests the validity of utilizing NRT to distinguish DLD from TD across a range of linguistically diverse populations, even when similar methods are not used to create NRT lists.

Finally, the high LH values that we found across our four groups may be a result of (1) the unpredictable presentation order of the nonwords for the Catalan and European Portugese NRT; and (2) differences in the wordlikeness ratings of the nonwords in the NRT-Cat and NRT-EPort stimuli. Specifically, in our study, the Catalan and European Portuguese nonwords were not presented sequentially in accordance to syllable length (i.e., all 2-syllable nonwords followed by all 3-syllable nonwords, etc.). Instead, the nonwords in our lists were presented in a fixed random order, such that participants could not predict the syllable length or wordlikeness of an upcoming nonword. The unpredictability of stimuli presentation has been found to impact task difficulty and subsequent performance in studies of adult cognition [80] and may have increased the difficulty in the NRT that effectively negated the ceiling effect for our four groups. However, it is worth noting that few studies have reported whether the presentation of their nonwords were sequential or randomized, and as such, this interpretation of results must remain somewhat speculative. Future research should investigate the influence of the order of nonword stimuli presentation on accuracy to ensure that behavioral results are truly representative of the cognitive processes that underlie nonword repetition itself, and not an artifact of stimuli presentation methodology.

Taken together, the findings from this study suggest that in the absence of pre-existing standardized NRT, to use NRT as a potential identifier of DLD, an NRT task can be developed from scratch in a child's native language using the method outlined in this study for the NRT-Cat and NRT-EPort. Furthermore, the findings from this study show that using 3-, 4-, and 5-syllable nonwords that are presented in a random order is a valuable presentation approach. Finally, using nonwords that vary in wordlikeness may result in the most sensitive version of an NRT task. Surprisingly, and encouragingly, for languages where large-scale word databases are not available to calculate measures such as word frequency, phonotactic probability, etc., the findings from this study show that using a native speaker's ranking of wordlikeness of nonwords results in a valuable measure that appears to increase the diagnostic sensitivity of an NRT task.

There has been a growing interest in the development of behavioral markers of DLD. For the majority of children with DLD, the deficits extend beyond the language system to include deficits in the nonverbal and cognitive domains as well. If unidentified or left untreated, these deficits persist into adulthood, putting these children at risk for failure in academic and work settings, lower standard of living, social isolation, and significant secondary socio-emotional and stress-related health issues. The variability in the DLD language deficit profile coupled with a diagnostic classification system based on exclusion criteria makes the identification and effective treatment of DLD a significant challenge to both researchers and clinicians. The identification of markers for the objective diagnosis of DLD have broad health-related implications. Specifically, tasks that are not language-specific and that have the ability to differentiate DLD from typically developing children have the potential to aid in the refinement of the diagnostic accuracy of DLD, help identify potential subtypes of the disorder, resulting in better estimations of prognosis for the disorder and predictive validity of individual symptoms of DLD, and aid in the development of more effective therapeutic approaches that are more tightly linked to the underlying cause(s) of the disorder.

Author Contributions: Conceptualization, J.L.E., A.B.-B., A.C.-S. and N.A.; methodology, A.C.-S., N.A. and L.F.; validation, A.C.-S., N.A. and L.F.; formal analysis, J.L.E., A.C.-S. and N.A.; investigation, A.C.-S., N.A. and L.F.; resources J.L.E., L.A. and M.S.-T.; data curation, A.C.-S., N.A. and L.F.; writing—original draft preparation, J.L.E., A.B.-B., A.C.-S. and N.A.; writing—review and editing, J.L.E., A.B.-B., A.C.-S. and N.A.; visualization, J.L.E. and N.A.; supervision, J.L.E., L.A. and M.S.-T.; project administration, J.L.E., L.A. and M.S.-T.; funding acquisition, J.L.E., A.C.-S., L.A. and M.S.-T. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Institute on Deafness and Other Communication Disorders (NIDCD) Grant [R0105263] awarded to Dr. Evans, from The National Institute of Child Health and Human Development (NICHD) Grant [P30HD03352] awarded to the Waisman Center, The University of Wisconsin Madison; by the Ministerio de Economía y Competitividad Grant [2016EDU2016-75368-P/BES-2014-070511] awarded from Dr. Andreu and from the Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR) de la Generalitat de Catalunya [2017SGR387] awarded to the Grup de Recerca en Cognició i Llenguatge (UOC-UB); by the CIEd—Research Centre on Education, Institute of Education, University of Minho, projects [UIDB/01661/2020] and [UIDP/01661/2020], through national funds of FCT/MCTES-PT.

Institutional Review Board Statement: The Catalan-Spanish protocol was approved by the Ethics Committee of Universitat Oberta de Catalunya (Project identification code: 2016EDU2016-75368-P). The European Portuguese protocol was approved by the Ethics Committee of the University of Minho (Project identification code: SFRH/BD/9604/ 2004).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments: We especially want to thank the children and their families who generously contributed their time. We would also like to thank the teachers, speech-language pathologists, and other professionals that contributed to the study. We thank Associació del Trastorn Específic del Llenguatge de Catalunya (ATELCA) and all the schools that participated in the study. Especially, we thank the implication of Escola Santa Anna (Premià de Dalt) and the Centres de Recursos per a Deficients Auditius de Catalunya (CREDA) and the Equips d'Assessorament Psicopedagògic de Catalunya (EAP). Finally to Coral Pérez and Joan Tarrida who contributed to the data collection in Catalonia.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. The phonetic transcriptions in International Phonetic Transcription (IPA) of the nonwords for the Catalan version of the Nonword Repetition task (NRT-Cat) by syllable length and wordlikeness.

2-Syllable	3-Syllable	4-Syllable	5-Syllable	6-Syllable
		High wordlikeness nonwords		
Marlut	Grimpola	Palimatrenc	Balatenesta	Pomalimenasa
[mɐr. ˈlut]	[ɡɾim.ˈpa.lə]	[pɐ.li.mɐ.ˈtɾɛ^{ɲ}k]	[bɐ.lɐ.tə.ˈnɛs.ta]	[pʊ.mɐ.li.mɐ.ˈna.zɐ]
Calim	Llendira	Xipileta	Jolistarista	Desperdiculandre
[ka.ˈlim]	[ʎən.ˈdi.ɾɐ]	[ɕi.pi.ˈlɛ.tɐ]	[ʑʊ.lis.tɐ.ˈɾis.tə]	[dəs.pəɾ.ði.ku.ˈlan.dɾə]
Padot	Vifatull	Calisota	Xepunetura	Parmicaginosa
[pɐ.ˈðɔt]	[vi.fa.ˈtuʎ]	[kɐ.li.ˈsɔ.tə]	[ɕə.pʊ.nə.ˈtu.ɾɐ]	[pɐr.mi.kɐ.ʑi.ˈno.zə]
Galmet	Dubalet	Clomitura	Parantalaresc	Situcalinomi
[gɐl.ˈmɛt]	[du.βɐ.ˈlɛt]	[klu.mi.ˈtu. ɾa]	[pa.ɾən.tɐ.lɐ.ˈɾɛsk]	[si.tʊ.kɐ.li.ˈnɔ.mi]
		low Wordlikeness nonwords		
Pidop	Satompa	Becuradoc	Lamerquitorma	Diraculmestici
[pi.ˈðɔp]	[sɐ.ˈtum.pa]	[bɐ.kʊ.ɾɐ.ˈðɔk]	[la.mar.ki.ˈtor.mɐ]	[di.ɾɐ.kʊl.mas.ˈti.si]
Rolma	Xolopi	Daltrosqueti	Lifortamasuc	Purmesidocata
[ˈrol.ma]	[ɕʊ.ˈlɔ.pi]	[dɐl.tɾʊs.ˈkɛ.ti]	[li.for.tɐ.mɐ.ˈsuk]	[pur.mə.si.ðʊ.ˈca.tə]
Becra	Calmepi	Castretuma	Lemulticada	Casitedilafa
[ˈbɛ.kɾɐ]	[kɐl.ˈmɛ.pi]	[kɐs.tɾɐ.ˈtu.mə]	[lɐ.mul.ti.ˈka.ðə]	[ka.si.ta.di.ˈla.fɐ]
Dapa	Dalumi	Milusota	Zicuparamal	Garelisupota
[ˈðæ.pa]	[ða.ˈlu.mi]	[mi.lʊ.ˈsɔ.tɐ]	[zi.ku.pɐ.ɾɐ.ˈmal]	[gɐ.ɾɐ.li.su.ˈpɔ.tə]

Appendix B

Table A2. The phonetic transcriptions in International Phonetic Transcription (IPA) of the nonwords for the European Portuguese version of the Nonword Repetition Task (NRT-EPort) by syllable length and wordlikeness.

2-Syllable	3-Syllable	4-Syllable	5-Syllable	6-Syllable
High wordlikeness nonwords				
Naca	Lofena	Covilado	Melanifito	Turamisalato
[ˈna.kɐ]	[lu.ˈfe.nɐ]	[ku.vi.ˈla.du]	[mɨ.lɐ.nɨ.ˈfi.tu]	[tu.rɐ.mi.zɐ.ˈla.tu]
Fopa	Banita	Fenerade	Bonifadade	Detagapalico
[ˈfo.pɐ]	[bɐ.ˈni.tɐ]	[fɨ.nɨ.ˈɾa.dɨ]	[bɔ.nɨ.fɐ.ˈda.dɨ]	[dɨ.tɐ.gɐ.pɐ.ˈli.ku]
Trana	Praleta	Trapilado	Craletonina	Prinalvenioso
[ˈtrɐ.nɐ]	[pra.ˈle.tɐ]	[trɐ.pi.ˈla.du]	[krɐ.lɨ.tu.ˈni.nɐ]	[pri.ɐɫ.vɨ.ˈnjo.zu]
Prota	Bramato	Cravastado	Versatranista	Volturacidade
[ˈprɔ.tɐ]	[brɐ.ˈma.tu]	[krɐ.vɐʃ.ˈta.du]	[vɛɾ.sɐ.trɐ.ˈniʃ.tɐ]	[vɔɫ.tu.rɐ.si.ˈda.dɨ]
Low wordlikeness nonwords				
Cafo	Mafopa	Lemanado	Nocafozano	Rolinicistato
[ˈka.fu]	[mɐ.ˈfo.pɐ]	[lɨ.mɐ.ˈna.du]	[nɔ.kɐ.fu.zɔ.nʊ]	[rɔ.lɨ.nɨ.siʃ.ˈta.tu]
Tuma	Dopeta	Dilomopa	Lodanapito	Fatatuviricho
[ˈtu.mɐ]	[du.ˈpe.tɐ]	[di.lu.ˈmɔ.pɐ]	[lɔ.dɐ.na.ˈpi.tu]	[fɐ.tɐ.tu.vɨ.ˈɾi.ʃu]
Grapa	Gremata	Dragamato	Defermicato	Satopogatico
[ˈgra.pɐ]	[grɨ.ˈma.tɐ]	[drɐ.gɐ.ˈma.tu]	[dɨ.fɨɾ.mɨ.ˈka.tu]	[sɐ.to.po.ga.ˈti.ku]
Trila	Tramafa	Trafeleste	Promoflicada	Cremoforosada
[ˈtri.lɐ]	[trɐ.ˈma.fɐ]	[tra.fɨ.ˈlɛʃ.tʃɨ]	[prɔ.mɔ.fli.ˈka.dɐ]	[krɛ.mu.fu.ɾu.ˈza.dɐ]

References

1. Adams, A.-M.; Gathercole, S.E. Phonological Working Memory and Speech Production in Preschool Children. *J. Speech Lang. Hear. Res.* **1995**, *38*, 403–414. [CrossRef] [PubMed]
2. Coady, J.A.; Evans, J.L. Uses and interpretations of non-word repetition tasks in children with and without specific language impairments (SLI). *Int. J. Lang. Commun. Disord.* **2008**, *43*, 1–40. [CrossRef] [PubMed]
3. Estes, K.G.; Evans, J.L.; Else-Quest, N.M. Differences in the Nonword Repetition Performance of Children with and without Specific Language Impairment: A Meta-Analysis. *J. Speech Lang. Hear. Res.* **2007**, *50*, 177–195. [CrossRef]
4. Melby-Lervåg, M.; Lervåg, A. Oral Language Skills Moderate Nonword Repetition Skills in Children with Dyslexia: A Meta-Analysis of the Role of Nonword Repetition Skills in Dyslexia. *Sci. Stud. Read.* **2012**, *16*, 1–34. [CrossRef]
5. Laws, G. The use of nonword repetition as a test of phonological memory in children with Down syndrome. *J. Child Psychol. Psychiatry* **1998**, *39*, 1119–1130. [CrossRef] [PubMed]
6. Leonard, L.B. *Children with Specific Language Impairment*; MIT Press: Cambridge, MA, USA, 2014. [CrossRef]
7. Tager-Flusberg, H.; Cooper, J. Present and Future Possibilities for Defining a Phenotype for Specific Language Impairment. *J. Speech Lang. Hear. Res.* **1999**, *42*, 1275–1278. [CrossRef]
8. Tomblin, J.B.; Records, N.L.; Zhang, X. A System for the Diagnosis of Specific Language Impairment in Kindergarten Children. *J. Speech Lang. Hear. Res.* **1996**, *39*, 1284–1294. [CrossRef]
9. Adani, F.; Forgiarini, M.; Guasti, M.T.; Van Der Lely, H.K.J. Number dissimilarities facilitate the comprehension of relative clauses in children with (Grammatical) Specific Language Impairment. *J. Child Lang.* **2013**, *41*, 811–841. [CrossRef]
10. Bishop, D.V.M. Why is it so hard to reach agreement on terminology? The case of developmental language disorder (DLD). *Int. J. Lang. Commun. Disord.* **2017**, *52*, 671–680. [CrossRef]
11. Bishop, D.V.; Snowling, M.J.; Thompson, P.A.; Greenhalgh, T.; Catalise-2 Consortium; Adams, C.; Archibald, L.; Baird, G.; Bauer, A.; Bellair, J.; et al. Phase 2 of CATALISE: A multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *J. Child Psychol. Psychiatry* **2017**, *58*, 1068–1080. [CrossRef]
12. Lee, J.C.; Tomblin, J.B. Reinforcement learning in young adults with developmental language impairment. *Brain Lang.* **2012**, *123*, 154–163. [CrossRef] [PubMed]
13. Botting, N.; Conti-Ramsden, G. Non-word repetition and language development in children with specific language impairment (SLI). *Int. J. Lang. Commun. Disord.* **2001**, *36*, 421–432. [CrossRef]
14. Ellis Weismer, S.; Edwards, J. The role of phonological storage deficits in specific language impairment: A reconsideration. *Appl. Psycholinguist.* **2006**, *27*, 556. [CrossRef]
15. Evans, J.L.; Gillam, R.B.; Montgomery, J.W. Cognitive Predictors of Spoken Word Recognition in Children with and without Developmental Language Disorders. *J. Speech Lang. Hear. Res.* **2018**, *61*, 1409–1425. [CrossRef] [PubMed]

16. Bishop, D.V.M.; McArthur, G. Individual Differences in Auditory Processing in Specific Language Impairment: A Follow-Up Study using Event-Related Potentials and Behavioural Thresholds. *Cortex* **2005**, *41*, 327–341. [CrossRef]
17. Tallal, P.; Stark, R.E.; Mellits, D. The relationship between auditory temporal analysis and receptive language development: Evidence from studies of developmental language disorder. *Neuropsychologia* **1985**, *23*, 527–534. [CrossRef]
18. Brumbach, A.C.D.; Goffman, L. Interaction of Language Processing and Motor Skill in Children with Specific Language Impairment. *J. Speech Lang. Hear. Res.* **2014**, *57*, 158–171. [CrossRef]
19. Baddeley, A. Working Memory: Theories, Models, and Controversies. *Annu. Rev. Psychol.* **2012**, *63*, 1–29. [CrossRef]
20. Perrachione, T.K.; Ghosh, S.S.; Ostrovskaya, I.; Gabrieli, J.D.E.; Kovelman, I. Phonological Working Memory for Words and Nonwords in Cerebral Cortex. *J. Speech Lang. Hear. Res.* **2017**, *60*, 1959–1979. [CrossRef]
21. Delage, H.; Frauenfelder, U.H. Relationship between working memory and complex syntax in children with Developmental Language Disorder. *J. Child Lang.* **2019**, *47*, 1–33. [CrossRef]
22. Just, M.A.; Carpenter, P.A. A capacity theory of comprehension: Individual differences in working memory. *Psychol. Rev.* **1992**, *99*, 122–149. [CrossRef] [PubMed]
23. Alt, M. Phonological working memory impairments in children with specific language impairment: Where does the problem lie? *J. Commun. Disord.* **2011**, *44*, 173–185. [CrossRef] [PubMed]
24. Gathercole, S.E. Nonword repetition and word learning: The nature of the relationship. *Appl. Psycholinguist.* **2006**, *27*, 513–543. [CrossRef]
25. Munson, B.; Kurtz, B.A.; Windsor, J. The influence of vocabulary size, phonotactic probability, and perwordlikeness on nonword repetitions of children with and without specific language impairment. *J. Speech Lang. Hear. Res.* **2005**, *48*, 1033–1047. [CrossRef]
26. Munson, B.; Edwards, J.; Beckman, M.E. Relationships Between Nonword Repetition Accuracy and Other Measures of Linguistic Development in Children with Phonological Disorders. *J. Speech Lang. Hear. Res.* **2005**, *48*, 61–78. [CrossRef]
27. Barry, C.; Morrison, C.M.; Ellis, A.W. Naming the Snodgrass and Vanderwart pictures: Effects of age of acquisition, frequency, and name agreement. *Q. J. Exp. Psychol. Sect. A* **1997**, *50*, 560–585. [CrossRef]
28. Luce, P.A.; Pisoni, D.B. Recognizing Spoken Words: The Neighborhood Activation Model. *Ear Hear.* **1998**, *19*, 1–36. [CrossRef]
29. Mainela-Arnold, E.; Evans, J.L. Beyond capacity limitations: Determinants of word recall performance on verbal working memory span tasks in children with SLI. *J. Speech Lang. Hear. Res.* **2005**, *48*, 897–909. [CrossRef]
30. Mainela-Arnold, E.; Evans, J.L.; Coady, J. Beyond Capacity Limitations II: Effects of Lexical Processes on Word Recall in Verbal Working Memory Tasks in Children with and Without Specific Language Impairment. *J. Speech Lang. Hear. Res.* **2010**, *53*, 1656–1672. [CrossRef]
31. Adams, A.M. Phonological working memory and spoken language development in young children. *Q. J. Exp. Psychol. Sect. A* **1996**, *49*, 216–233. [CrossRef]
32. Girbau, D.; Schwartz, R.G. Phonological working memory in Spanish–English bilingual children with and without specific language impairment. *J. Commun. Disord.* **2008**, *41*, 124–145. [CrossRef] [PubMed]
33. Thordardottir, E.; Kehayia, E.; Mazer, B.; Lessard, N.; Majnemer, A.; Sutton, A.; Trudeau, N.; Chilingaryan, G. Sensitivity and Specificity of French Language and Processing Measures for the Identification of Primary Language Impairment at Age 5. *J. Speech Lang. Hear. Res.* **2011**, *54*, 580–597. [CrossRef]
34. Windsor, J.; Kohnert, K.; Lobitz, K.F.; Pham, G.T. Cross-Language Nonword Repetition by Bilingual and Monolingual Children. *Am. J. Speech Lang. Pathol.* **2010**, *19*, 298–310. [CrossRef]
35. Melby-Lervåg, M.; Lervåg, A.; Lyster, S.-A.H.; Klem, M.; Hagtvet, B.; Hulme, C. Nonword-Repetition Ability Does Not Appear to Be a Causal Influence on Children's Vocabulary Development. *Psychol. Sci.* **2012**, *23*, 1092–1098. [CrossRef] [PubMed]
36. Kalnak, N.; Peyrard-Janvid, M.; Forssberg, H.; Sahlén, B. Nonword Repetition—A Clinical Marker for Specific Language Impairment in Swedish Associated with Parents' Language-Related Problems. *PLoS ONE* **2014**, *9*, e89544. [CrossRef] [PubMed]
37. Radeborg, K.; Barthelom, E.; Sjöberg, M.; Sahlén, B. A Swedish non-word repetition test for preschool children. *Scand. J. Psychol.* **2006**, *47*, 187–192. [CrossRef] [PubMed]
38. Pham, G.; Ebert, K.D. Diagnostic Accuracy of Sentence Repetition and Nonword Repetition for Developmental Language Disorder in Vietnamese. *J. Speech Lang. Hear. Res.* **2020**, *63*, 1521–1536. [CrossRef] [PubMed]
39. Shaalan, S. Nonword Repetition Skills in Gulf Arabic–Speaking Children with Developmental Language Disorder. *J. Speech Lang. Hear. Res.* **2020**, *63*, 3700–3713. [CrossRef]
40. Dispaldro, M.; Leonard, L.B.; Deevy, P. Real-Word and Nonword Repetition in Italian-Speaking Children with Specific Language Impairment: A Study of Diagnostic Accuracy. *J. Speech Lang. Hear. Res.* **2013**, *56*, 323–336. [CrossRef]
41. Thordardottir, E. Language-Specific Effects of Task Demands on the Manifestation of Specific Language Impairment: A Comparison of English and Icelandic. *J. Speech Lang. Hear. Res.* **2008**, *51*, 922–937. [CrossRef]
42. Santos, F.H.; Bueno, O.F. Validation of the Brazilian Children's Test of Pseudoword Repetition in Portuguese speakers aged 4 to 10 years. *Braz. J. Med. Biol. Res.* **2003**, *36*, 1533–1547. [CrossRef] [PubMed]
43. Summers, C.; Bohman, T.M.; Gillam, R.B.; Peña, E.D.; Bedore, L.M. Bilingual performance on nonword repetition in Spanish and English. *Int. J. Lang. Commun. Disord.* **2010**, *45*, 480–493. [CrossRef] [PubMed]
44. Thordardottir, E.; Brandeker, M. The effect of bilingual exposure versus language impairment on nonword repetition and sentence imitation scores. *J. Commun. Disord.* **2013**, *46*, 1–16. [CrossRef] [PubMed]
45. Tucker, L.R.; Lewis, C. A reliability coefficient for maximum likelihood factor analysis. *Psychometrika* **1973**, *38*, 1–10. [CrossRef]

46. Weismer, S.E.; Tomblin, J.B.; Zhang, X.; Buckwalter, P.; Chynoweth, J.G.; Jones, M. Nonword repetition performance in school-age children with and without language impairment. *J. Speech Lang. Hear. Res.* **2000**, *43*, 865–878. [CrossRef] [PubMed]
47. Girbau, D. The Non-word Repetition Task as a clinical marker of Specific Language Impairment in Spanish-speaking children. *First Lang.* **2016**, *36*, 30–49. [CrossRef]
48. Girbau, D.; Schwartz, R.G. Non-word repetition in Spanish-speaking children with Specific Language Impairment (SLI). *Int. J. Lang. Commun. Disord.* **2007**, *42*, 59–75. [CrossRef] [PubMed]
49. Dollaghan, C.; Campbell, T.F. Nonword Repetition and Child Language Impairment. *J. Speech Lang. Hear. Res.* **1998**, *41*, 1136–1146. [CrossRef] [PubMed]
50. Ahufinger, N.; Ferinu, L.; Pacheco-Vera, F.; Sanz-Torrent, M.; Andreu, L. El trastorno del desarrollo del lenguaje (TDL) más allá de las dificultades lingüísticas: Memoria y atención. *Revista de Logopedia Foniatría y Audiología* **2020**, *41*, 4–16. [CrossRef]
51. Statistical Institute of Catalonia (Idescat). Survey on Language Uses of the Population (Catalan Initials: EULP) 2018. Available online: https://www.idescat.cat/indicadors/?id=anuals&n=10364&tema=cultu&lang=en (accessed on 9 September 2020).
52. Arnau, J.; Boada, H. Languages and school in Catalonia. *J. Multiling. Multicult. Dev.* **1986**, *7*, 107–122. [CrossRef]
53. Ferrer, F. Languages, Minorities and Education in Spain: The case of Catalonia. *Comp. Educ.* **2000**, *36*, 187–197. [CrossRef]
54. Alarcón, A.; Garzón, L.G. (Eds.) *Language, Migration and Social Mobility in Catalonia*; Brill: Leiden, The Netherlands, 2011.
55. Sanz-Torrent, M.; Badia, I.; Serra, M. *Contributions from Bilingual Specific Language Impairment in Catalan and Spanish to the Understanding of Typical and Pathological Language Acquisition*; Multilingual Matters: Bristol, UK, 2007; pp. 135–158. [CrossRef]
56. Sanz-Torrent, M.; Serrat, E.; Andreu, L.; Serra, M. Verb morphology in Catalan and Spanish in children with Specific Language Impairment: A developmental study. *Clin. Linguist. Phon.* **2008**, *22*, 459–474. [CrossRef] [PubMed]
57. Aguado, G.; Coloma, C.; Martínez, A.; Mendoza, E.; Montes, A.; Navarro, R.; Serra, M. Documento de consenso elaborado por el comité de expertos en TEL sobre el diagnóstico del trastorno. *Revista de Logopedia Foniatría y Audiología* **2015**, *35*, 147–149. [CrossRef]
58. Kaufman, A.S.; Kaufman, N.L. *Test Breve de Inteligencia Kaufman: K-BIT Spanish Adaptation*; Cordero-Pando, A., Calonge Romano, I., Eds.; Pearson: Madrid, Spain, 2004.
59. Wiig, E.H.; Semel, E.M.; Secord, A.W. *Clinical Evaluation of Language Fundamentals 4: Spanish Edition*; CELF-4-Spanish; Pearson: Madrid, Spain, 2006.
60. American National Standard Institute. *American National Standard: Methods for Calculation of the Speech Intelligibility Index*; Acoustical Society of America: Melville, NY, USA, 1997.
61. Wechsler, D. *Escala de Inteligência de Wechsler Para Crianças—Terceira Edição (WISC III)*; Cegoc: Lisbon, Portugal, 2003.
62. Gathercole, S.E.; Willis, C.S.; Baddeley, A.D.; Emslie, H. The children's test of nonword repetition: A test of phonological working memory. *Memory* **1994**, *2*, 103–127. [CrossRef] [PubMed]
63. Bailey, T.M.; Hahn, U. Determinants of Wordlikeness: Phonotactics or Lexical Neighborhoods? *J. Mem. Lang.* **2001**, *44*, 568–591. [CrossRef]
64. Aguado, G. Contribuciones Al Diagnóstico Del Trastorno Específico del Lenguaje Por Medio de la Repetición de Pseudopalabras. Prueba Y Baremos. 2005. Available online: http://dadun.unav.edu/bitstream/10171/19266/1/rpp%20WEB.pdf (accessed on 20 November 2020).
65. Alameda, J.R.; Cuetos, F. *Diccionario de Frecuencias de las Unidades Lingüísticas Del Castellano*; Universidad de Oviedo: Oviedo, Spain, 1995.
66. Nascimento, M.; Rodrigues, M.C.; Gonçalves, J.B. *Actas do XI Encontro Nacional da Associação Portuguesa de Linguística*; Colibri: Lisbon, Portugal, 1996.
67. Gomes, I. Ler E Escrever Em Português Europeu. Ph.D. Thesis, Universidad do Porto, Porto, Portugal, 2001.
68. Cunha, C.; Cintra, L. *Nova Gramática do Português Contemporâneo*; Editora Nova Fronteira: Rio de Janeiro, State of Rio de Janeiro, Brazil, 1985.
69. Martins, M.R. *Ouvir Falar: Introdução à Fonética do Português*; Editorial Caminho: Alfragide, Portugal, 1988.
70. Rigolet, S.A. *Para Uma Aquisição Precoce e Optimizada da Linguagem: Linhas de Orientação Para Crianças Até 6 Anos*; Porto Editora: Porto, Portugal, 2015.
71. Rigolet, S.A. *Os Três p Comunicação e Linguagem para Uma Plena Expressão*; Porto Editora: Porto, Portugal, 2009.
72. Sim Sim, I. *Avaliação da Linguagem Oral*; Fundação Calouste Gulbenkian: Lisboa, Portugal, 2014.
73. Da Silva, A.C. Até à Descoberta Do Princípio Alfabético. Master's Thesis, ISPA—Instituto Universitário de Ciências Psicológicas, Sociais e da Vida, Alfama, Portugal, 2003.
74. Grimm, L.G.; Yarnold, P.R. *Reading and Understanding Multivariate Statistics*; American Psychological Association: Washington, DC, USA, 1995.
75. Florkowski, C.M. Sensitivity, Specificity, Receiver-Operating Characteristic (ROC) Curves and Likelihood Ratios: Communicating the Performance of Diagnostic Tests. *Clin. Biochem. Rev.* **2008**, *29* (Suppl. 1), S83–S87. [PubMed]
76. Sackett, D.L.; Haynes, R.B.; Tugwell, P. *Clinical Epidemiology: A Basic Science for Clinical Medicine*; Little, Brown and Company: New York, NY, USA, 1985.
77. Haynes, R.B.; Sackett, D.L.; Guyatt, G.H.; Tugwell, P. How to do clinical practice research: A new book and a new series in the Journal of Clinical Epidemiology. *J. Clin. Epidemiol.* **2006**, *59*, 873–875. [CrossRef]
78. Gathercole, S.E.; Willis, C.; Emslie, H.; Baddeley, A.D. The influences of number of syllables and wordlikeness on children's repetition of nonwords. *Appl. Psycholinguist.* **1991**, *12*, 349–367. [CrossRef]

79. Gutiérrez-Clellen, V.F.; Simon-Cereijido, G. Using Nonword Repetition Tasks for the Identification of Language Impairment in Spanish-English-Speaking Children: Does the Language of Assessment Matter? *Learn. Disabil. Res. Pract.* **2010**, *25*, 48–58. [CrossRef]
80. O'Connell, M.A.; Basak, C. Effects of task complexity and age-differences on task-related functional connectivity of attentional networks. *Neuropsychologia* **2018**, *114*, 50–64. [CrossRef]

children

Article

Association between CCC-2 and Structural Language, Pragmatics, Social Cognition, and Executive Functions in Children with Developmental Language Disorder

Clara Andrés-Roqueta *, **Irene Garcia-Molina** and **Raquel Flores-Buils**

Department of Developmental, Educational, Social and Methodological Psychology, Universitat Jaume I de Castelló, 12071 Castelló de la Plana, Spain; imolina@uji.es (I.G.-M.); flores@uji.es (R.F.-B.)
* Correspondence: candres@uji.es

Abstract: (1) Background: Developmental Language Disorder (DLD) is diagnosed when the child experiences problems in language with no known underlying biomedical condition and the information required for its correct evaluation must be obtained from different contexts. The Children's Communication Checklist (CCC-2) covers aspects of a child's communication related to structural language and pragmatic skills, which are linked to social cognition or executive functions. The aim of this article is to examine parents' reports using the Spanish version of the CCC-2 questionnaire and its association with different formal assessments related to communication. (2) Methods: 30 children with DLD (3; 10–9 years old) and 39 age-matched (AM) children with typical development were assessed using formal measures of structural language, pragmatics, social cognition, and executive functions. Parents of children with DLD answered the Spanish version of the CCC-2. (3) Results: The performance of children with DLD was lower in all the formal assessments in comparison to AM children. The CCC-2 was significantly correlated with all the direct child assessments, although only formal measures of structural language predicted both the structural language and pragmatics scales of the CCC-2. (4) Conclusions: The CCC-2 answered by parents was consistent with formal assessments in children with DLD, and structural language seemed to be the best predictor of all the subscales.

Keywords: Children's Communication Checklist (CCC-2); Developmental Language Disorder (DLD); parents' reports; formal measures; pragmatics communicative profile; social cognition; executive functions

Citation: Andrés-Roqueta, C.; Garcia-Molina, I.; Flores-Buils, R. Association between CCC-2 and Structural Language, Pragmatics, Social Cognition, and Executive Functions in Children with Developmental Language Disorder. *Children* **2021**, *8*, 123. https://doi.org/10.3390/children8020123

Academic Editors: Eva Aguilar Mediavilla, Miguel Pérez Pereira, Elisabet Serrat-Sellabona and Daniel Adrover-Roig

Received: 31 December 2020
Accepted: 4 February 2021
Published: 9 February 2021

1. Introduction

Developmental Language Disorder (DLD) is diagnosed when the child may be experiencing problems in both the content (semantics) and the form of language (phonology, morphology, and syntax) with no known underlying biomedical condition such as brain injury or autism spectrum disorders, ASD [1]. It is important to highlight that DLD is a heterogeneous condition. Deficits in the acquisition and use of language can be noticed in the reduced vocabulary, limited structure of the sentences, and errors in discourse. The areas commonly affected in DLD (focusing on oral language) are phonology, syntax, word finding and semantics, discourse, verbal learning, and memory, but difficulties also occur in reading and writing. Children with DLD may have problems with language use in social interactions, i.e., pragmatics [2,3]. These problems can have secondary effects in other areas related to social contexts, such as socialization, communication, socioemotional problems, or academic achievement [4–6]. According to Bishop and Norbury [7], there is a particular group of children who have fluent, complex, and clearly articulated expressive language, despite showing severe problems in the way the language is used socially. This condition is commonly known as Pragmatic Language Impairment (PLI), or what has more recently been called Social Communication Disorder (SCD; Diagnostic and Statistical Manual of Mental Disorders, DSM-5 [8]; see Davies, Andrés-Roqueta and Norbury [2] for a discussion on whether pragmatic language skills of children with SCD may elucidate sources of pragmatic breakdown in other developmental populations such as autism). With all these difficulties, early diagnosis is crucial because children with DLD have a greater risk of experiencing poor social, emotional, and mental health outcomes [9–12], which in turn increases the probabilities of them becoming victims of bullying [13].

Information about children's language and communication development can be obtained in different ways: on the one hand, information about a child's performance in natural contexts can be provided by parents or teachers and, on the other hand, practitioners also assess children with different formal measures. Standardized tests are important because they are a valid way to find the profile of the individual and any possible deficits. In this sense, both reports and formal assessments are considered an important source of information for establishing a meaningful and contextualized diagnosis [14,15]. Nonetheless, it is worth mentioning that formal assessments require controlled situations, and this therefore represents a contextual limitation for professionals who do not have access to the information about the children's daily life in more natural contexts, such as at home or school. In these situations, children usually display more natural behaviors, which afford clinical observations and/or assessments [16].

There are different reports to be answered by parents that provide valuable information when children are not yet able to answer for themselves (e.g., Rossetti Infant Toddler Language Scale [17]) or, as in the case of the Social Communication Questionnaire (SCQ, [18]), that are widely available as screening tools; however, the SCQ was designed specifically for detecting risk for ASD, not for DLD. Most language tests or reports assess language structure and content, but are less well suited for evaluating how children use and interpret language or for identifying communication problems, e.g., poor turn-taking or over-literal comprehension [19,20].

In this regard, there is a widely used instrument that employs indirect measures to assess communication behaviors in children: the Children's Communication Checklist revised (CCC-2). This questionnaire is answered by parents, although it can be completed by any adult who has had regular contact with the child at least 3–4 days per week for at least 3 months, and it provides information about the language and communication profiles in different contexts of interaction (i.e., home or school). Originally, the CCC was designed by Bishop [21] to obtain a quick and easy evaluation measure of children with DLD, predominantly with pragmatic language impairment. Subsequently, the CCC was revised by Bishop in 2003 and standardized with 542 typically developing (TD) children, resulting in the CCC-2. One year later, the CCC-2 was validated with children with diagnoses such as DLD, PLI, and different levels of autism [3].

The CCC–2, with its 70 items spanning 10 subscales, was built to assess different measures of language and communication skills in children from 4 to 16 years of age [22]. Comparisons and results of the CCC-2 have usually been performed based on two overall scores: a general communication composite (GCC), which evaluates children's communication skills, and a social interaction difference index (SIDI), which specifically assesses children's pragmatic language. Bishop [22] demonstrated that children with DLD are expected to have difficulty with scales related to structural language, and to do relatively better with scales concerning pragmatics and autistic behavior. Moreover, they would be more likely to obtain a mean GCC below 100. Given the differences in the main composite scores, most studies have employed the CCC-2 as a useful questionnaire that discriminates children with clinical communication problems from typically developing children, as well as differentiating children with DLD from children with pragmatic language impairments. To our knowledge, no research has correlated the subscales with other formal measures or tests.

The CCC-2 was adapted and translated into different languages and went on to become an international tool. The Dutch version of the CCC-2 was used to distinguish the language profiles of children with ASD, DLD, and attention deficit hyperactivity disorder (ADHD) [23]. Moreover, the Dutch version of the CCC-2 added a new pragmatic score called the General Pragmatic Score (GenPragS), which is comparable to the Pragmatic composite score of the original CCC. However, the subscale (D) coherence was added, which was previously classified as a communication skill, but researchers realized that some items referred to pragmatics [20]. The Norwegian version of the CCC-2 successfully differentiated between a group of language-impaired and non-language-impaired children [24]. The Serbian version of the CCC-2 [25] was redefined by Glumbić and Brojčin [26] to obtain three subscales—General Communication Ability, Pragmatics, and Structural Language Aspects—that would distinguish between clinical samples (ASD, ADHD, DLD). The reliability and validity of the CCC-2 was also found in French in Quebec [27], Portuguese [28], and Spanish [29]. Finally, the Spanish version of the CCC-2 was also used to determine whether the CCC-2 was able to identify pragmatic profiles and discriminate between normative and clinical profiles such as significant language difficulties) or Down syndrome [30].

As mentioned above, the CCC-2 has been widely used with the aim of identifying different communication profiles in children with DLD, ADHD, and ASD [21,31–38]. However, the CCC-2 has also been used as a quick screening tool for identifying language disorders in children with sex chromosome trisomies [39], auditory processing disorder [40,41], schizophrenia [42], sleep problems [43], William syndrome [44], and in deaf children with cochlear implants [45]. However, none of these studies have used structural language, pragmatics, social cognition (SC, the ability to attribute mind to others and ourselves) or executive function (EF, a set of cognitive processes that are necessary for the cognitive control of behavior) tests as predictive measures of the CCC-2 scales, although they are considered key predictors of pragmatics [46].

It has been shown that the CCC-2 is a quick and reliable tool that allows professionals to gain a global vision of the problems of language and communication skills that the child may have. In this respect, although they are widely used, formal measures to assess children's sociocognitive and linguistic profile are usually time-consuming. However, the assessment of pragmatics is difficult and complex [47]. It is a complex area because it is related not only to the use of language in context, but also to SC [46]. Pragmatics and SC are related by definition, given that pragmatic ability underlies the capacity to use and interpret language appropriately in social situations. Studies using screening instruments and conversational analysis have reported pragmatic deficits among children with DLD [3]. Similarly to ASD (but with fewer empirical studies conducted in the area), research based on experimental tasks has described problems in specific areas of pragmatics such as understanding figurative language [48–50], sensibility to conversational maxims like "quantity" [51,52] and other Gricean maxims [53], using the preceding context to resolve ambiguous utterances and in narrative production [54,55], or understanding graphic humor [56]. Their pragmatic skills are usually in keeping with the levels of their structural language, as they perform as well as younger TD children matched for language level in experimental tasks [51]. Moreover, their level of SC also matters for these pragmatic problems when the tasks are socially oriented [57].

One of the most widely used SC measures is the Strange Stories task [58]. This task allows pragmatic impairment to be detected in children with DLD and ASD, because children must grasp the speaker's real intention by understanding his/her utterance (sometimes non-literal) and contextual aspects concerning the communicative situation [46]. The Strange Stories task involves stories about sarcasm, irony, and white lies, among other things. For example, in the ironic story, a character uses an utterance ("Well, that's very nice, isn't it!") to remark that a person is being rude. Thus, the participant must use his or her structural language competence to understand the literal meaning, and may understand the context and pragmatic maxims for the "hidden" intended meaning [51,57,59]. Interestingly, there is no empirical evidence from studies using the CCC-2 pragmatic subscales and establishing relationships with SC tasks.

Likewise, very little work has explored the relationships between pragmatics and EF in children with DLD. The rules of conversation change depending on the context in which they occur, and therefore they allow us to adapt our speech flexibly to the dynamic demands of the context by being flexible, tight, and effective [60]. Furthermore, maintaining a coherent reciprocal conversation requires paying attention and remembering what our speaker is saying (therefore using our attention and working memory skills). At the same time, we also need to inhibit excessive talking and ensure that our contribution is relevant (thus using our inhibition, organizing, monitoring, and planning skills) (see Green et al. [61] for a review).

The novelty of the present study is the fact that it has broken the CCC-2 down into simple scales (structural language, pragmatics, and autistic subscores) for associations and predictions with formal measures of structural language, pragmatics, SC, and EF. The present paper thus aims to examine whether parental information provided by the CCC-2 questionnaire is significantly associated with formal measures of structural language, pragmatics, EF, and SC in children with DLD. In this regard, the following hypotheses are stated:

Hypotheses 1 (H1). *First, the responses given by the parents are expected to show the language and communication problems of their children according to their age.*

Hypotheses 2 (H2). *Second, the CCC-2 scales (structural language, pragmatics, and autistic subscores) are expected to be associated with clinical-related tests (i.e., the scales of structural language will correlate to phonetics, syntax, and vocabulary tests; the pragmatic and autistic scales will correlate with SC, EF, and pragmatics) [51].*

Hypotheses 3 (H3). *Finally, formal measures are also expected to predict the CCC-2-related scales (i.e., the scales of structural language will correlate with the phonetics, syntax, and vocabulary tests; the pragmatic and autistic scales will correlate with SC, EF, and pragmatics).*

2. Materials and Methods

2.1. Participants

2.1.1. DLD Group

The DLD sample comprised 30 Spanish-speaking children (8 girls and 22 boys) diagnosed with DLD with ages between 3;10 and 9;0 years (mean age 70.50 months, range = 46–108). Children came from different ordinary schools in Spain.

- Diagnosis of DLD: Children with DLD had an updated diagnosis by a qualified educational psychologist from the ordinary schools where the sample was recruited. All these psychologists from the different schools belonged to the same local health services, and so they followed the same diagnostic protocols, thereby ensuring that the criteria used for all these diagnoses were homogeneous. In this regard, their records confirmed that the children had substantial language disability as the main cause for receiving speech and language therapy, while presenting a typical nonverbal intelligence based on standardized language and cognitive tests. Moreover, participants were recruited for the DLD group if they were native speakers of Spanish receiving speech and language therapy in the school at the time of the study; had language difficulties (discarding possible auditory disorders, neuro-sensory and intellectual disability); and finally, their clinical record had to be free of any medical condition that was likely to affect language, such as a diagnosis of ASD. Additionally, DLD condition was confirmed by the research group. The research team assessed each selected participant using two standardized grammar measures: on the one hand, Comprensión de Estructuras Gramaticales [62], and on the other hand, a Memory subtest on the Evaluación del Lenguaje Infantil, which is a sentence repetition task that measures expressive language ability and short-term auditory memory (ELI; [63]), both of which have been reported as a valid formal measure in the diagnosis of DLD (e.g., [64]). Participants were recruited only if they scored one standard deviation below the mean in at least one of these two tests. This threshold has been used in similar papers to conduct in-depth studies in children with DLD [52]. Moreover, Raven's Progressive

Matrices revised version [65] was used to ensure that the children's IQ was within the normal range (above the 15th percentile). In this sense, the DLD group obtained a mean of 28.57 (SD = 72.04, range = 25–99).

2.1.2. Chronological Age-Matched (AM) Group

The AM sample comprised 39 Spanish-speaking children (11 girls and 28 boys) aged between 3;7 and 9;00 years old (mean age 70.50 months, range = 43–109). The children came from the same ordinary schools as those in the DLD group. They were age- and gender-matched to children with DLD within ± 3 months of age. Furthermore, the educational psychologists were consulted to ensure that the children were not receiving speech and language therapy at the school when the study was being conducted or prior to it.

The AM group was introduced as a control to be compared with the DLD group on formal measures of structural language, pragmatics, SC, and EF. However, it was not used to compare the CCC-2 scores because: (1) This was not one of the main aims of the present study; and (2) The CCC-2 provides centile scores to compare a child with a normative mean, but the other measures (language, SC, and EF) are used with raw scores to be introduced in the correlational and predictive analyses. Therefore, the AM group was used to show whether the DLD group also had age-impaired performance on those measures.

2.2. Measures

2.2.1. Parent Reports

Children's Communication Checklist (CCC-2)

Parents of the participants with DLD answered the Spanish version of the CCC-2 (CCC-2—Spain/Spanish—Version 2 of 22 Jun 2012—MAPI Institute, provided by Pearson to the research group). The CCC–2 is used for children between the ages of 4;0 and 16;11 [19]. It is a 70-item questionnaire that caregivers rate using a 4-point numeric frequency scale ranging from 0 to 3, where 0 = less than once a week (or never); 1 = at least once a week, but not every day (or occasionally); 2 = once or twice a day (or frequently); 3 = several times (more than twice) a day (or always). It measures ten scales of communication, each with seven items: A. Speech, B. Syntax, C. Semantics, D. Coherence, E. Inadequate Initiation, F. Stereotyped Language, G. Context, H. Non-verbal Communication, I. Social Relations, and, J. Interests. The first four scales (A–D) measure structural language aspects; the four intermediate scales (E–H) measure pragmatic aspects, and the last two areas (I–J) measure autistic behavior. Internal consistency values between 0.66 and 0.80 and inter-rater reliability between parents and teachers ranged from 0.16 to 0.79 for the CCC-2.

Table 1 shows, by way of example, one item from each of the communication aspects measured in the CCC-2.

The CCC–2 produces the General Communication Composite (GCC) and the Social Interaction Difference Index (SIDI).

The GCC is a norm-referenced standard composite score that reflects overall communication skills (M = 100, SD = 15), and it is used to identify clinically significant communication problems. It is computed by adding scaled scores from Scales A through H.

The SIDI is an index score that reflects the difference between the structural language scales (A, B, C, and D) and the pragmatic language scales (E, H, I, and J), so it provides an index of mismatch between structural language skills and pragmatic and social skills [50]. SIDI scores ranging from -10 to 10 are considered typical, and scores within this range were obtained by 90% of the CCC–2 normative sample. Scores ≥ 11 suggest that speech/syntactic/semantic skills are deficient and relatively poorer than pragmatic skills, whereas scores ≤ -11 suggest that pragmatic language skills are deficient and relatively poorer than syntactic/semantic skills; this profile is associated with ASD [22].

However, in the present paper, the SIDI score was not used for correlations and the regression analysis, because it is used to ascertain the nature of an identified communication impairment and should, therefore, usually only be taken into account when the GCC is less than 55 (which ranged from 30 to 100 in the DLD in the present study, see Section 3) [19,66].

Moreover, other scores can be formed for more information by adding scales, such as structural language skills (A + B + C + D) [33,67,68]. Geurts [23] also proposed a second composite score that can be calculated as the sum of the scales D–H (GenPragS). This index was also used in the present study due to the fact that item (D) Coherence mixed aspects of structural language and also pragmatic skills.

Table 1. Examples of items from the communication aspects measured in the Children's Communication Checklist-2 (CCC-2).

Aspects of Communication	Sample Items
A. Speech	He/she speaks fluently and clearly, producing all speech sounds accurately and without hesitating.
B. Syntax	He/she produces long and complicated phrases like: "When we were in the park I went to see the ducks" or "I saw that man standing on the corner".
C. Semantics	He/she uses words that refer to classes of objects, rather than specific things; e.g., talks about the table, chair and drawers as "furniture", or calls bananas, apples and pears "fruit".
D. Coherence	It is difficult to know whether he/she is talking about something real or something invented.
E. Inappropriate initiation	He/she talks repeatedly about things that no-one is interested in.
F. Stereotyped language	When he/she answers a question, he/she provides sufficient and relevant information, without being overly precise if it is not necessary.
G. Context	His/her ability to communicate is different according to the situation. He/she may not have any trouble talking one-on-one with a familiar adult, but may find it difficult to express him/herself with a group of children of his/her own age.
H. Non-verbal communication	He/she doesn't look at people when he/she talks to them.
I. Social Relations	He/she hurts or disturbs other children without realizing it, unintentionally.
J. Interests	He/she leads the conversation towards his/her favorite topics, even when others are not interested.

In this regard, the three main subscales *Structural language* (A + B + C + D, scores ranging from 0 to 28), *Pragmatic* (D + E + F + G + H, scores ranging from 0 to 35), and *Autistic* (I + J, scores ranging from 0 to 14) composite scores were created for the correlational and predictive analyses by adding up the raw scores. It must be highlighted that the raw scores on the CCC-2 indicate the amount of difficulty on the different scales (in contrast to the scaled scores).

2.2.2. Formal measures

- Non-verbal reasoning

Raven's Colored Progressive Matrices scale was administered in order to have a non-verbal reasoning score for each DLD and AM participant [65]. Raw scores in this test range from 0 to 36. The standardization study yielded a value of 0.80 in test-re-test reliability [69].

- Structural language

Phonetics: Phonetics subtest from Evaluación de Lenguaje Infantil, ELI [63]. The phonetic subtest measures the level of articulation, the ability to imitate phonemes, and the difficulties to pronounce words containing "he studied" phonemes in initial, middle or final position. Children pronounced these words that could be represented as images or also by imitating adults' pronunciation. As the phonetics subtest is qualitative, the scoring has been processed emphasizing the type of errors that the child exhibited in the following way: No problems = 0; With substitutions in simple structures of (Consonant–Vowel) = 1; With mistakes in complex structures (Consonant–Consonant and Vowel–Vowel) = 2; With all the types of mistakes = 3. Scores were based on how complex the phonetic problems were: the more complex the problems were, the higher the score was. The maximum score was 3.

Receptive grammar: Comprensión de Estructuras Gramaticales, CEG [62]. This is an instrument designed to evaluate the participant's capacity to understand different types of grammatical constructions (grammatical comprehension) through drawings and sentences, with sentences of varying lengths and degrees of complexity. The CEG is a Spanish adaptation of the Test for Reception of Grammar, TROG [70], which assesses English grammar comprehension in children. Raw scores in this test range from 0 to 80.

Expressive grammar: Sentence recall subtest from Evaluación de Lenguaje Infantil, ELI [63]. This subscale was used to evaluate grammatical expressive skills. It measures the average sentence length and short-term auditory memory. Raw scores in this subtest range from 0 to 10.

Receptive vocabulary: Receptive Vocabulary subtest from Evaluación de Lenguaje Infantil, ELI [63]. This subscale measures the level of knowledge of vocabulary and receptive vocabulary. There are 30 sheets of increasing difficulty (from more concrete to more abstract semantic fields), and the child should point to the drawing that was asked for. Raw scores in this subtest range from 0 to 30.

Expressive vocabulary: Expressive Vocabulary subtest from Evaluación de Lenguaje Infantil, ELI [63]. This subscale measures access to vocabulary, expressive vocabulary, semantic knowledge, and information. It consists of 30 drawings, which the child must name (words from the close context in growing difficulty). Raw scores in this subtest range from 0 to 30.

Structural language composite score: These last five language measures were used to examine structural language as a complex construct. However, as these measures use different scoring ranges, a composite score was created. Raw scores, rather than standardized scores, were used because they are a direct indicator of how many correct responses each child achieved in each test. For phonetics, we created an inverse variable. The measures were weighted equally and combined to form the language composite score. This approach was taken to address issues arising from different scaling within language measures and between the other non-standardized measures used in the study. Specifically, the language composite variable was obtained by adding together all five linguistic raw scores. The final score ranged from 0 to 100, with each language measure representing 1/3 of the new composite score (1/3 phonetics, 1/3 grammar and 1/3 vocabulary).

- Pragmatics

The pragmatics subscale from the ELI battery [63] was used to obtain a formal assessment of functional communication. This subscale has both *receptive* items related to gesture–speech integration (e.g., the examiner tells the child a sentence and the child must decide if there is a discrepancy between statements and gestures), and *expressive* items related to figurative language understanding and politeness (e.g., the examiner asks questions about politeness, like "What do you say when somebody gives you a present?" or idiom questions like "What does 'you're a pig' mean?').

Pragmatic composite score: The two pragmatic measures (expressive and receptive) were used to examine pragmatics by adding the two scores (as stated in the ELI test). Raw scores in this subtest range from 0 to 14.

**A note about the use of ELI:* This test has two versions (Catalan and Spanish) and it can be administered in both languages. Nevertheless, the participants of the present study were assessed with the Spanish version since their mother tongue was Spanish. Furthermore, it must be noted that the ELI test is usually used for children from 3 to 6 years old. However, practitioners also use it for older children with DLD. In this sense, the research group selected ELI because the DLD sample included children from 3;10 to 9;00 years old, and so it allowed us to test smaller children (e.g., the Batería de Lenguaje Objetiva y Criterial - BLOC - test starts at 5 years old, and the Prueba de Lenguaje Oral Navarra - PLON - covers similar age ranges as ELI). Moreover, regarding the aims of the study, we were interested in raw scores to carry out the correlational and predictive studies.

*A note on the reliability of ELI and CEG

The ELI test is widely used to assess children's language. ELI has adequate Psychometric properties: Validity, correlation values: ELI Comprehensive Vocabulary–Peabody

($r = 0.75$); ELI Expressive Vocabulary–Kaufman Brief Intelligence Test (K-BIT) ($r = 0.85$); ELI Sentences– Wechsler Preschool and Primary Scale of Intelligence (WPPSI) Sentences ($r = 0.51$) (K-BIT [71]; WPPSI [72]). Reliability, split-half reliability and Test-Retest were used: Expressive-Vocabulary (0.83), Sentences (0.71), and Comprehensive–Vocabulary (0.70).

CEG has adequate Psychometric properties: Reliability: 0.91; Validity, correlations values: CEG-Peabody ($r = 0.809$, $p = 0.00$); and CEG– Illinois Test of Psycholinguistic Abilities (ITPA) ($r = 0.644$, $p = 0.00$) (Peabody, [73]; ITPA, [74]). Discrimination: more than half of the elements provide a discrimination index greater than 0.3 among subjects with higher scores and lower scores in the test [75].

- Social cognition

Strange Stories [58]. The aim of this task is to assess the understanding of other people's communicative intentions when non-literal language is used. Six of the original stories were used: pretense, lie, white lie, irony, joke, and idiom. The participants saw and listened to the Strange Story, and then, based on the reply, the scores were calculated as follows: 0 = inappropriate without mentalist aspects; 1 = inappropriate with mentalist aspects—distinct intention; 2 = correct with explicit aspects; and 3 = mentalist intention correct with expected intention [51]. Raw scores in this test range from 0 to 18. Individual mental state stories exhibited moderate to strong inter-rater reliability (M $\kappa = 0.82$, range = $0.79 \leq \kappa \leq 0.85$, all $p < 0.01$), as did individual control stories (M $\kappa = 0.84$, range = $0.74 \leq \kappa \leq 0.95$, all $p < 0.01$) [76].

- Executive Functions

The *Matching Familiar Figures Test* (MFFT; [77]) was administered to measure sustained attention and inhibitory skills. The MFFT consists of 12 items where the children were shown a picture (person or object) and six similar stimuli and the children were required to pick the picture that was identical to (e.g., that matched) the person/object given. Two variables were provided: the total number of errors committed until the correct one was found (*sustained attention*), and the mean latency prior to the first response (*response latency*). Correlations were 0.91 for latency and 0.89 for errors [77].

EF composite score: The two measures (sustained attention and response latency) were used to examine executive functions. Nevertheless, as these measures have different scoring ranges, a composite score was created. For sustained attention, we created an inverse variable because it indicates the number of errors committed (and not capacity), whereas response latency indicates the number of seconds before the first response (capacity to inhibit first response). The measures were weighted equally and combined to form the EF score. The final score ranged from 0 to 100 with each EF measure representing 1/2 of the new composite score 1/2 sustained attention, 1/2 response latency).

2.3. Procedure

Permissions were requested from the Regional government and school authorities to select the ordinary schools that children with DLD attended. Four schools agreed to take part in this study. Parents were then informed about the aims of the study, and written consent to participate was obtained.

Each child was assessed with the study instruments by the research group during two 40 min sessions (approx.) in a quiet room provided by the school. Tasks were administered individually, in random order. In parallel, parents were interviewed individually by the research group, and they completed the CCC-2 in the same schools, one hour before they picked their children up from school.

The original sample included in this study included 35 parents. One was lost because those parents failed to answer the questionnaire and four of them did not pass the Consistency check, which shows that the respondent has understood how to complete the CCC-2 with regard to positively and negatively formulated items [19].

2.4. Data Analysis

Data analysis was conducted using the statistical package SPSS (version 27). When the sample was subdivided into two groups (DLD and AM), the data failed the Shapiro–Wilk test of normality, showing several unequal variances across groups for the different scores. Moreover, there was a different number of participants in each group. Therefore, Mann–Whitney U (two-tailed, significance threshold of 0.05) was used to examine differences between groups on key measures. Effect sizes of group comparisons were calculated using r with the formula: $r = (z)/(\sqrt{N})$, because according to Fritz, Morris, and Richler [78], when between-group comparisons are performed with Mann–Whitney U tests, size effects must be calculated using "*r*" and not "d", where a value of 0–0.1 is considered a small effect; 0.2–0.4 is considered a medium effect; and 0.5–1 is considered a large effect.

Moreover, zero-order nonparametric correlations (Spearman) between key measures were conducted within the DLD groups. Finally, in order to further investigate the contribution of age, structural language skills, social pragmatics, linguistic pragmatics, and executive functions on the CCC-2 scores, a hierarchical linear regression analysis was conducted for the whole clinical sample (n = 30), and a bootstrapping method was implemented using 2000 bootstrap samples to derive robust estimates of standard errors, confidence intervals, and p values of the regression model.

It must be stated that the AM group was only used to compare the DLD group in the scores from the formal measures, but not in the CCC-2. This was because, as explained in the Participants section, the main aim of the present study was not to see the association of the CCC-2 with the related variables in typically developing children. In this sense, correlations and regressions are only carried out within the group of children with DLD and not among the AM children.

3. Results

3.1. Descriptive Statistics and between-Group Comparisons on Formal Measures

Table 2 reports the descriptive statistics of the four groups for grammar, age, structural language, pragmatics, SC, and EF formal measures.

Table 2. Descriptive statistics of Developmental Language Disorder (DLD) and chronological Age-Matched (AM) groups and between-group comparisons on formal measures.

	DLD (n = 30)		**AM (n = 39)**		***U***	***p***	***r***
	M (SD)	**Range**	**M (SD)**	**Range**			
Age and Gender							
Months	70.50 (17.86)	46–108	68.62 (14.26)	43–109	560.5	0.767	
Gender (M:F)	22:08		28:11:00		-	-	
Structural language measures							
Phonetics (range: 0–3)	1.07 (0.91)	0–3	0 (0)	0–0	156	0	0.761
Receptive grammar (range: 0–80)	50.43 (10.62)	17–68	64.69 (7.56)	48–77	146	0	0.682
Expressive grammar (range: 0–10)	5.13 (1.83)	2–8	7.97 (1.31)	4–9	126	0	0.64
Receptive vocabulary (range: 0–30)	21.03 (6.22)	5–30	23.92 (4.91)	10–29	390	0.018	0.285
Expressive vocabulary (range: 0–30)	18.47 (6.31)	7–27	22.64 (6.20)	10–30	351	0.005	0.341
Structural language	53.71 (16.62)	14.54–75.53	75.62 (6.24)	60.62–83.7	89	0	0.722

Table 2. *Cont.*

	DLD (n = 30)		AM (n = 39)		U	p	r
	M (SD)	Range	M (SD)	Range			
Pragmatics							
Receptive pragmatics (range: 0–6)	4.8 (1.32)	1–6	5.59 (0.91)	2–6	332.5	0.001	0.414
Expressive Pragmatics (range: 0–8)	2.8 (1.91)	0–8	5.10 (1.77)	1–8	213.5	0	0.547
Pragmatics (range: 0–14)	7.6 (2.66)	3–14	10.69 (2.33)	5–14	221.5	0	0.533
Social cognition measure							
Speaker's intention (range: 0–18)	4.8 (3.68)	0–12	10.87 (3.87)	1–16	152.5	0	0.632
Executive function measures							
Sustained attention	22.14 (12.58)	7–53	15.74 (7.14)	4–32	405	0.046	0.24
Response latency	7.3 (5.71)	1.71–20.30	15.40 (12.57)	2.82–75.15	260	0	0.456
Executive function	33.53 (14.64)	2.28–56.75	44.67 (12.56)	20.3–82.46	323	0.003	0.362

Note: *Raw scores* on all measures; *Age* = chronological age (months); *Structural Language* = Language composite score; *Pragmatics* = Pragmatic composite score; *Executive function* = Executive function composite score.

The Mann–Whitney U-test showed that the groups did not differ in age (U = 560.500, $p < 0.001$, r = 0.036) and small effect sizes were observed. Moreover, a Chi-squared test showed that groups did not differ according to gender ($\chi^2 = 0.20$, p = 0.887).

As regards the formal measures (raw scores) of structural language, pairwise comparisons revealed that the DLD and AM groups differed in all the measures: phonetics, expressive grammar, receptive grammar, receptive vocabulary, expressive vocabulary, and also in the structural language composite score. Large-size effects were observed for grammar and phonetics measures, as well as for the composite score, whereas medium-size effects were observed for vocabulary measures.

Regarding pragmatic scores, between-group comparisons revealed differences between the DLD and the AM groups: receptive pragmatics, expressive pragmatics, and pragmatics composite score. Large-size effects were observed for all cases.

Regarding the social cognition measure, between-group comparisons revealed differences between the DLD and the AM groups, and large-size effects were observed.

Finally, turning to the executive function measures, again, between-group comparisons showed that the AM group achieved a significantly lower performance than the DLD group: sustained attention, latency of response, and EF composite variable, and large- and medium-size effects were observed.

3.2. Descriptive Statistics on CCC-2

Table 3 reports the descriptive statistics on the CCC-2 for the DLD group.

As regards the GCC score (scaled scores), the DLD group obtained a mean of 67.53, which is almost the cutoff range of language impairments.

As regards the SIDI index (scaled scores), the DLD group obtained a mean of 10.07, which is a value that is included within normal limits (the normal range of SIDI scores is from −10 to 10).

As regards Structural language (scaled scores), the DLD group obtained a mean of 28.57; as regards Pragmatics (scaled scores), the DLD group obtained a mean of 28.57; and as regards the Autistic index (scaled scores), the DLD group obtained a mean of 18.53.

Individual scores demonstrate that the DLD group showed better performance on scales E, F; then on D, G, H; then I and C; and the lowest means were obtained on A and B.

Table 3. Descriptive statistics of Developmental Language Disorder (DLD) participants on Children's Communication Checklist-2 (CCC-2) scores.

	DLD (n = 30)	
	M (SD)	**Range**
GCC score	67.53 (16.71)	30–100
SIDI index	10.07 (9.28)	−7–39
Structural Language (A + B + C + D) scaled score	28.57 (8.07)	14–49
Pragmatics (D + E + F + G + H) scaled score	47.93 (12.45)	28–73
Autistic index (I + J) scaled score	18.53–5.91	11–35
A. Speech	6.47 (3.39)	0–13
B. Syntax	5.6 (3.05)	0–12
C. Semantics	7.53(2.36)	4–15
D. Coherence	8.97 (3.2)	5–16
E. Inappropriate initiation	11.37 (4.06)	4–21
F. Stereotyped language	10.37 (3.38)	5–16
G. Context	8.50 (3.19)	4–20
H. Non-verbal communication	8.73 (3.99)	3–21
I. Social relations	7.37 (4.09)	2–21
J. Interests	11.17 (4.55)	1–22
Structural Language (A + B + C + D) raw score (range: 0–28)	21.37 (12.54)	2–52
Pragmatics (D + E + F + G + H) raw score (range: 0–35)	21 (12.14)	5–45
Autistic index (I + J) raw score (range: 0–14)	8.20 (3.75)	3–18

Note: scaled scores on all measures except for the last three measures.

3.3. Correlations between CCC-2 Scores and Age, Structural Language, Linguistic Pragmatics, Social Pragmatics and Executive Functions

Zero-order nonparametric correlations (Spearman) between key measures are presented in Table 4.

Table 4. Zero-order nonparametric Spearman correlations between formal measures and Children's Communication Checklist-2 (CCC-2) scores within the Developmental Language Disorder (DLD) group (n = 30).

	Age	**St. Lang**	**Prag**	**SC**	**EF**	**St. Lang CCC-2**	**Prag CCC-2**
Age							
St. lang	0.624 **						
Prag	0.382 *	0.557 **					
SC	0.292	0.434 *	0.384 *				
EF	0.563 **	0.593 **	0.552 **	0.250			
St. lang CCC-2	−0.705 **	−0.703 **	−0.445 *	−0.354	−0.685 **		
Prag CCC-2	−0.231	−0.442 *	−0.401 *	−0.457 *	−0.510 **	0.663 **	
Aut CCC-2	−0.276	−0.444 *	−0.219	−0.280	−0.472 **	0.568 **	0.527 **

Note 1: * $p < 0.05$; ** $p < 0.01$; Note 2: all scores are raw scores; Note 3: Age = chronological age (months); St. lang = Language composite score (formal measures); Prag = Pragmatic composite score (formal measures); SC = Social Cognition (formal measure); EF = Executive functions composite score (formal measures); St. lang CCC-2 = A + B + C + D raw scores on CCC-2; Prag CCC-2 = D + E + F + G + H raw scores on CCC-2; Aut CCC-2 = I + J raw scores on CCC-2.

First, the three scores derived from the CCC-2 proved to be positively and strongly correlated: Structural Language–Pragmatics ($p < 0.001$); Structural Language–Autistic index ($p = 0.001$); Pragmatics–Autistic index ($p = 0.003$).

Strong and negative correlations were observed between Structural Language measured with the CCC-2 and age ($p < 0.001$), the structural language composite score based on formal measures ($p < 0.001$), and the executive functions composite score ($p < 0.001$). Moreover, a medium and negative correlation was observed with pragmatics ($p = 0.014$). However, it was not correlated with social cognition ($p = 0.055$). The correlations are negative because raw scores on the CCC-2 indicate the degree of difficulty on the different scales (in contrast to scaled scores).

For the Pragmatic measure of the CCC-2, a negative and non-significant correlation was observed with age ($p = 0.219$). Medium and negative associations were observed with formal measures of structural language ($p = 0.014$), pragmatics ($p = 0.028$), social cognition ($p = 0.011$), and executive functions ($p = 0.005$). Again, the correlations are negative because raw scores on the CCC-2 indicate the degree of difficulty on the different scales.

Finally, for the Autistic measure of the CCC-2, a negative and non-significant correlation was observed with formal measures of age ($p = 0.139$), pragmatics ($p = 0.246$), and social cognition ($p = 0.134$). Medium and negative associations were observed with formal measures of structural language ($p = 0.014$) and executive functions ($p = 0.010$). Again, the correlations are negative because raw scores on the CCC-2 indicate the degree of difficulty on the different scales.

3.4. Predictive Analysis of the CCC-2 Scores of Structural Language, Pragmatics and Autistic Indexes

As shown in the correlation analyses, several variables were correlated in the DLD group, making it difficult to identify the independent contribution each of them makes to the CCC-2 measures. To further investigate the contribution of these variables, three hierarchical linear regression analyses were conducted for the DLD group ($n = 30$).

Structural Language (CCC-2: A + B + C + D), Pragmatics (CCC-2: D + E + F + G + H), and Autistic score (CCC-2: I + J) were the outcome variables in the regression, and five predictor variables were entered in the following order (Tables 5 and 6): age was entered first, as raw scores had been used, and also because there was an important age difference between some participants in the sample (ranging from 3 to 9 years old). The Structural Language composite score (formal measure) was entered next, because structural language deficits are fundamental factors for the pragmatic deficits observed in children with DLD (e.g., [51]). Pragmatic and social cognition scores were entered after structural language to investigate whether they make a specific contribution to each CCC-2 score when structural language skills have been taken into account. Finally, the executive function composited variable was introduced in the final stage, to check whether other features related to sustained attention and latency of response are relevant.

Table 5. Summary of regression coefficients for Structural Language CCC-2 scores (A + B + C + D) within the DLD group (bootstrap results based on 2000 bootstrap samples).

Structural Language CCC-2 (A + B + C + D)						
Predictor	**R^2 Adjusted**	**B**	**B 95% CI (LL, UL)**	**SE B**	**t**	***p***
Step 1	0.369					
Constant		51.672	[36.841, 67.801]		6.592	0.001
Age		−0.427	[−0.635, −0.248]	−0.607	−3.972	0.003
Step 2	0.226					
Constant		56.748	[42.854, 70.537]		8.680	0.000
Age		−0.141	[−0.305, 0.149]	−0.200	−1.216	0.167
St. Language		−0.472	[−0.848, −0.208]	−0.626	−3.807	0.013
Step 3	0.026					
Constant		59.296	[44.281, 72.738]		8.804	0.000
Age		−0.137	[−0.292, 0.202]	−0.196	−1.206	0.192
St. Lang		−0.395	[−0.832, −0.162]	−0.524	−2.917	0.031
Prag		−0.907	[−2.346, 0.570]	−0.192	−1.312	0.214

Table 5. *Cont.*

Structural Language CCC-2 (A + B + C + D)						
Predictor	**R^2 Adjusted**	**B**	**B 95% CI (LL, UL)**	**SE B**	**t**	***p***
Step 4	0.003					
Constant		59.018	[43.206, 73.028]		8.581	0.000
Age		−0.136	[−0.291, 0.202]	−0.194	−1.176	0.203
St. Language		−0.381	[−0.833, −0.133]	−0.505	−2.685	0.042
Prag		−0.852	[−2.580, 0.661]	−0.181	−1.193	0.300
SC		−0.207	[−1.200, 0.825]	−0.061	−0.430	0.653
Step 5	0.018					
Constant		57.804	[42.260, 72.465]		8.311	0.000
Age		−0.092	[−0.254, 0.323]	−0.131	−0.752	0.453
St. Language		−0.356	[−0.839, −0.109]	−0.472	−2.480	0.060
Prag		−0.568	[−2.315, 1.163]	−0.121	−0.747	0.518
SC		−0.219	[−1.043, 0.838]	−0.064	−0.455	0.648
EF		−0.159	[−0.599, 0.104]	−0.183	−1.061	0.356

Note 1: all are raw scores; Note 2: Age = chronological age (months); St. lang = Language composite score (formal measures); Prag = Pragmatic composite score (formal measures); SC = Social Cognition (formal measure); EF = Executive functions composite score (formal measures).

Table 6. Summary of regression coefficients for Pragmatic CCC-2 scores (D + E + F + G + H) within the DLD group bootstrap results based on 2000 bootstrap samples).

Pragmatic CCC-2 (D + E + F + G + H)						
Predictor	**R^2 Adjusted**	**B**	**B 95% CI (LL, UL)**	**SE B**	**t**	***p***
Step 1	0.071					
Constant		34.013	[16.291, 52.549]		3.725	0.003
Age		−0.180	[−0.432, 0.055]	−0.266	−1.435	0.155
Step 2	0.155					
Constant		38.048	[21.054, 55.890]		4.385	0.001
Age		0.048	[−0.282, 0.361]	0.071	0.313	0.755
St. Language		−0.375	[−0.746, −0.003]	−0.518	−2.280	0.037
Step 3	0.060					
Constant		41.751	[21.742, 58.248]		4.703	0.001
Age		0.052	[−0.242, 0.379]	0.078	0.349	0.707
St. Lang		−0.264	[−0.711, 0.053]	−0.364	−1.477	0.138
Prag		−1.318	[−3.045, 0.789]	−0.291	−1.447	0.158
Step 4	0.054					
Constant		40.602	[21.339, 58.148]		4.639	0.001
Age		0.057	[−0.255, 0.357]	0.085	0.387	0.676
St. Language		−0.204	[−0.664, 0.208]	−0.282	−1.130	0.295
Prag		−1.092	[−2.960, 1.203]	−0.241	−1.202	0.295
SC		−0.855	[−2.391, 0.538]	−0.262	−1.395	0.272
Step 5	0.030					
Constant		39.076	[18.758, 56.919]		4.412	0.003
Age		0.112	[−0.191, 0.444]	0.166	0.718	0.433
St. Language		−0.172	[−0.648, 0.191]	−0.238	−0.944	0.377
Prag		−0.735	[−2.722, 1.758]	−0.162	−0.759	0.515
SC		−0.870	[−2.401, 0.461]	−0.266	−1.421	0.252
EF		−0.200	[−0.864, 0.167]	−0.240	−1.047	0.346

Note 1: all are raw scores; Note 2: Age = chronological age (months); St. lang = Language composite score (formal measures); Prag = Pragmatic composite score (formal measures); SC = Social Cognition (formal measure); EF = Executive functions composite score (formal measures).

For Structural Language (CCC-2), the general model was significant and accounted for 64% of the variance, F (5, 23) = 8.225, $R^2 = 0.641$, $p < 0.001$ (see Table 5). Higher Structural Language scores on the CCC-2 were negatively and significantly associated with Age, which explained 37% of the variance, and higher structural language scores of formal measures, which explained 23% of the variance. No single association was found with the formal assessments on pragmatics, social cognition, or executive functions.

For Pragmatics (CCC-2), the general model was significant and accounted for 37% of the variance, F (5, 23) = 2.692, $R^2 = 0.369$, $p = 0.047$ (see Table 6). Higher Pragmatic scores on the CCC-2 were negatively and significantly associated with structural language formal measures, which explained 16% of the variance. No single association was found with the formal assessments on age, pragmatics, social cognition or executive functions.

Finally, for the Autistic index (CCC-2), the general model was not significant: F (5, 23) = 1.296, $R^2 = 0.220$, $p = 0.300$. In this sense, no single association was found with formal measures of age, structural language composited score, pragmatics, social cognition or executive functions.

4. Discussion

The present study attempted to determine whether the Spanish version of the CCC-2, applied to parents of children with DLD, agrees with clinical information when linguistic, pragmatic, EF, and SC areas are assessed through direct measures. Furthermore, this information from parents allows us to better understand the children's problems, in a more detailed and contextualized way than with only the results of direct measures.

Regarding the first hypothesis, the responses given by parents were expected to allow us to better understand the problems of language and communication that children with DLD have. First, all the formal measures discriminated between the two groups (DLD–AM), demonstrating that the performance of those with DLD was lower than that of their AM peers in structural language, pragmatics, SC, and EF measures, in line with previous studies (for structural language and pragmatics: [46,49–54]; for SC [46,51,57,59]; and for EF [60,61]). Also, regarding previous findings using other versions of the CCC-2 (e.g., Norwegian sample), a cutoff at or below a scaled score of 64 on the GCC was selected to identify children with language impairments [24]. Our sample was closer to the cutoff, which would corroborate their problems in structured language. Moreover, taking a closer look at the scales, their lower means were found in (A) Speech and (B) Syntaxis. However, the children with DLD obtained a positive score (within normal limits) in the SIDI, indicating no disproportionate pragmatic impairments according to their structural language abilities. Again, a more detailed observation shows that the best scores of the children with DLD were found on the pragmatic scales (E > F > D > G > H).

With regard to Hypothesis 2, the CCC-2 scales were associated with the different clinical-related tests (formal measures). In this sense, Structural Language (CCC-2) was correlated with age, structural language, pragmatics, and EF (but not with SC); Pragmatics (CCC-2) was correlated with structural language, pragmatics, SC, and EF (but not with age); and the Autistic index (CCC-2) was correlated with structural language and EF (but not with age, pragmatics, or SC). All the CCC-2 correlations were in line with what was expected for Structural Language and Pragmatic scores, but not for the Autistic index, which was expected to also correlate with SC. SC and Pragmatics would go hand in hand for the complete understanding of the contextual aspects around the communicative situation [46]. In the same sense, EF would seem to be crucial to remember the interlocutor's information, monitoring, and planning the speech or to know when it would be better to initiate a conversation [60,61].

It is important to note the significant correlations found on comparing parents' opinions and the clinical measures. In this regard, it was found that parents' and professionals' criteria ran in the same direction, which is useful for detecting and recognizing their children's language and communication handicaps. It should be highlighted that parents' information would be useful not only in the aspects referring to the structural language (phonology, morphosyntax, and semantics), but also in the pragmatics areas, which are less easily measurable in the clinical setting (inadequate initiation, context, non-verbal communication). Thus, it is concluded that the parents' opinions agree with the clinical assessments.

Together, these findings identify structural language, pragmatics, SC, and EF as key skills for communication in the DLD group.

Finally, in Hypothesis 3, results corroborated the idea that formal measures predicted the CCC-2 scales related to structural language and pragmatics. On the one hand, age and structural language predicted the structural language index (CCC-2), which makes sense as the errors in phonetics, syntax and vocabulary affect the structural language index (CCC-2), and, in the same line, the formal measures of structural language. On the other hand, structural language predicted the Pragmatic Index (CCC-2), but not age or formal measures of pragmatics, EF, and SC. Again, this could be explained by the need to use the structural language competence to understand the literal meaning at first to fully comprehend the hidden intention or the context where the utterance is made [51,57,59]. Finally, none of the formal measures predicted the Autistic Index, which could be explained by the fact that the sample consisted only of children with DLD (non-autistic children) with specific difficulties in structural language and pragmatics, rather than SC and EF.

Specifically regarding pragmatic competence, between-group comparisons confirmed that children with DLD faced significant difficulties on the formal measures (including structural language, pragmatics, SC, and EF tasks) compared with their AM peers. However, difficulties with pragmatics were in keeping with the participants' structural language (but not SC or EF). This issue emphasizes the role of structural language in pragmatic development for children with DLD, obtaining similar findings to those of previous research (e.g., [2,49,50]; for a complete review, see [46].

As a limitation, we must be cautious with our results. First, a larger sample size would be needed in future studies to be able to conduct a meticulous study of the overlap between structural language, SC, EF, and pragmatics. Moreover, most of the participants in the present study are still developing some of the skills measured, so the results of the present study are applicable to children with DLD from 4 to 9 years old, but not for older ones. In this sense, most of the studies cited have assessed samples of older children with DLD, and pragmatic problems are perhaps more salient and best predicted by SC or EF, instead of only structural language. Furthermore, the Spanish language entails greater difficulties in morphological inflections compared to English (gender and number concordance, or verb conjugations). Other cultural influences that are more related to pragmatic aspects could exert an influence (e.g., accepted distance between interlocutors, improper initiation, or turn-taking behavior accepting more interruptions). Finally, there is a need for further research on the exact characterization of the different pragmatic skills and whether they depend on the phenomenon under study (e.g., figurative understanding, gesture–speech integration or understanding irony), on the structural aspects linked to understanding the linguistic context, on the SC strategies of the listener, or on the EF skills to maintain and initiate communication with other people [46,51]. In this respect, after examining the results, training of structural language skills seems essential for the development of pragmatic skills in children with DLD between 4 and 9 years of age. Furthermore, it is important to remember that the pragmatics subtest of ELI may not be enough to measure the child's pragmatic competences in an exact and complete way. In fact, it is a quick measure that can give us an initial idea. However, this subtest cannot be compared with pragmatic tests in which the evaluation is based on natural observation, through the behavior of the participants in a specific context. In Spanish, professionals and practitioners need more complex formal batteries to assess pragmatic competences in a comprehensive way.

5. Conclusions

In sum, the results of the present study responded to our main goal. The parents' reports on the CCC-2 were consistent with the professionals' formal evaluations. Another conclusion that could be drawn is that the extent and underlying causes of general communication difficulties of children with DLD correlate not only on the children's competence with structural language and pragmatics, but also on SC and EF. Nevertheless, structural language seems to be the best predictor of all the subscales measured with the CCC-2.

It is important to highlight that, on the one hand, parents' responses have been seen as important and complementary cues to complete the important information about children's communicative skills in different contexts. Furthermore, it should be highlighted that the CCC-2 is an informative, fast, and cost-effective tool to measure and anticipate language impairments in preschool and school-age children. Since parents can participate, clinicians can have access to the children's daily life in a more natural context than in a clinical setting. It should be added that, according to our data, the information provided by the parents seems to be precise in structural language aspects (the most visible in communication), but they do not seem to be aware of the actual pragmatic implications/difficulties, as some pragmatic skills develop during later childhood. In this regard, clinicians should make parents aware of these difficulties and help them with guidelines for intervention in pragmatic aspects. Future research should investigate the use of these kinds of tools with larger samples, while also adding more sophisticated items to address pragmatic components, given that they are more difficult to evaluate in designs involving children.

Moreover, as demonstrated in the present study, structural language skills affect the general communication of a child with DLD. Therefore, focusing only on pragmatics without taking into account other structural language will not reveal the actual communication needs of a child with DLD and might result in treatment goals that are too tightly defined [34].

Consequently, multi-disciplinary assessments of the communication profile of a child with DLD are necessary to design an adapted and individualized intervention (e.g., include formal assessments together with parents' reports on aspects of communication). Moreover, it seems very important to include structural language contents in interventions that aim to improve pragmatic competence [51]. However, it would also be important to include aspects of SC (e.g., regarding the understanding of a speaker's intention) and EF (e.g., regarding attention to cues from context, or impulsive or quick responses to the speaker) to improve other social communication skills that have not been addressed with the CCC-2 (e.g., understanding irony). Early diagnosis of pragmatic difficulties is crucial due to the fact that children with DLD are at greater risk of experiencing poor social, emotional, and mental health outcomes [9–12], which increases the probabilities of them being the victim of bullying [13].

Author Contributions: Conceptualization, C.A.-R. and I.G.-M.; methodology, C.A.-R., I.G.-M. and R.F.-B.; formal analysis, C.A.-R., I.G.-M. and R.F.-B.; investigation, C.A.-R., R.F.-B. and I.G.-M.; resources, C.A.-R. and R.F.-B.; writing—original draft preparation, C.A.-R., I.G.-M. and R.F.-B.; writing—review and editing, C.A.-R., I.G.-M. and R.F.-B.; supervision, C.A.-R.; project administration, C.A.-R.; funding acquisition, C.A.-R. and R.F.-B. All authors have read and agreed to the published version of the manuscript.

Funding: This study was funded by the Conselleria de Educación, Cultura y Deporte of the Generalitat Valenciana (Spain) (GV/2015/092) and the Universitat Jaume I of Castellón (UJI-A2016-12).

Institutional Review Board Statement: The study was conducted in accordance with the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board (or Ethics Committee) of the Universitat Jaume I of Castellón (date: 25 November 2013).

Informed Consent Statement: Informed consent was obtained from the parents of all the participants involved in the study. Written informed consent to publish this paper was obtained from the children's parents.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments: The researchers would like to thank the children and families who took part in this study.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bishop, D.V.M.; Snowling, M.J.; Thompson, P.A.; Greenhalgh, T. Phase 2 of CATALISE: A multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *J. Child Psychol. Psychiatry* **2017**, *58*, 1068–1080. [CrossRef]
2. Davies, C.; Andrés-Roqueta, C.; Norbury, C.F. Referring expressions and structural language abilities in children with specific language impairment: A pragmatic tolerance account. *J. Exp. Child Psychol.* **2016**, *144*, 98–113. [CrossRef] [PubMed]
3. Norbury, C.F.; Nash, M.; Baird, G.; Bishop, D.V. Using a parental checklist to identify diagnostic groups in children with communication impairment: A validation of the Children's Communication Checklist–2. *Int. J. Lang. Commun. Disord.* **2004**, *39*, 345–364. [CrossRef] [PubMed]
4. Dockrell, J.E.; Lindsay, G.; Palikara, O. Explaining the academic achievement at school leaving for pupils with a history of language impairment: Previous academic achievement and literacy skills. *Child Lang. Teach. Ther.* **2011**, *27*. [CrossRef]
5. Durkin, K.; Toseeb, U.; Botting, N.; Pickles, A.; Conti-Ramsden, G. Social confidence in early adulthood among young people with and without a history of language impairment. *J. Speech Lang. Hear. Res.* **2017**, *60*, 1635–1647. [CrossRef]
6. Snowling, M.J.; Duff, F.J.; Nash, H.M.; Hulme, C. Language profiles and literacy outcomes of children with resolving, emerging, or persisting language impairments. *J. Child Psychol. Psychiatry* **2016**, *57*, 1360–1369. [CrossRef]
7. Bishop, D.V.M.; Norbury, C.F. Exploring the borderlands of autistic disorder and specific language impairment: A study using standardised diagnostic instruments. *J. Child Psychol. Psychiatry* **2002**, *43*, 917–929. [CrossRef]
8. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*, 5th ed.; American Psychiatric Association: Washington, DC, USA, 2013. [CrossRef]
9. Conti-Ramsden, G.; Botting, N. Emotional health in adolescents with and without a history of specific language impairment (SLI). *J. Child Psychol. Psychiatry* **2008**, *49*, 516–525. [CrossRef]
10. Kilpatrick, T.; Leitão, S.; Boyes, M. Mental health in adolescents with a history of developmental language disorder: The moderating effect of bullying victimisation. *Autism Dev. Lang. Impair.* **2019**, *4*. [CrossRef]
11. Lindsay, G.; Dockrell, J.E. Longitudinal patterns of behavioral, emotional, and social difficulties and self-concepts in adolescents with a history of specific language impairment. *Lang. Speech Hear. Serv. Sch.* **2012**, *43*, 445–460. [CrossRef]
12. Snowling, M.J.; Bishop, D.V.; Stothard, S.E.; Chipchase, B.; Kaplan, C. Psychosocial outcomes at 15 years of children with a preschool history of speech-language impairment. *J. Child Psychol. Psychiatry* **2006**, *47*, 759–765. [CrossRef] [PubMed]
13. van den Bedem, N.P.; Dockrell, J.E.; van Alphen, P.M.; Kalicharan, S.V.; Rieffe, C. Victimization, Bullying, and Emotional Competence: Longitudinal Associations in (Pre)Adolescents With and Without Developmental Language Disorder. *J. Speech Lang. Hear. Res.* **2018**, *61*, 2028–2044. [CrossRef]
14. Amador, J.A.; Forns, M.; Guàrdia, J.; Peró, M. Estructura factorial y datos descriptivos del perfil de atención y del cuestionario TDAH para niños en edad escolar [Factor structure and descriptive data from the Attention Profile and the ADHD Questionnaire for school-age children]. *Psicothema* **2006**, *18*, 696–703.
15. Zhang, Y.; Xu, X.; Jiang, Y.; Sun, W.; Wang, Y.; Song, Y.; Dong, S.; Zhu, Q.; Jiang, F.; Sheng, L. Early language and communication development in Chinese children: Adaptation and validation of a parent report instrument. *Int. J. Speech Lang. Pathol.* **2020**, 1–10. [CrossRef] [PubMed]
16. Thudge, J.; Hogan, D. An ecological approach to observations of children's everyday lives. In *Researching Children's Experience: Approaches and Methods*; Greene, S., Hogan, D., Eds.; SAGE Publications: London, UK, 2004.
17. Rosetti, L. *The Rossetti Infant-Toddler Language Scale*; LinguiSystems: East Moline, IL, USA, 2006.
18. Rutter, M.; Bailey, A.; Lord, C. *SCQ: The Social Communication Questionnaire*; Western Psychological Services: Torrance, CA, USA, 2003.
19. Bishop, D.V.M. *The Children's Communication Checklist–2*; Psychological Corp.: London, UK, 2003.
20. Bishop, D.V.; Norbury, C.F. Executive functions in children with communication impairments, in relation to autistic symptomatology. 1: Generativity. *Autism* **2005**, *9*, 7–27. [CrossRef]
21. Bishop, D.V.M. Development of the Children's Communication Checklist (CCC): A method for assessing qualitative aspects of communicative impairment in children. *J. Child Psychol. Psychiatry* **1998**, *39*, 879–891. [CrossRef] [PubMed]
22. Bishop, D.V.M. What Causes Specific Language Impairment in Children? *Curr. Dir. Psychol. Sci.* **2006**, *15*, 217–221. [CrossRef]
23. Geurts, H.M. *CCC-2-NL: Children's Communication Checklist-2*; Harcourt Assessment B.V.: Amsterdam, The Netherlands, 2007.
24. Helland, W.A.; Biringer, E.; Helland, T.; Heimann, M. The usability of a Norwegian adaptation of the Children's Communication Checklist Second Edition (CCC-2) in differentiating between language impaired and non-language impaired 6- to 12-year-olds. *Scand. J. Psychol.* **2009**, *50*, 287–292. [CrossRef]
25. Glumbić, N.; Brojčin, B.; Đorđević, M. Pouzdanost Komunikacione čekliste za decu [Reliability of the Communication Checkpoint for Children]. *Beogradska Defektološka Škola* **2010**, *16*, 209–216. [CrossRef]
26. Glumbić, N.; Brojčin, B. Factor structure of the Serbian version of the Children's Communication Checklist-2. *Res. Dev. Disabil.* **2012**, *33*, 1352–1359. [CrossRef] [PubMed]
27. Vézina, M.; Samson-Morasse, C.; Gauthier-Desgagné, J.; Fossard, M.; Sylvestre, A. Développement de la version québécoise francophone du Children's Communication Checklist—2 (CCC-2): Traduction, adaptation et équivalence conceptuelle. *Revue Canadienne d'orthophonie et d'audiologie* **2011**, *35*, 244–253.

28. Costa, V.B.; Harsányi, E.; Martins-Reis, V.O.; Kummer, A. Translation and cross-cultural adaptation into Brazilian Portuguese of the Children's Communication Checklist-2. *Codas* **2013**, *25*, 115–119. [CrossRef] [PubMed]
29. Crespo-Eguílaz, N.; Magallón, S.; Sánchez-Carpintero, R.; Narbona, J. La adaptación al castellano de la Children's Communication Checklist permite detectar las dificultades en el uso del lenguaje pragmático y diferenciar subtipos clínicos. [The Spanish adapted version of the Children's Communication Checklist identifies disorders of pragmatic use of language and differentiates between clinical subtypes]. *Rev. Neurol.* **2016**, *62*, S49–S57. [CrossRef]
30. Mendoza, E.; Garzón, M. Puede el CCC-2 diferenciar perfiles pragmáticos? [Can CCC-2 differentiate pragmatic profiles?]. *Rev. Chil. Fonoaudiol.* **2012**, *11*, 37–56. [CrossRef]
31. Andrés-Roqueta, C.; Adrian, J.E.; Clemente, R.A.; Villanueva, L. Social cognition makes an independent contribution to peer relations in children with Specific Language Impairment. *Res. Dev. Disabil.* **2016**, *49*, 277–290. [CrossRef] [PubMed]
32. Bignell, S.; Cain, K. Pragmatic aspects of communication and language comprehension in groups of children differentiated by teacher ratings of inattention and hyperactivity. *Br. J. Dev. Psychol.* **2007**, *25*, 499–512. [CrossRef]
33. Bishop, D.V.M.; McDonald, D. Identifying language impairment in children: Combining language test scores with parental report. *Int. J. Lang. Commun. Disord.* **2009**, *44*, 600–615. [CrossRef] [PubMed]
34. Geurts, H.; Embrechts, M. Pragmatics in pre-schoolers with language impairments. *Int. J. Lang. Commun. Disord.* **2010**, *45*, 436–447. [CrossRef] [PubMed]
35. Grzadzinski, R.; Di Martino, A.; Brady, E.; Mairena, M.A.; O'Neale, M.; Petkova, E.; Castellanos, F.X. Examining autistic traits in children with ADHD: Does the autism spectrum extend to ADHD? *J. Autism Dev. Disord.* **2011**, *41*, 1178–1191. [CrossRef]
36. Volden, J.; Phillips, L. Measuring pragmatic language in speakers with autism spectrum disorders: Comparing the children's communication checklist—2 and the test of pragmatic language. *Am. J. Speech Lang. Pathol.* **2010**, *19*, 204–212. [CrossRef]
37. Volden, J.; Coolican, J.; Garon, N.; White, J.; Bryson, S. Brief report: Pragmatic language in autism spectrum disorder: Relationships to measures of ability and disability. *J. Autism Dev. Disord.* **2009**, *39*, 388–393. [CrossRef]
38. Whitehouse, A.; Barry, J.G.; Bishop, D.V. Further defining the language impairment of autism: Is there a specific language impairment subtype? *J. Commun. Disord.* **2008**, *41*, 319–336. [CrossRef] [PubMed]
39. Bishop, D.V.M.; Hardiman, M.J.; Barry, J.G. Is auditory discrimination mature by middle childhood? A study using time-frequency analysis of mismatch responses from 7 years to adulthood. *Dev. Sci.* **2011**, *14*, 402–416. [CrossRef] [PubMed]
40. Dawes, P.; Bishop, D. Auditory processing disorder in relation to developmental disorders of language, communication and attention: A review and critique. *Int. J. Lang. Commun. Disord.* **2009**, *44*, 440–465. [CrossRef] [PubMed]
41. Ferguson, M.A.; Hall, R.L.; Riley, A. Communication, listening, cognitive and speech perception skills in children with auditory processing disorder (APD) or Specific Language Impairment (SLI). *J. Speech Lang. Hear. Res.* **2011**, *54*, 211–227. [CrossRef]
42. Solomon, M.; Olsen, E.; Niendam, T.; Ragland, J.D.; Yoon, J.; Minzenberg, M.; Carter, C.S. From lumping to splitting and back again: Atypical social and language development in individuals with clinical-high-risk for psychosis, first episode schizophrenia, and autism spectrum disorders. *Schizophr. Res.* **2011**, *131*, 146–151. [CrossRef]
43. Quach, J.; Hiscock, H.; Canterford, L.; Wake, M. Australian population longitudinal study: Outcomes of child sleep problems over the school-transition period. *Pediatrics* **2009**, *123*, 287–292. [CrossRef] [PubMed]
44. Hoffmann, A.; Martens, M.A.; Fox, R.; Rabidoux, P.; Andridge, R. Pragmatic language assessment in Williams syndrome: A comparison of the Test of Pragmatic Language-2 and the Children's Communication Checklist-2. *Am. J. Speech Lang. Pathol.* **2013**, *22*, 198–204. [CrossRef]
45. Ramirez-Inscoe, J.; Moore, D.R. Processes that influence communicative impairments in deaf children using cochlear implants. *Ear Hear.* **2011**, *32*, 690–698. [CrossRef]
46. Andrés-Roqueta, C.; Katsos, N. The contribution of grammar, vocabulary and theory of mind in pragmatic language competence in children with autistic spectrum disorders. *Front. Psychol.* **2017**, *8*, 996. [CrossRef] [PubMed]
47. Adams, C. Intervention for children with pragmatic language impairments: Frameworks, evidence and diversity. In *Understanding Developmental Language Disorders: From Theory to Practice*; Norbury, C.F., Tomblin, J.B., Bishop, D.V.M., Eds.; Psychology Press: London, UK, 2008; pp. 189–204.
48. Andrés-Roqueta, C.; Clemente, R.A. Idiom understanding competence of Spanish children with Specific Language Impairment. In *Atypical Language Development in Romance Languages*; Aguilar-Mediavilla, E., Buil-Legaz, L., López-Penadés, R., Sanchez-Azanza, V.A., Adrover-Roig, D., Eds.; John Benjamins Publishing Company: Amsterdam, The Netherlands, 2019; pp. 114–126. [CrossRef]
49. Norbury, C.F. Factors supporting idiom comprehension in children with communication disorders. *J. Speech Lang. Hear. Res.* **2004**, *47*, 1179–1193. [CrossRef]
50. Norbury, C.F. The relationship between theory of mind and metaphor: Evidence from children with language impairment and autistic spectrum disorders. *Br. J. Dev. Psychol.* **2005**, *23*, 383–399. [CrossRef]
51. Andrés-Roqueta, C.; Katsos, N. A distinction between linguistic and social pragmatics helps the precise characterization of pragmatic challenges in children with Autism Spectrum Disorders and Developmental Language Disorder. *J. Speech Lang. Hear. Res.* **2020**, *63*, 1494–1508. [CrossRef]
52. Katsos, N.; Andrés-Roqueta, C.; Clemente, R.A.; Cummins, C. Are children with Specific Language Impairment challenged by linguistic-pragmatics? *Cognition* **2011**, *119*, 43–57. [CrossRef] [PubMed]
53. Surian, L.; Baron-Cohen, S.; Van der Lely, H. Are children with autism deaf to Gricean Maxims? *Cogn. Neuropsychiatr.* **1996**, *1*, 55–71. [CrossRef]

54. Brock, J.; Norbury, C.; Einav, S.; Nation, K. Do individuals with autism process words in context? Evidence from language-mediated eye-movements. *Cognition* **2008**, *108*, 896–904. [CrossRef]
55. Norbury, C.F.; Gemmell, T.; Paul, R. Pragmatics abilities in narrative production: A cross-disorder comparison. *J. Child. Lang.* **2014**, *41*, 485–510. [CrossRef] [PubMed]
56. Andrés-Roqueta, C.; Ballester-Manuel, L. Comprensión del Humor Gráfico en niños y niñas con Trastorno Específico del Lenguaje y con Trastorno del Espectro Autista [Graphic Humour understanding in children with Specific Language Impairment and children with Autistic Spectrum Disorder]. In *Psicología y educación: Presente y futuro*; Castejón, J., Ed.; ACIPE—Asociación Científica de Psicología y Educación: Alicante, Spain, 2016; pp. 1803–1811.
57. Andrés-Roqueta, C.; Clemente, R.A. Dificultades pragmáticas en el Trastorno Específico del Lenguaje. El papel de las tareas mentalistas [Pragmatic difficulties in children with specific language impairment. The role of mentalistic tasks]. *Psicothema* **2010**, *22*, 677–683.
58. Happé, F. An advanced test of theory of mind: Understanding of story characters thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *Autism Dev. Disord.* **1994**, *2*, 129–154. [CrossRef]
59. Botting, N.; Conti-Ramsden, G. The role of language, social cognition, and social skill in the functional social outcomes of young adolescents with and without a history of SLI. *Br. J. Dev. Psychol.* **2008**, *26*, 281–300. [CrossRef]
60. Martin, I.; McDonald, S. Weak coherence, no theory of mind, or executive dysfunction? Solving the puzzle of pragmatic language disorders. *Brain Lang.* **2003**, *85*, 451–466. [CrossRef]
61. Green, B.C.; Johnson, K.A.; Bretherton, L. Pragmatic language difficulties in children with hyperactivity and attention problems: An integrated review. *Int. J. Lang. Commun. Disord.* **2014**, *49*, 15–29. [CrossRef] [PubMed]
62. Mendoza, E.; Carballo, G.; Muñoz, J.; Fresneda, M.D. *CEG: Test de Comprensión de Estructuras Gramaticales [Grammatical Structures Comprehension Test]*; TEA Ediciones: Madrid, Spain, 2005.
63. Saborit, C.; Julián, J.P. *L'avaluacio del Llenguatge Infantil, ELI [Assessment of Children's Language]*; Universitat Jaume I. Col·leccio Educació: Castellón, España, 2005.
64. Conti-Ramsden, G.; Botting, N.; Faragher, B. Psycholinguistic markers for specific language impairment (SLI). *J. Child Psychol. Psychiatry* **2001**, *42*, 741–748. [CrossRef]
65. Raven, J.C. *Manual for the Coloured Progressive Matrices (Revised)*; NFERNelson: Windsor, UK, 1984.
66. Norbury, C.; Bishop, D. Children's Communication Checklist—2: A validation study. *Rev. Tranel (Travaux Neuchâtelois de Linguistique)* **2005**, *42*, 83-63.
67. Bishop, D.V.M.; Laws, G.; Adams, C.; Norbury, C.F. High heritability of speech and language impairments in 6-year-old twins demonstrated using parent and teacher report. *Behav. Genet.* **2006**, *36*, 173–184. [CrossRef]
68. Bishop, D.V.M.; Maybery, M.; Wong, D.; Maley, A.; Hallmayer, J. Characteristics of the broader phenotype in autism: A study of siblings using the Children's Communication Checklist-2. *Am. J. Med. Genet.* **2006**, *141*, 117–122. [CrossRef]
69. Raven, J.C.; Court, J.H.; Raven, J. *Section 2: Coloured Progressive Matrices (1990 Edition, with US Norms). Manual for the Raven's Progressive Matrices and Vocabulary Scales*; Oxford Psychologist Press: Oxford, UK, 1990.
70. Bishop, D.V.M. *Test for Reception of Grammar. Medical Research Council*; Chapel Press; University of Manchester: Manchester, UK, 1983.
71. Kaufman, A.S.; Kaufman, N.L. *K-BIT. Test Breve de Inteligencia de Kaufman (K-BIT)*, 2nd ed.; TEA Ediciones: Madrid, Spain, 2000.
72. Wechsler, D. *Manual for the Wechsler Intelligence Scale for Children*, 3rd ed.; The Psychological Corporation: San Antonio, TX, USA, 1991.
73. Dunn, L.M.; Dunn, L.M.; Arribas, D. *PPVT-III Peabody, Test de Vocabulario en Imágenes*; TEA Ediciones: Madrid, Spain, 2006.
74. Kirk, S.; McCarthy, J.; Kirk, W. *Illinois Test of Psycholinguistic Abilities. Examiner's Manual*; University of Illinois Press: Champaign, IL, USA, 1968; pp. 3–14.
75. Muñoz, J.; Fresneda, M.D.; Mendoza, E.; Carballo, G. Propiedades psicométricas de una prueba de comprensión gramatical [Psychometric properties of a grammatical comprehension test]. *Rev. Neurol.* **2008**, *47*, 21–23.
76. White, S.; Hill, E.; Happé, F.; Frith, U. Revisiting the Strange Stories: Revealing mentalising impairments in autism. *Child Dev.* **2009**, *80*, 1097–1117. [CrossRef] [PubMed]
77. Cairns, E.; Cammock, T. Development of a more reliable version of the Matching Familiar Figures Test. *Dev. Psychol.* **1978**, *14*, 555–560. [CrossRef]
78. Fritz, C.O.; Morris, P.E.; Richler, J.J. Effect size estimates: Current use, calculations, and interpretation. *J. Exp. Psychol. Gen.* **2012**, *141*, 2–18. [CrossRef] [PubMed]

children

Article

Learning Disabilities in Reading and Writing and Type of Delivery in Twin Births

María-José González-Valenzuela [1,*], Dolores López-Montiel [2], Olga Cazorla-Granados [3] and Ernesto-Santiago González-Mesa [4]

1 Department of Development and Educational Psychology, Faculty of Psychology, University of Málaga, Teatinos University Campus, s/n, 29071 Málaga, Spain
2 Department of Psychobiology and Methodology of Behavioral Sciences, Faculty of Psychology, University of Málaga, Teatinos University Campus, s/n, 29071 Málaga, Spain; dlopezm@uma.es
3 Santa Mª de la Victoria Teaching Foundation, Santa Rosa de Lima School, Argentea 19, 29006 Málaga, Spain; olgacazorla@gmail.com
4 Department of Surgical Specialties, Biochemistry and Immunology, Faculty of Medicine, University of Málaga, Teatinos University Campus, s/n, 29071 Málaga, Spain; egonzalezmesa@gmail.com
* Correspondence: valenzu@uma.es

Citation: González-Valenzuela, M.-J.; López-Montiel, D.; Cazorla-Granados, O.; González-Mesa, E.-S. Learning Disabilities in Reading and Writing and Type of Delivery in Twin Births. *Children* **2021**, *8*, 834. https://doi.org/10.3390/children8100834

Academic Editors: Eva Aguilar Mediavilla and Marco Carotenuto

Received: 20 July 2021
Accepted: 20 September 2021
Published: 23 September 2021

Abstract: The aim of this study was to analyse the relationship between the type of delivery (vaginal or caesarean), as a risk factor, and the likelihood of having learning disabilities in reading (reading accuracy) and writing (phonetic and visual orthography), controlling for the interaction and/or confounding effect of gestational, obstetric, and neonatal variables (maternal age at delivery, gestational age, foetal presentation, Apgar 1, and newborn weight) among six-year-old children born in twin births. In this retrospective cohort study, the exposed and non-exposed cohorts consisted of children born by caesarean section and vaginal delivery, respectively. A total of 124 children born in twin births were evaluated in year one of primary education. Intelligence was measured using the K-BIT test; reading and writing variables were evaluated using the Evalúa-1 battery of tests, and clinical records were used to measure gestational, obstetric, and neonatal variables. Binary logistic regressions applied to each dependent variable indicated that caesarean delivery is a possible independent risk factor for difficulties in reading accuracy and phonetic and visual orthography. Future research using larger samples of younger children is required to analyse the relationship between obstetric and neonatal variables and the different basic indicators of reading and writing.

Keywords: learning disabilities; reading; writing; type of delivery; twin births

1. Introduction

Twins have been considered in some studies as a psychologically and academically vulnerable population, even though this population is subject to the influence of certain prenatal and perinatal factors—hence, the interest of this study [1]. Multiple births are associated with prematurity, low Apgar scores, neonatal sepsis, pulmonary hypertension, hyperbilirubinaemia, and restricted intrauterine growth, among others. Some of these complications can on occasion lead to neuropsychological difficulties, academic difficulties, and even death [2–6].

Specific learning disabilities in reading and writing present disorders in cognitive processes and associative behaviours, which are explained by hypotheses on the biological origin thereof [7–12]. Pupils who display such difficulties have below the expected reading and writing level for their age, despite receiving adequate instruction for at least six months and having the intelligence to be a good reader [8,12]. They are characterised by difficulties in the precise and fluid recognition of words and by problems with spelling—in other words, difficulties with the basic components of reading and writing, such as reading and writing accuracy [7]. These specific difficulties affect lexical and sub-lexical processes (visual and phonological processing) implicit in both reading and writing accuracy, and they result from a deficit in the phonological component that is neurobiological in origin [7]. Estimates of the prevalence of learning disabilities vary by definition and language, but generally affect between 7% and 16% of school-age children [13,14]. There are studies that highlight the influence of pre- and perinatal factors in the appearance of cerebral dysfunctions, which would justify the existence of these difficulties in academic learning among children born in single and multiple births. Specifically, these factors could justify the characteristic cognitive and linguistic problems displayed [15–18], among others. Differences in the volume of grey matter in the cortex, reduced cortical activity, and disorders in cortical plasticity manifested by children with learning disabilities might be conditioned by the influence of these peri- and prenatal factors [19–21].

Prematurity would be one such perinatal risk factor [22–31]. Some of these studies find that children with low birth weight present reading difficulties, which vary according to weight gain and loss [24,25,30,31]. Other research states that children who weigh less than 1500 g at birth and are born before 34 weeks of gestation then go on to obtain lower scores than children born full term with regard to spelling, accuracy, and reading speed at young ages. These results are associated with the difficulties they display in terms of speech, phonological awareness, visual perception, rapid naming, and executive functions [23,26,27,32]. Furthermore, the *DSM-5* [7] indicates that one of the possible risk factors for specific learning disabilities is low birth weight.

Maternal age and foetal presentation are also considered risk factors in multiple births for psychological development and academic learning. Advanced maternal age increases the likelihood of multiple gestation [33], and extreme maternal age (teenagers and women over 35) has been described as an independent risk factor associated with adverse perinatal results [34]. Furthermore, the risk of complications such as preterm delivery is also higher in multiple pregnancies [35]. Foetal presentation conditions the type of delivery, with a high rate of caesarean sections when either of the foetuses is presenting non-cephalically. Vaginal delivery of the second twin is associated with lower scores on the Apgar test and lower umbilical cord pH in vaginal deliveries, depending on the interval between the births of both twins [36].

Another of the prenatal and obstetric factors that also seem to be related to academic learning and difficulties in twin births is type of delivery. From an obstetric point of view, there are no recommendations based on the analysis of prospective data in relation to the best delivery option for the second twin. In fact, current recommendations are based on retrospective studies, and the monitoring of children is based on the study of grave morbidity or mortality in the perinatal period [37]. One of the few prospective studies published about the influence of delivery type concludes that there are no differences in maternal or foetal morbidity in twins born vaginally or by caesarean section if they are both programmed and attended by qualified professionals, although the foetus born second does present a higher risk of morbidity than the first-born twin [38]. In other words, some studies indicate that there is no evidence that proves programmed caesarean delivery for twin births to be better than vaginal ones [3] and that it includes a risk of low Apgar scores, neonatal respiratory morbidity, perinatal mortality caused by the rupture of the uterus or by placenta previa, and placenta abruption in subsequent pregnancies [6,39,40]. However, other studies note that caesarean delivery reduces the risk of low scores in the Apgar 5 test in the first twin in breech presentation, foetal breech presentation, and intrapartum foetal death, but it increases the risk of both maternal and neonatal death in the event of cephalic presentation [3,6,41] and is associated with severe motor delay in early ages when performed under general anaesthesia [40]. Along these lines, previous studies have found that caesarean delivery is a risk factor for psychological development problems [41,42] and difficulties in reading (lexical access and comprehension), writing (phonetic and visual lexical access), and arithmetic in twin births [41,43]. However, these studies did not examine in depth the separate components of reading and writing, such as reading accuracy and phonetic and visual orthography. It would be interesting to analyse which of these components related to lexical and sub-lexical processes (visual and phonological processing) are influenced by the type of delivery since they are different cognitive processes.

Therefore, the main aim of this study of children born in twin births once they reached the age of six was to analyse the relationship between type of delivery (vaginal or caesarean), as a risk factor, and the probability of presenting learning disabilities in reading accuracy, phonetic orthography, and visual orthography, controlling for the interaction and/or confounding effect of other gestational, obstetric, and neonatal variables such as maternal age at delivery, gestational age, foetal presentation, Apgar 1 score, and newborn weight on account of their importance in neuropsychological development during childhood and specific academic learning.

2. Material and Methods

2.1. Design

To achieve the objective of this paper, the authors designed a retrospective epidemiological cohort study in which type of delivery (risk) predicted the learning of reading accuracy and visual and phonetic orthography among six-year-old children born in twin births (outcome). The exposed cohort comprised children born by caesarean section, and children from the sample selected who were born by natural or induced vaginal delivery comprised the non-exposed cohort

2.2. Participants

As in previous studies [41–43], the study population consisted of Caucasian children born in twin births at the Hospital Materno-Infantil deMálaga, once they had reached the age of six, who were born in 2005, and were in Year 1 of Primary Education. Of the 7120 births registered in the year 2005 in this hospital, 135 were twin births. Of the 270 children born of twin births that same year, 64 were excluded since they were born prior to 32 weeks of gestation, as were 14 who were still in Early Years Education and 68 who could not be recruited for the study owing to difficulties locating their families or because they did not wish to take part. Therefore, of the population selected, a total of 124 children could be evaluated, born in 62 twin births [41–43] (see Figure 1).

The age of the children was between 74 and 86 months (M = 79.42 and SD = 3.44), and the sample comprised 62 male (50%) and 62 female participants (50%) (Table 1).

The 124 children included in the final sample presented an average intelligence quotient (standard score = 101 and Enneatype type = 5). Of the 124 children included in the final sample, those with normal levels of intelligence (standard score $\geq$ 101 and E $\geq$ 5) who presented scores for reading and writing accuracy (phonetic and visual orthography) below the 25th percentile were classified as children with specific learning disabilities. The instruments used to evaluate intellectual quotient and reading and writing accuracy are defined in the instruments section. This criterion has been considered by different research studies to diagnose these subjects and coincides with the criterion established by the instrument used for the assessment of such disabilities [42–47]. In other words, performance in reading accuracy and phonetic and visual orthography was defined in terms of the following dichotomy: presence of difficulty in the learning of reading accuracy, phonetic orthography, or visual orthography, if the score achieved by the child in specific learning tests is below the 25th percentile; absence of difficulty in the learning of reading accuracy, phonetic orthography, or visual orthography, if the child achieves a grade of at least the 25th percentile in each measurement. A total of 93 children (75%) out of the total sample selected presented no difficulties in reading accuracy, 92 (74.2%) in phonetic orthography, and 93 (75%) in visual orthography. A total of 31 children (25%) presented difficulties in reading accuracy, 32 (25.8%) in phonetic orthography, and 31 (25%) in visual orthography (see Tables 2–4).

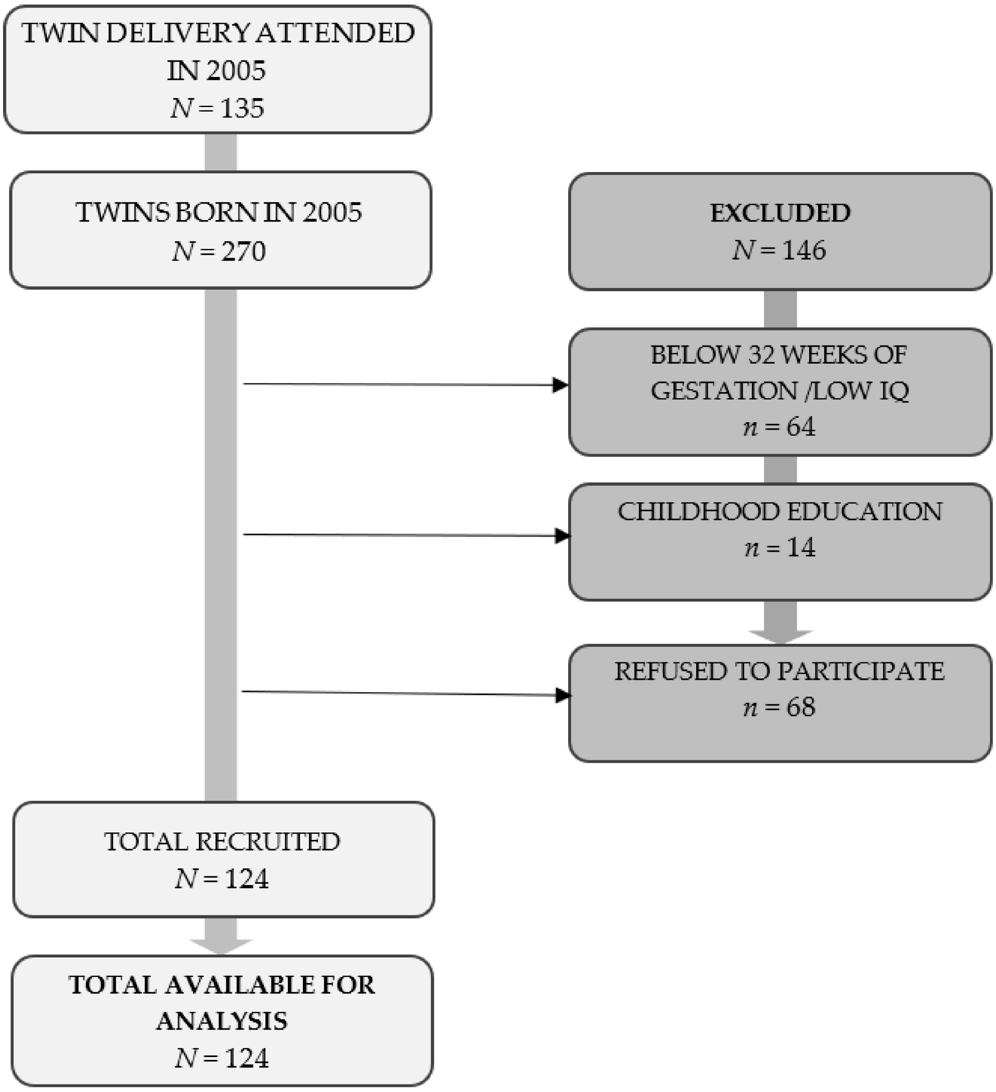


Figure 1. Flow diagram of participants through each stage of the study.

2.3. Instruments

To evaluate specific learning difficulties, different instruments were used.

To verify that the children did not have any intellectual disabilities, we measured their intelligence by means of the Kaufman Brief Intelligence Test (K-BIT) [48]. The K-BIT test analyses verbal and non-verbal intelligence from the age of four onwards and comprises two subtests: Vocabulary and Matrices. The first evaluated verbal abilities related to learning at school, using knowledge of words, and forming concepts (expressive vocabulary and definitions). The child had to say the name of a figure that was presented to him/her and find the word that best fit two clues that were given (a descriptive phrase and a word in which some letters had been deleted). The second test measured non-verbal abilities and the capacity to resolve problems, complete analogies, and perceive relations. It included two types of exercises consisting of 48 items: the first consisted of drawings of objects (e.g., a moon), and the child had to select from among 5 drawings that were placed (trousers, a sun, an apple, a car, and a heater) the one that best related to the stimulus drawing; the second demanded that the child complete a visual analogy from either figurative drawings or abstract figures. The total score for Intelligence was the sum total of the scores attained in each of the subtests (number of correct answers given). The reliability coefficients for the different tests ranged between 0.80 and 0.90 [48].

Table 1. Characteristics of the samples.

Variables	*N* = 124 (Study Sample)			*N* = 62 (Validation Sample)		
	n (%)	*M (SD)*	Range	*n* (%)	*M (SD)*	Range
Age (months)		79.42 (3.44)	74–86		79.42 (3.46)	74–86
Gender						
Female	62 (50)			34 (50)		
Male	62 (50)			28 (45.2)		
Mother's level of education						
Low	51 (41.1)			25 (40.3)		
Middle	38 (30.6)			19 (30.6)		
High	35 (28.2)			18 (29)		
Father's level of education						
Low	58 (46.8)			29 (46.8)		
Middle	40 (32.3)			20 (32.3)		
High	26 (21)			13 (21)		
Type of delivery						
Vaginal	84 (67.7)			42 (67.7)		
Caesarean	40 (32.3)			20 (32.3)		
Foetal presentation						
Cephalic	80 (64.5%)			41 (66.1%)		
Non-cephalic	44 (35.5%)			21 (33.9%)		
Maternal Age (years)		33.24 (4.29)	22–45		33.24 (4.29)	22–45
Gestational Age (weeks)		35.14 (2.09)	32–41		35.14 (2.09)	32–41
Weight of Newborn (grams)		2137.76 (432.79)	1179–3080		2170.90 (433.66)	1250–3080
Apgar 1		8.41 (1.18)	4–10		8.36 (1.25)	5–9
Reading Accuracy		110.18 (27.21)	25–144		107.82 (29.25)	25–144
Phonetic Orthography		61.60 (12.67)	11–78		60.29 (15.31)	11–78
Visual Orthography		16.12 (6.26)	0–28		16.23 (6.36)	0–26

Note. *M* = mean; *SD* = standard deviation; Low: primary and pre-secondary studies; Middle: secondary or high school (technical and non-technical); High: university and graduate.

To analyse performance in the reading and writing measures applied here, we used different subtests from the Evalúa-1 psycho-pedagogical battery of tests [45]. The Evalúa-1 psycho-pedagogical battery is an instrument for children aged 6–7, which provides information about the cognitive foundations of learning, basic instrumental learning, and affective and behavioural aspects, in order to facilitate decision-making with regard to education processes. The three tests used encompassed reading accuracy, phonetic orthography, and visual orthography.

The Reading Accuracy test measured lexical access processes. These processes refer to knowledge of the main grapheme–phoneme conversion rules and fluency and pace in reading. Lexical access processes were measured by means of tasks to identify letters, syllables, words, pseudowords, and phrases. The child had to perform different tasks: to mark in each box the letter that was dictated to him (12 items), to match each letter from a column with that same letter from another column (14 items), to join with arrows each word from a column with the same word that was to its side (10 items), and to read aloud syllables, words and pseudowords (50 items), and phrases (2 items). The total score was obtained by adding together the number of correct answers given in the different tasks. The reliability of the test according to Cronbach's alpha coefficient was 0.93 [45].

Table 2. Characteristics of the samples for both vaginal delivery and caesarean delivery.

Variables	**Type of Delivery**													
	Vaginal Delivery					**Caesarean**								
***N*= 124 Study Sample**	***n*= 84 (66.7%)**					***n* = 40 (32.3%)**					**Statistical Test**			
	***n*(%)**	***M***	***SD***	**Range**	**MR**	***n*(%)**	***M***	***SD***	**Range**	**MR**	***U***	***Z***	***p***	**ES**
Reading Accuracy		114.73	25.35	26–144	69.17		100.63	28.81	25–144	48.49	1119.50	−2.99	0.003	0.26 [1]
Phonetic Orthography		64.06	11.68	11–78	70.30		56.43	13.26	25–78	46.13	1025.00	−3.50	0.000	0.31 [1]
Visual Orthography		17.06	5.46	0–26	67.48		14.15	7.36	0–28	52.05	1262.00	−2.24	0.025	0.20 [1]
Maternal Age (years)		32.71	4.03	22–40	59.21		34.35	4.42	28–45	69.40	1404.00	−1.48	0.139	0.13 [1]
Gestational Age (weeks)		35.09	1.99	32–40	62.31		35.25	2.27	32–41	62.90	1664.00	−0.08	0.932	0.01 [1]
Apgar 1		8.36	1.23	4–10	61.11		8.53	1.06	5–9	63.85	1586.00	−0.51	0.612	0.04 [1]
											t	*df*		
Weight of Newborn (grams)		2154.48	445.26	1310–3080	–		2102.65	408.59	1179–2905	–	0.62	122	0.535	0.12 [2]
Foetal presentation											χ^2			
Cephalic	71 (84.5)					9 (22.5)					45.53	1	0.000	0.60 [3]
Non-cephalic	13 (15.5)					31 (77.5)								
***N* = 62 Validation Sample**	***n*= 42 (66.7%)**					***n*= 20 (32.3%)**								
	***n*(%)**	***M***	***SD***	**Range**	**MR**	***n*(%)**	***M***	***SD***	**Range**	**MR**	***U***	***Z***	***p***	**ES**
Reading Accuracy		112.69	27.88	26–143	35.11		97.60	30.13	25–144	23.93	268.50	−2.28	0.022	0.29 [1]
Phonetic Orthography		64.05	14.56	11–78	36.96		52.40	14.07	25–76	20.03	190.50.	−3.46	0.001	0.44 [1]
Visual Orthography		17.48	5.19	3–26	34.74		13.60	7.81	0–25	24.70	284.00	−2.05	0.040	0.26 [1]
Maternal Age (years)		32.81	3.20	22–40	30.31		34.15	4.66	28–45	34.00	370.00	−0.75	0.450	0.09 [1]
Gestational Age (weeks)		35.00	2.03	32–40	30.43		35.44	2.22	32–41	33.75	375.00	−0.68	0.497	0.08 [1]
Apgar 1		8.37	1.22	5–9	31.07		8.53	1.35	5–9	30.85	407.00	−0.06	0.951	0.01 [1]
											t	*df*		
Weight of Newborn (grams)		2140.93	456.88	1310–3080	–		2233.85	383.65	1250–2905	–	−0.78	60	0.435	0.22 [2]
Foetal presentation		112.69	27.88	26–143	35.11		97.60	30.13	25–144	23.93	χ^2			
Cephalic	37 (88.1)					4 (20)					28.05	1	.000	0.67 [3]
Non-cephalic	5 (11.9)					16 (80)								

Note. *M* = mean; *SD* = standard deviation; MR = rean rank; *U* = Mann–Whitney U-test; *t* = Student's *t*-test; Pearson χ^2; ES = effect size. [1] Correlation coefficient *r* (Cohen's reference values: small = 0.10; medium = 0.30; large = 0.50; very large = 0.70). [2] Cohen's *d* (Cohen's reference values: small = 0.20; medium = 0.50; large = 0.80). [3] Cramer's *V* (Cramer's *V* reference values for *df less* = 1: small = 0.10; medium = 0.30; large = 0.50).

Table 3. Contingency tables and bivariate associations between reading accuracy, type of delivery, and the control variables.

Variables	Categories	Total N = 124	Dependent Variable Reading Accuracy		Test [1]	Sig.	ES [2]	OR	95% CI	
			No RAD n = 93 (75%)	RAD n = 31 (25%)					Lower	Upper
Independent **Type of delivery**	Vaginal Caesarean	84 (67.7%) 40 (32.3%)	70 (83.3%) 23 (57.5%)	14 (16.7%) 17 (42.5%)	9.64 [a]	0.002	0.28	3.69	1.58	8.64
Control Maternal age (years)	Under 35 Over 35	88 (71%) 36 (29%)	66 (75%) 27 (75%)	22 (25%) 9 (25%)	0.00 [a]	0.999	0.00	1.00	0.41	2.45
Gestational age of newborn (weeks)	Over 37 Under 37	40 (32.3%) 84 (67.7%)	29 (72.5%) 64 (76.2%)	11 (27.5%) 20 (23.8%)	0.19 [a]	0.657	0.04	0.82	0.35	1.94
Foetal presentation	Cephalic Non-cephalic	80 (64.5%) 44 (35.5%)	62 (77.5%) 31 (70.5%)	18 (22.5%) 13 (29.5%)	0.75 [a]	0.386	0.08	1.44	0.63	3.32
Weight of newborn (grams)	Over 1500 Under 1500	112 (90.3%) 12 (9.7%)	83 (74.1%) 10 (83.3%)	29 (25.9%) 2 (16.7%)	0.49 [b]	0.729	0.06	0.57	0.12	2.77
Apgar 1	Over 7 Under 7	99 (79.8%) 25 (20.2%)	75 (75.8%) 18 (72%)	24 (24.2%) 7 (28%)	0.15 [a]	0.698	0.03	1.21	0.45	3.26

Note. RAD = Reading Accuracy Difficulty; ES = effect size; OR = odds ratio; CI = confidence interval. [1] Pearson's χ^2. [2] Cramer's *V* (Cramer's *V* reference values for *df less* = 1: small = 0.10; medium = 0.30; large = 0.50). [a] 0% of cells with an expected frequency less than 5. [b] 25% of cells with an expected frequency less than 5.

Table 4. Contingency tables and bivariate associations between phonetic orthography, type of delivery, and the control variables.

Variables	Categories	Total N = 124	Dependent Variable Phonetic Orthography		Test [1]	Sig.	ES [2]	OR	95% CI	
			No POD n = 92 (74.2%)	POD n = 32 (25.8%)					Lower	Upper
Independent **Type of delivery**	Vaginal Caesarean	84 (67.7%) 40 (32.3%)	69 (82.1%) 23 (57.5%)	15 (17.9%) 17 (42.5%)	8.59 [a]	0.003	0.26	3.40	1.47	7.87
Control Maternal age (years)	Under 35 Over 35	88 (71%) 36 (29%)	69 (78.4%) 23 (63.9%)	19 (21.16%) 13 (36.1%)	2.81 [a]	0.093	0.15	2.05	0.88	4.79
Gestational age ofnewborn (weeks)	Over 37 Under 37	40 (32.3%) 84 (67.7%)	31 (77.5%) 61 (72.6%)	9 (22.5%) 23 (27.4%)	0.33 [a]	0.561	0.05	1.30	0.53	3.14
Foetal presentation	Cephalic Non-cephalic	80 (64.5%) 44 (35.5%)	66 (82.5%) 26 (59.1%)	14 (17.5%) 18 (40.9%)	8.12 [a]	0.004	0.25	3.26	1.42	7.51
Weight of newborn (grams)	Over 1500 Under 1500	112 (90.3%) 12 (9.7%)	82 (73.2%) 10 (83.3%)	30 (26.8%) 2 (16.7%)	0.58 [b]	0.729	0.07	0.54	0.11	2.64
Apgar 1	Over 7 Under 7	99 (79.8%) 25 (20.2%)	75 (75.8%) 17 (68%)	24 (24.2%) 8 (32%)	0.63 [a]	0.478	0.07	1.47	0.56	3.83

Note. POD = Phonetic Orthography Difficulty; ES = effect size; OR = odds ratio; CI = confidence interval. [1] Pearson's χ^2. [2] Cramer's *V* (Cramer's *V* reference values for *df less* = 1: small = 0.10; medium = 0.30; large = 0.50). [a] 0% of cells with an expected frequency less than 5. [b] 25% of cells with an expected frequency less than 5.

Phonetic orthography and visual orthography tests evaluate the phonetic and visual processes of lexical access which are involved in dictation, copying, and spontaneous writing. The phonetic orthography test measured knowledge of the phoneme–grapheme conversion rules through tasks involving the dictation of different linguistic units such as letters, syllables, words, and phrases; syllables, words, and phrases copying; and completing words in a short text. The sum total of correct answers given when carrying out the different tasks was the total score. The reliability of the test according to the Cronbach alpha coefficient was 0.97 [45]. The visual orthography test measured word recognition by means of a task involving completing graphemes that were missing everyday words with significant reference drawings. The total score was the sum total of correct answers given when carrying out the task. The reliability of the test according to the Cronbach alpha coefficient was 0.87 [45].

By analysing the clinical records of the mothers and their children during gestation and birth, we assessed gestational, obstetric, and neonatal variables. We then dichotomised the control variables analysed, in line with previous studies [41–43]: Maternal age at the time of delivery, over or under the age of 35; gestational age of the newborn, over or under 37 weeks; foetal presentation, cephalic or non-cephalic (breech or transverse); Minute-1 Apgar score, above or below seven points; and newborn weight, above or below 1500 grams.

2.4. Procedure

First, we requested authorisation from the Research Ethics Committee at the Hospital Materno Infantil (Comité Ético de Investigación-CEI), on 30 January 2014, in order to begin compiling data.

Second, we contacted the hospital administration directly to obtain the telephone details of mothers of twin births. Having contacted them, we explained how the research would be developed and then asked whether they would agree for their children to be subjected to psychological evaluation. At the beginning of each evaluation session, an informed consent form was signed by the mothers.

Third, the Kaufman Intelligence Test [48], and subsequently the reading and writing tests, were individually administered by three experienced psychologists. The approximate time taken was 40 min.

Finally, some of the authors involved in the study compiled the obstetric and neonatal data of the selected cases by reviewing the clinical records held at the hospital and through the identification of the mothers selected from all the records of twin births registered in the same year.

2.5. Statistical Analysis

According to the objective, design, and nature of the data, we chose regression models as the main statistical analysis procedure. To verify whether linear models were suitable for the data properties, we conducted descriptive and exploratory analyses of all the variables, performed a bivariate analysis, and estimated multiple linear regression models. For the bivariate analysis, Pearson's correlation coefficients and biserial correlation coefficients were calculated according to the measurement scales of the variables and their corresponding significance tests. They were considered small ($r = |0.10|$), moderate ($r = |0.30|$), or strong ($r = |0.50|$ or greater) according to Cohen's criterion [49] for effect size. The independent variables were selected for each model when in the previous bivariate analyses they had an associated probability lower than 0.05 and an effect size equal to or greater than $|0.20|$. The non-fulfilment of the assumptions of linear regression (linearity, normality, and homoscedasticity) was verified both a priori and a posteriori. Therefore, in accordance with the properties of the data, the main statistical analysis technique chosen was binary logistical regression. To ensure the correct application of this technique, the dependent variable must be dichotomous categorical (measured on a nominal scale), whereas independent variables could be categorical or continuous. For continuous variables, the assumption of linearity between each continuous predictive variable and the logarithm of the response variable must be fulfilled. Having confirmed non-fulfilment, we included these dichotomised variables in the regression models.

Hence, to ensure efficient analysis and a clear interpretation of the results, we dichotomised all the variables that were originally continuous, both the dependent and the control variables, in accordance with the criteria set out previously, in line with previous studies [42,43].

Having explored all the variables, we applied different types of analyses to examine the study variables: preliminary analysis, binary logistic regression analysis, and complementary analysis to validate the regression models eventually estimated.

As part of the preliminary analyses, we carried out an initial descriptive analysis of the sociodemographic characteristics of the sample, the dependent variables, and the potential predictors analysed, in terms of means and standard deviations or in terms of frequencies and percentages, depending on the categorical or continuous nature of the variable according to its original scale. Subsequently, we compared the means or percentages of all the variables according to the type of delivery, with a view to evaluating the main relationships studied, detecting possible masking variables, and selecting the most appropriate ones for the regression models. We applied Student's *t* tests for independent samples and Mann–Whitney U or Pearson's $\chi 2$ tests depending on the nature of the variables, according to their original scale and the fulfilment of the assumptions for the application of parametric tests. We also studied the bivariate relationship between the study variables once they were dichotomised, by means of contingency tables and Pearson's χ^2 independence tests. For these preliminary analyses, we calculated the effect size of the statistics using Cohen's *d*, the *r* correlation coefficient, and Cramer's *V*, respectively [49,50]. Furthermore, to measure the degree to which type of delivery and each control variable increased or decreased the risk of having learning difficulties with regard to reading accuracy, phonetic orthography, and visual orthography, we estimated the unadjusted odds ratios (OR), along with their confidence intervals at 95% (95% CI).

In the exploratory analysis, the presence of outliers in the dependent variables was also analysed by means of the typified residuals and graphic analysis. When finding cases around two and three standard deviations above or below the mean, the logistic regression model was adjusted both with and without these cases. If there were no significant differences in the regression coefficients, in the ORs, and in the adjustment values between both models, the cases could be part of the sample.

In the binary logistic regressions for each dependent variable (difficulties with reading accuracy, phonetic orthography, and visual orthography), when statistically possible we assessed the possible interaction (modification of the principal effect studied) between the control variables and the independent variable type of delivery, as well as the possible confusion between the control variables and the main relationship evaluated (effect of the type of delivery on reading accuracy, phonetic orthography, and visual orthography). Variables were included when the previous bivariate analyses had more than 10% of cases for each cell in the contingency tables and a probability associated with Pearson's χ^2 statistic of less than 0.05.

The researchers [51,52] estimated the regression models based on a maximum hierarchical model, conserving statistically significant interactions and the variables involved, when possible. Having eliminated non-significant interactions sequentially from the model, we then went on to study possible confounding factors, examining the possible bias in the regression coefficients, the accuracy (amplitude) of their confidence intervals, and their standard error, as well as non-statistical criteria, such as change in the OR magnitude. This magnitude evaluated the strength of association between the independent and the dependent variable. The potential for confounding was observed when the magnitude of the OR clinically changed to a substantial degree (10% between the gross and adjusted measures of association) when eliminating one variable from the equation, with regard to the initial model. We also evaluated the effect size of the OR according to their transformation to Cohen's *d* [53]. The variables retained were taken into account in the construction of the most suitable model. In order to evaluate the goodness of fit of the models, we used the Likelihood Ratio and Hosmer–Lemeshow tests. To compare the statistical significance of the regression coefficients, we applied Wald's chi-squared test. For the global evaluation of the validity of the models, we used Nagelkerke's adjusted coefficient of determination, along with the percentage of correct classifications. Since incorrect inferences can be drawn if correlations between observations with samples of twins are ignored, we validated the binary logistic regression models using a random sample selected from the total sample, comprising twins from different couples. Bearing in mind that, by halving the sample, statistical power is lost, we compared the results of these analyses with those obtained using bootstrapping techniques [54], with 1000 samples per analysis to simulate sampling and with robust estimations of standard errors, statistical significance, and confidence intervals for regression coefficients. All analyses were executed using version 23 of the Statistical Package for the Social Sciences (IBM SPSS).

3. Results

Table 1 provides the descriptive statistics for the variables from the study sample. A primary/pre-secondary level of education was shown by 51 mothers (41.1%) and 58 fathers (46.8%); an intermediate level of education (secondary or high school, technical and non-technical) was shown by 38 mothers (30.6%) and 40 fathers (32.3%); and 35 mothers (28.2%) and 26 fathers (21%) had a higher education (university and graduate).

A total of 84 children were born vaginally (67.7%) and 40 via caesarean section (32.3%). An elected caesarean was indicated in 11 deliveries on account of maternal problems (preeclampsia, premature rupture of membranes, pregnancy following one or more caesarean births previously, abnormal contractions of the myometrium, or prolonged pushing), whereas in 29 deliveries, caesarean was indicated on account of foetal problems (first twin in wrong position, non-progression, or decline in foetal wellbeing). Foetal presentation was cephalic in 80 deliveries (64.5%) and non-cephalic (breech and transverse) in 44 deliveries (35.5%).

Maternal age at the time of delivery ranged from 22 to 45 years of age (M = 33.2 and SD = 4.29). Gestational age of the newborn was between 32 and 41 weeks (M = 35.14 and SD = 2.09). The score obtained in the 1-minute Apgar was between 4 and 10 points (M = 8.41 and SD = 1.18), and newborn weight was between 1179 and 3080 g (M = 2137.76 and SD = 432.79)

Table 1 also provides descriptions of the variables from the study validation sample, showing that the characteristics of both samples are equivalent.

Table 2 summarises the description and comparisons between the means of the originally quantitative dependent variables (reading accuracy, phonetic orthography, and visual orthography) and the control variables (maternal age, gestational age, 1-minute Apgar, newborn weight, and foetal presentation); in accordance with the independent variable type of delivery (vaginal or caesarean), we observed statistically significant and moderate differences for reading accuracy (U = 1119.50, $z = -2.99$, $p < 0.01$, r = 0.26) between the mean rank of children born vaginally (MR = 69.17) and those born by caesarean section (MR = 48.49). The same was true for phonetic orthography (U = 1025.00, $z = -3.50$, $p < 0.001$, r = 0.31) between the mean ranks of the groups made up of the variable type of delivery, (MR = 70.30) vs. (MR = 46.13). Additionally, groups were different with respect to visual orthography (U = 1262.00, $z = -2.24$, $p < 0.05$, r = 0.20), (MR = 67.48) vs. (MR = 52.05).

With regard to the control variables, there were no statistically significant differences regarding maternal age (U = 1404.00, $z = -1.48$, p = 0.139, r = 0.13) between the mean rank of mothers who delivered by caesarean section (MR = 69.40) and those who delivered vaginally (MR = 59.21); for gestational age (U = 1664.00, $z = -0.08$, p = 0.932, r = 0.01) between the mean rank of children born vaginally (MR = 62.31) and those born by caesarean section (MR = 62.90); for Apgar 1 (U = 1586.00, $z = -0.51$, p = 0.612, r = 0.04) between the mean ranks of the two different groups, (MR = 61.11) vs. (MR = 63.85); or in newborn weight (t (122) = 0.62, p = 0.535, d = 0.12), as found by González-Valenzuela, González-Mesa et al. [42] and González-Valenzuela, López-Montiel et al. [43].

To detect possible interactions between the independent variable type of delivery (vaginal/caesarean) and the potentially masking control variables for the main effect (maternal age, gestational age, foetal presentation, newborn weight, and Apgar 1), we conducted bivariate analyses as seen in González-Valenzuela, González-Mesa et al. [42] and González-Valenzuela, López-Montiel et al. [43]. The only statistically significant relationship found was between type of delivery and foetal presentation ($\chi^2(2, N = 124) = 45.53, p < 0.001, V = 0.60$), with a large magnitude observed for this association (see Table 2).

To assess whether the type of delivery of one twin was related to the type of delivery of the other twin, a contingency table was also performed, considering the Pearson χ^2 test of independence as well as the unadjusted OR and its 95% confidence interval. A total of 12 children (28.6%) were delivered vaginally after their brother was delivered by caesarean section, and another 12 children (60%) had a caesarean delivery after their brother had a vaginal delivery. On the other hand, 30 siblings were born both by vaginal delivery (71.4%) and 8 by caesarean section (40%). No statistically significant relationship was found between the type of birth of both children ($\chi^2(1, N = 62) = 0.81, p = 0.368, V = 0.11$). Therefore, the probability of being born by one type of delivery as a function of the other was not significant, OR = 1.66, 95% CI (0.54, 5.09).

In short, there were statistically significant differences observed related to the type of delivery in each one of the dependent variables (reading accuracy, phonetic orthography, and visual orthography). The mean values were significantly higher and with a medium effect size in children born vaginally compared with children born by caesarean section. For the control variables, the mean value was only significantly higher, with a large effect size, for the variable foetal presentation, where cephalic presentation was more frequent in vaginal deliveries, and non-cephalic delivery was more frequent in caesarean deliveries.

With the validation sample, the results of these preliminary analyses were also very similar to those of the study sample (see Table 2). The robust estimations of the statistics calculated by means of the 1000 sampling simulation samples also confirmed these results.

Tables 3–5, below, summarise the bivariate analyses between the independent variable, which is the type of delivery, and the control variables with each dependent variable (presence/absence of difficulty in reading accuracy, phonetic orthography, and visual orthography), respectively, having previously dichotomised the latter variables. In each table, the frequency distributions are presented for each level of the independent variables according to each of the dependent variables, the percentage of cells with expected frequencies below five, Pearson's χ^2 statistic, with its statistical significance and effect size, and the unadjusted OR and CI. ORs with intervals that did not contain the null value were considered significant (OR = 1). The percentage of cells with expected frequencies not less than five was only found in the relationship between newborn weight and each dependent variable.

As shown in Table 3, between the independent variable and the criterion variable and the control variables, the relationship between type of delivery and reading accuracy was statistically significant and moderate ($\chi^2(2, N = 124) = 9.64, p < 0.\ 01, V = 0.28$). Of the 40 (32.3%) children born by caesarean section, 17 (42.5%) did not pass the reading accuracy test, whereas of the 84 (67.7%) born by vaginal delivery, 14 (16.7%) did not pass. We estimated the OR to evaluate the frequency of difficulties in learning reading accuracy present in children who were born by caesarean section in comparison with children born vaginally. This raw measure seems to indicate that birth by caesarean triples the likelihood of presenting difficulties in reading accuracy, OR = 3.69, 95% CI (1.58, 8.64). We did not observe any statistically significant relationships between reading accuracy and the other control variables evaluated.

Table 5. Contingency tables and bivariate associations between visual orthography, type of delivery and the control variables.

Variables	Categories	Total N = 124	Dependent Variable Visual orthography		Test [1]	Sig.	ES [2]	OR	95% CI	
			No VOD n = 93 (75%)	VOD n = 31 (25%)					Lower	Upper
Independent **Type of delivery**	Vaginal Caesarean	84 (67.7%) 40 (32.3%)	69 (82.1%) 24 (60%)	15 (17.9%) 16 (40%)	7.08 [a]	0.008	0.24	3.06	1.32	7.13
Control Maternal age (years)	Under 35 Over 35	88 (71%) 36 (29%)	67 (76.1%) 26 (72.2%)	21 (23.9%) 10 (27.8%)	0.21 [a]	0.648	0.04	1.23	0.51	2.95
Gestational age of newborn (weeks)	Over 37 Under 37	40 (32.3%) 84 (67.7%)	29 (72.5%) 64 (76.2%)	11 (27.5%) 20 (23.8%)	0.19 [a]	0.657	0.04	0.82	0.35	1.94
Foetal presentation	Cephalic Non-cephalic	80 (64.5%) 44 (35.5%)	64 (80%) 29 (65.9%)	16 (20%) 15 (34.1%)	3.00 [a]	0.083	0.15	2.07	0.90	4.74
Weight of newborn (grams)	Over 1500 Under 1500	112 (90.3%) 12 (9.7%)	85 (75.9%) 8 (66.7%)	27 (24.1%) 4 (33.3%)	0.49 [b]	0.493	0.06	1.57	0.44	5.64
Apgar 1	Over 7 Under 7	99 (79.8%) 25 (20.2%)	77 (77.8%) 16 (64%)	22 (22.2%) 9 (36%)	2.00 [a]	0.155	0.13	1.97	0.76	5.06

Note. VOD = Visual Orthography Difficulty; ES = effect size; OR = odds ratio; CI = confidence interval. [1] Pearson's χ^2. [2] Cramer's V (Cramer's V reference values for *df less* = 1: small = 0.10; medium = 0.30; large = 0.50). [a] 0% of cells with an expected frequency less than 5.

Table 4 shows that the main relationship between type of delivery and phonetic orthography was also statistically significant and the magnitude of the relationship moderate ($\chi^2(2, N = 124) = 8.59, p < 0.01, V = 0.26$). Of the 40 (32.3%) children born by caesarean section, 17 (42.5%) did not pass the phonetic orthography test, and of the 84 (67.7%) children born vaginally, 15 (17.9%) did not pass. The OR seems to indicate that caesarean delivery triples the likelihood of presenting difficulties in this dependent variable, OR = 3.40, 95% CI (1.47, 7.87) (see Table 3). The same was true between foetal presentation and phonetic orthography ($\chi^2(2, N = 124) = 8.12, p < 0.01, V = 0.25$). Of the 44 (35.5%) children born following non-cephalic presentation, 18 (40.9%) did not succeeded phonetic orthography test, and 14 (17.5 %) out of the 80 (64.5%) born following cephalic presentation did not pass. The OR was statistically significant, OR = 3.26, 95% CI (1.42, 7.51), indicating once again a likelihood of presenting difficulties with regard to phonetic orthography approximately 3 times greater among children who were born following non-cephalic presentation.

Table 5 also shows a significant and moderate relationship between type of delivery and visual orthography ($\chi^2(2, N = 124) = 7.08, p < 0.01, V = 0.24$). Of the 40 (32.3%) children born by caesarean section, 16 (40%) did not pass the visual orthography test, and of the 84 (67.7%) born vaginally, 15 (17.9%) did not pass. In this case, the probability of presenting difficulties in visual orthography is approximately 3 times higher when the type of delivery was by caesarean section rather than vaginal, OR = 3.06, 95% CI (1.32, 7.13). We did not find any statistically significant relationships between visual orthography and the other control variables studied here. We observed similar results in the validation sample with regard to these bivariate analyses. All the relationships were statistically significant and moderate between type of delivery and difficulties with reading accuracy ($\chi^2(1, N = 62) = 4.58, p < 0.05, V = 0.27$), type of delivery and difficulties with phonetic orthography ($\chi^2(1, N = 62) = 5.20, p < 0.05, V = 0.29$), foetal presentation and difficulties with phonetic orthography ($\chi^2(1, N = 62) = 7.06, p < 0.01, V = 0.34$), as well as between type of delivery and difficulties with visual orthography ($\chi^2(1, N = 62) = 7.56, p < 0.05, V = 0.35$). However, on account of the sample size, in the distribution of frequencies, percentages of cells with expected frequencies no lower than five were found in the relationship between maternal age, newborn weight, and 1-minute Apgar and each dependent variable. The unadjusted ORs were also statistically significant, with a moderate effect size, according to their CI (see the ORs calculated in the univariate logistic regression analysis shown in Table 6).

Table 6. Multivariate logistic regression analysis for reading accuracy and phonetic and visual orthography disabilities, adjusted by potential interaction and confounding factors.

***N* = 124**	*Variables*	*b*	*SE*	*Wald* χ^2	*df*	*p*	**OR**	**95% CI**	
Study Sample								*Lower*	*Upper*
RAD Model 1	**Type of delivery** [(a)]	1.31	0.43	9.08	1	0.003	3.69 [1]	1.58	8.64
	Constant	−1.61	0.29	30.22	1	0.000	0.20		
* $\chi^2(1, N = 124) = 9.21$, $p = 0.002$; $R^2 = 0.10$; GPC = 75%									
POD Model 1	Type of delivery [(a)]	1.00	0.77	1.65	1	0.198	2.72	0.59	12.56
	Foetal presentation [(b)]	0.88	0.68	1.67	1	0.196	2.42	0.63	9.27
	Type of delivery x Foetal presentation	−0.38	1.05	0.13	1	0.712	0.68	0.08	5.30
	Constant	−1.69	0.33	26.75	1	0.000	0.18		
* $\chi^2(3, N = 124) = 10.20$, $p = 0.017$; $\chi^2(3, N = 124) = 0.00$, $p = 0.999$; $R^2 = 0.11$; GPC = 74.2%									
Model 2	Type of delivery [(a)]	0.79	0.53	2.19	1	0.139	2.20	0.77	6.25
	Foetal presentation [(b)]	0.72	0.53	1.85	1	0.173	2.05	0.73	5.81
	Constant	−1.66	0.31	28.90	1	0.000	0.19		
* $\chi^2(2, N = 124) = 10.06$, $p = 0.007$; $\chi^2(2, N = 124) = 0.13$, $p = 0.934$; $R^2 = 0.11$; GPC = 74.2%									
Model 3	**Type of delivery** [(a)]	1.22	0.43	8.16	1	0.004	3.40 [1]	1.47	7.87
	Constant	−1.52	0.28	28.69	1	0.000	0.21		
* $\chi^2(1, N = 124) = 8.23$, $p = 0.004$; $R^2 = 0.09$; GPC = 74.2%									
VOD Model 1	**Type of delivery** [(a)]	1.12	0.43	6.77	1	0.009	3.06 [1]	1.32	7.13
	Constant	−1.52	0.28	28.69	1	0.000	0.21		
* $\chi^2(1, N = 124) = 6.79$, $p = 0.009$; $R^2 = 0.08$; GPC = 75%									
***N*= 62**	*Variables*	*b*	*SE*	*Wald*χ^2	*df*	*p*	**OR**	**95% CI**	
Validation Sample								*Lower*	*Upper*
RAD Model 1	**Type of delivery** [(a)]	1.24	0.60	4.36	1	0.037	3.48 [1]	1.08	11.20
	Constant	−1.45	0.39	13.56	1	0.000	0.23		
* $\chi^2(1, N = 62) = 4.41$, $p = 0.036$; $R^2 = 0.10$; GPC = 72.6%									
POD Model 1	Type of delivery [(a)]	0.36	1.23	0.08	1	0.772	1.43	0.13	15.87
	Foetal presentation [(b)]	1.05	1.00	1.09	1	0.296	2.86	0.40	20.47
	Type of delivery x Foetal presentation	0.30	1.61	0.03	1	0.852	1.35	0.06	31.77
	Constant	−1.69	0.33	26.75	1	0.000	0.18		
* $\chi^2(3, N = 62) = 7.36$, $p = 0.006$; $\chi^2(3, N = 62) = 0.00$, $p = 0.999$; $R^2 = 0.16$; GPC = 72.6%									
Model 2	Type of delivery [(a)]	0.53	0.78	0.46	1	0.498	1.70	0.37	7.85
	Foetal presentation [(b)]	1.17	0.77	2.28	1	0.131	3.21	0.70	14.67
	Constant	−1.47	0.41	13.15	1	0.000	0.23		
* $\chi^2(2, N = 62) = 7.32$, $p = 0.026$; $\chi^2(2, N = 62) = 0.03$, $p = 0.983$; $R^2 = 0.16$; GPC = 72.6%									
Model 3	**Type of delivery** [(a)]	1.30	0.58	4.94	1	0.026	3.67 [1]	1.17	11.52
	Constant	−1.30	0.37	11.94	1	0.001	0.27		
* $\chi^2(1, N = 62) = 5.04$, $p = 0.025$; $R^2 = 0.11$; GPC = 69.4%									
VOD Model 1	**Type of delivery** [(a)]	1.61	0.61	6.97	1	0.008	5.00[1]	1.51	16.51
	Constant	−1.61	0.41	15.11	1	0.000	0.20		
* $\chi^2(1, N = 62) = 7.26$, $p = 0.007$; $R^2 = 0.16$; GPC = 72.6%									

Note. RAD = Reading Accuracy Difficulty; POD = Phonetic Orthography Difficulty; VOD = Visual Orthography Difficulty; OR = odds ratio; CI = confidence interval. Variables reference categories: [(a)] = Vaginal delivery; [(b)] = Cephalic. [1] OR effect size as a function of the transformation to Cohen's *d* (Cohen's reference values: insignificant = less than 1.68; small = 1.68–3.47; moderate = 3.47–6.71; large = greater than 6.71). * Goodness-of-fit tests for logistic regression models: global test χ^2; Hosmer–Lemeshow χ^2; Nagelkerke R^2; GPC = global percentage of correct classifications.

We subsequently carried out binary logistic regressions for reading accuracy, phonetic orthography, and visual orthography according to the type of delivery to evaluate the degree to which this factor increased or decreased the risk of having difficulties in these aptitudes, statistically controlling for possible interactions and confounding factors. The results are summarised in Table 6.

With the independent variable difficulties in reading accuracy, the estimated model was significant ($\chi^2(1, N = 124) = 9.21, p < 0.01$), including only the independent, with an OR = 3.69, 95% CI (1.58, 8.64). In the model presented, birth by caesarean section presented a risk factor for difficulties in reading accuracy: the risk was 3.69 times higher among those born by caesarean sector than those born vaginally in twin births once these children have reached the age of 6. The magnitude of the effect of this odds ratio was moderate, according to the transformation into Cohen's *d*. Regarding the explanatory capacity of this model, 10% of the variability observed in the response variable was explained by the estimated logistic regression model (Nagelkerke $R^2 = 0.10$). The percentage of cases that could be correctly predicted was 75% (see Table 6).

Secondly, we adjusted the main relationship studied between type of delivery and difficulties in phonetic orthography for foetal presentation and interaction between type of delivery and foetal presentation. Since foetal presentation was related statistically to the independent variable ($\chi^2(2, N = 124) = 45.53, p < 0.001$) and the dependent variable (see Table 4), it was considered a potential modifying variable of the main effect studied, as well as a potential confounding variable. According to this adjustment, having eliminated in two successive steps the interaction term and the variable foetal presentation, the third estimated model was statistically significant ($\chi^2(1, N = 124) = 8.23, p <0.01$), where the independent variable was significant with an OR = 3.40, 95% CI (1.47, 7.87). Model 2 shows that the effect of the variable type of delivery on difficulties in phonetic orthography was modified in the presence of the variable foetal presentation, but the ORs were not statistically significant (they included unity), and so it was not included in the final model. Caesarean birth was a risk factor for difficulties in learning phonetic orthography, making the risk 3.40 times higher among those born by caesarean section than the children born through vaginal delivery in twin births, once the children reached the age of 6. According to the OR, the effect size was considered small. In this model, the estimated model explained 9% of variance in the variable reading accuracy (Nagelkerke $R^2 = 0.09$). The percentage of cases it could correctly predict was 74.2% (see Table 6).

With the dependent variable visual orthography, two cases were detected close to 3 standard deviations below the mean. A posteriori, the regression model for visual orthography was adjusted both with these two cases and without them, not appreciating significant differences in the regression coefficients, in the OR, or in the adjustment values of the models between both procedures. Therefore, it was decided that these two cases could be part of the sample.

The estimated model was significant ($\chi^2(1, N = 124) = 6.79, p < 0.01$), including only the independent variable type of delivery, with an OR = 3.06 and 95% CI (1.32, 7.13). Therefore, the risk of having difficulties in the learning of visual orthography was 3.06 times higher among those born by caesarean than those children born via vaginal delivery in twin births, once the children reached the age of 6. According to the OR, the effect size was also considered small. In the estimated model, the variable type of delivery explained 8% of the variance in the variable visual orthography (Nagelkerke $R^2 = 0.08$). The percentage of cases correctly predicted was 75% (see Table 6).

With the random sample of 62 subjects, comprising one twin from each couple, we found similar results to those obtained with the total sample, for all the variables studied (see Table 6).

In summary, the differences between the mean values show significantly higher scores in the learning variables for children born vaginally. In bivariate associations, type of delivery had a significant effect on the probability of presenting difficulties in the learning of reading accuracy, phonetic orthography, and visual orthography. With regard to the other gestational, obstetric, and neonatal variables, only foetal presentation appeared as a potential modifying and confounding variable for the main effect between type of delivery and phonetic orthography. However, although related to both, when it was controlled statistically by means of logistic regression, the effect of the first on the second was not modified. Therefore, in line with the preliminary statistical analysis, only the variable type of delivery was shown as a possible risk factor for disabilities in the reading and writing measures taken into account in this study.

4. Discussion

The objective of this cohort study of children born in twin births once they reached the age of 6 was to analyse retrospectively the relationship between reading accuracy and phonetic and visual orthography with type of delivery, evaluating the degree to which type of delivery may be related to the risk of having difficulties in these basic and specific components of reading and writing, and controlling statistically for possible interacting and confounding factors.

We observed that type of delivery was related statistically to the learning disabilities found, constituting a risk factor. The risk of having difficulties in reading accuracy and phonetic and visual orthography can be around 3 times higher among the children born by caesarean section than those born through vaginal delivery. The magnitude of the effect of type of delivery was moderate in relation to reading accuracy, and small in relation to phonetic and visual orthography. The explanatory capacity of variance in reading and writing learning disabilities was discreet, as was the percentage of cases that the final models could predict adequately, in accordance with final models that only included one independent variable.

In other words, in multiple births where caesarean deliveries must be performed, there are certain obstetric and perinatal circumstances that might affect the neurological development of the baby [17]. These types of conditions would justify in the long term the linguistic and cognitive difficulties that characterise the difficulties in reading accuracy and in phonetic and visual orthography. Recent studies have found that type of delivery is related to and is also a risk factor for neuropsychological development disorders and intellectual alterations in twin births [41–43]. Furthermore, some studies highlight the existence of short-term neonatal morbidity related with caesarean delivery, describing high rates of neonatal hypoxia, transient tachypnoea, and respiratory distress syndrome in children born by caesarean section [38], with a potential influence on posterior neurocognitive development. Various studies also highlight an increase in long-term postnatal morbidity, with an increase in respiratory morbidity, such as asthma or obstructive apnoea, diabetes, and obesity [3,6,41]. Our results also support the existence of circumstances that unfavourably condition the development of children born by caesarean section and justify the cognitive and linguistic difficulties they present [3,6,19,41–45,55]. Although the physio-pathological mechanisms underlying the deficits described are not clear, it would appear that the most striking difference between children born vaginally and those born via a programmed caesarean section is the neuroendocrine response to the stress produced by contractions, conditioned by normal delivery [6,56]. These differences in the neuroendocrine response to stress have been linked, in the case of programmed caesarean births, with the existence of defective expressions of certain genes (UCP2) in the neurons on the foetal hippocampus [55], with differences in the concentrations of dopamine depending on the type of delivery in certain areas of the prefrontal cortex, the nucleus accumbens, and striatum [57,58] and with differences in concentrations of noradrenaline in the adult amygdala and thalamus [59].

It should be noted that, in our study, gestational age and newborn weight did not affect reading and writing variables, as other studies have found [42,45]. This might be due to the fact that the choice of the sample excluded extremely premature or very premature subjects, whose psychological development and academic performance might truly be affected [22,23,25,27,30].

In short, although in multiple births caesarean section delivery is a risk factor for neuropsychological development disorders and specific learning difficulties [41–45], the results found in this study also indicate risk in the basic components of reading and writing related to separate lexical and sub-lexical processes, such as reading accuracy and phonetic and visual orthography. These results could have important repercussions in the explanation of dyslexia and are in line with the findings put forward by some studies about the influence of perinatal factors on school learning [19,20,41–45]. However, they should be taken with caution since they do not take into account other potentially influential obstetric and perinatal variables.

Future studies should be conducted using broader samples in order to adequately control for variables, such as newborn weight, so that the findings can be generalized.. Furthermore, in order to analyse which other types of explanations could lead to the appearance of specific components of reading and writing difficulties, it would also be advisable to include other obstetric and perinatal variables as possible risk factors (e.g., congenital infection, antenatal drug/toxin exposure, respiratory distress, hyperbilirubinaemia, etc.). Some research has found that the risk of exhibiting reading and spelling problems among children with normal intelligence levels is increased when there is perinatal asphyxia [60] and that neonatal hyperbilirubinaemia is a risk factor for problems with reading, writing, and mathematics [61]. Along this line, a maternal exposure to nicotine during pregnancy is related with the DYX1C1 gene and would justify problems decoding words in reading and writing [8,62,63]. The role played by the reason for caesarean section should also be taken into consideration—and this is a limitation of the present study—since foetal stress caused by possible infection, deficient blood flow, etc., is probably what puts children at greater risk of developing learning disabilities, although the decision to opt for a caesarean is made the moment the attending physicians observe that this situation might occur, using their clinical judgement. Knowing the influence of sociodemographic variables (parents' level of education and profession, etc.) and their interaction with the aforementioned perinatal and obstetric variables would also be useful in terms of analysing which other types of explanations might give rise to the appearance of specific difficulties in reading and writing.

Furthermore, the findings of this study could be useful in clinical practice since they support the avoidance of caesarean section on demand or without specialised indication, in order to avoid in the long term the appearance of specific difficulties in reading (reading accuracy) and writing (phonetic and visual orthography). They also point to the advantages of vaginal delivery in multiple pregnancies, provided it is not contraindicated.

In conclusion, this study opens up new possibilities for research since the type of delivery has consequences in the learning of reading and writing among students born in twin births. Although, according to the results, clinical relevance is not high, it is also not insignificant and should not be ignored. Many factors are involved in the choice of delivery in twin births, and these must be studied (reason for caesarean, congenital infection, hyperbilirubinaemia, respiratory distress, etc.). Therefore, further research is needed with larger samples that will provide more accurate information about the relevance of such factors.

Author Contributions: Conceptualization, M.-J.G.-V., D.L.-M., O.C.-G. and E.-S.G.-M.; methodology, M.-J.G.-V., D.L.-M. and E.-S.G.-M.; software, M.-J.G.-V., D.L.-M., O.C.-G. and E.-S.G.-M.; formal analysis, D.L.-M.; investigation, M.-J.G.-V., D.L.-M., O.C.-G. and E.-S.G.-M.; resources, M.-J.G.-V., D.L.-M. and E.-S.G.-M.; data curation, M.-J.G.-V., D.L.-M., O.C.-G. and E.-S.G.-M.; writing—original draft preparation, M.-J.G.-V., D.L.-M. and E.-S.G.-M.; writing—review and editing, M.-J.G.-V. and D.L.-M.; visualization, M.-J.G.-V., D.L.-M., O.C.-G. and E.-S.G.-M.; supervision, M.-J.G.-V. and D.L.-M.; project administration, M.-J.G.-V. and O.C.-G.; funding acquisition, M.-J.G.-V. and E.-S.G.-M. All authors have read and agreed to the published version of the manuscript.

Funding: This study has been funded by the research group SEJ 521 of the Andalusian government and by the Department of Surgical Specialties, Biochemistry and Immunology of University of Málaga.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by Research Ethics Committee at the Hospital Materno Infantil (Comité Ético de Investigación-CEI), on 30 January 2014.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: No data availability. This manuscript does not have associated data in a data repository.

Conflicts of Interest: In the realization of this study, there is no conflict of interest.

References

1. Allin, A.B.; Fischbein, S. Twins: Are they at risk? A longitudinal study of twins and nontwins from birth to 18 years of age. *Acta Genet. Med. Gemellol.* **1991**, *40*, 29–40. [CrossRef]
2. Carlsson Wallin, M.; Ekström, P.; Marsál, K.; Källé, K. Apgar score and perinatal death after one previous caesarean delivery. *BJOG* **2010**, *117*, 1088–1097. [CrossRef] [PubMed]
3. Hogle, K.L.; Hutton, E.K.; McBrien, K.A.; Barrett, J.F.; Hannah, M.E. Cesarean delivery for twins: A systematic review and meta-analysis. *Am. J. Obstet. Gynecol.* **2003**, *188*, 220–227. [CrossRef] [PubMed]
4. Liston, F.A.; Allen, V.M.; O'Connell, C.M.; Jangaard, K.A. Neonatal outcomes with caesarean delivery at term. *Arch. Dis. Child Fetal Neonatal Ed.* **2008**, *93*, 176–182. [CrossRef]
5. Olusanya, B.O. Perinatal outcomes of multiple births in southwest Nigeria. *J. Health Popul. Nutr.* **2011**, *29*, 639–647. [PubMed]
6. Villar, J.; Carroli, G.; Zavaleta, N.; Donner, A.; Wojdyla, D.; Faundes, A.; Velazco, A.; Bataglia, V.; Langer, A.; Narváez, A.; et al. Maternal and neonatal individual risks and benefits associated with caesarean delivery: Multicentre prospective study. *BMJ* **2007**, *335*, 1025. [CrossRef]
7. American Psychiatric Association (APA). *Diagnostic and Statistical Manual of Mental Disorders*, 5th ed.; American Psychiatric Publishing: Arlington, VA, USA, 2014.
8. Becker, N.; Vasconcelos, M.; Oliveira, V.; Santos, F.; Bizarro, L.; Almeida, R.; Salles, J.F.; Carvalho, M. Genetic and environmental risk factors for developmental dyslexia in children: Systematic review of the last decade. *Dev. Neuropsychol.* **2017**, *42*, 423–445. [CrossRef]
9. Ramus, F.; Altarelli, I.; Jednoróg, K.; Zhao, J.; Scotto di Covella, L. Neruoanatomy of developmental dyslexia: Pitfalls and promise. *Neurosci. Biobehav. Rev.* **2018**, *84*, 434–452. [CrossRef] [PubMed]
10. Richlan, F.; Kronbichler, M.; Wimmer, H. Functional abnormalities in the dyslexic brain: A quantitative meta-analysis of neuroimaging studies. *Hum. Brain Mapp.* **2009**, *30*, 3299–3308. [CrossRef] [PubMed]
11. Richlan, F.; Kronbichler, M.; Wimmer, H. Meta-analyzing brain dysfunctions in dyslexic children and adults. *Neuroimage* **2011**, *56*, 1735–1742. [CrossRef] [PubMed]
12. Shaywitz, S.E.; Shaywitz, B.A. *Nelson Textbook of Pediatrics*, 21st ed.; Kliegman, R.M., Stanton, B.F., Geme, J.W., St., Schor, N.F., Eds.; Saunders Elsevier: Philadelphia, PA, USA, 2020; Volume 1, pp. 267–269.
13. Fletcher, J.M. Dyslexia: The evolution of a scientific concept. *J. Int. Neuropsychol. Soc.* **2009**, *15*, 501–508. [CrossRef]
14. Peterson, R.L.; Pennington, B.F. Developmental Dyslexia. *Annu. Rev. Clin. Psychol.* **2015**, *11*, 283–307. [CrossRef]
15. Colletti, L.F. Relationship between pregnancy and birth complications and the later development of learning disabilities. *J. Learn. Disabil.* **1979**, *12*, 25–29. [CrossRef]
16. Hill, S.K.; Cawthorne, V.; Dean, R.S. Utilitiy of the Maternal Perinatal Scale (MPS) in distinguishing normal from learning disabled children. *Int. J. Neurosci.* **1998**, *95*, 141–154. [CrossRef]
17. Liu, L.; Wang, J.; Shanshan, S.; Luo, X.; Kong, R.; Zhang, X.; Song, R. Descriptive epidemiology of prenatal and perinatal risk factors in a Chinese population with reading disorder. *Sci. Rep.* **2016**, *6*, 36697. [CrossRef] [PubMed]
18. Xue, J.A. Perinatal Complications as Predictors of Neuropsychological Outcome in Children with Learning Disabilities. Ph.D. Thesis, Ball State University, Muncie, IN, USA, 1996.
19. Labouesse, M.A.; Langhans, W.; Meyer, U. Long-term pathological consequences of prenatal infection: Beyond brain disorders. *Am. J. Physiol. Regul. Integr. Comp. Psysiol.* **2015**, *309*, 1–12. [CrossRef]
20. Ranasinghe, S.; Or, G.; Wang, E.Y.; Levins, A.; McLean, M.A.; Niell, C.M.; Chau, V. Reduced cortical activity impairs development and plasticity after neonatal hypoxia ischemia. *J. Neurosci.* **2015**, *35*, 11946–11959. [CrossRef] [PubMed]
21. Rosen, M.G.; Debanne, S.M.; Thompson, K.; Dickinson, J.C. Abnormal labor and infant brain damage. *Obstet. Gynecol.* **1992**, *80*, 961–965. [PubMed]
22. Chyi, L.J.; Lee, H.C.; Hintz, S.R.; Gould, J.B.; Sutcliffe, T.L. School outcomes of late preterm infants: Special needs and challenges for infants born at 32 to 36 weeks gestation. *J. Pediatr.* **2010**, *153*, 25–31. [CrossRef]
23. Guarini, A.; Sansavini, A.; Fabbri, C.; Savini, S.; Alessandroni, R.; Faldella, G.; Karmiloff-Smith, A. Long-term effects of preterm birth on language and literacy at eight years. *J. Child Lang.* **2010**, *37*, 865–885. [CrossRef]

24. Hall, A.; McLeod, A.; Counsell, C.; Thomson, L.; Mutch, L. School attainment, cognitive ability and motor function in a total Scottish very-low-birthweight population at eight years: A controlled study. *Dev. Med. Child Neurol.* **1995**, *37*, 1037–1050. [CrossRef]
25. Jansson-Verkasalo, E.; Korpilathi, P.; Jäntti, V.; Valkama, M.; Vainionpää, L.; Alku, P.; Suominen, K.; Näätänen, R. Neurophysiologic correlates of deficit phonological representations and object naming in prematurely born children. *Clin. Neurophysiol.* **2004**, *115*, 179–187. [CrossRef] [PubMed]
26. Johnson, S.; Wolke, D.; Hennessy, E.; Marlow, N. Educational outcomes in extremely preterm children: Neuropsychological correlates and predictors of attainment. *Dev. Neuropsychol.* **2011**, *36*, 74–95. [CrossRef] [PubMed]
27. Kovachy, V.N.; Adams, J.N.; Tamaresis, J.S.; Feldman, H.M. Reading abilities in school-aged preterm children: A review and meta-analysis. *Dev. Med. Child Neurol.* **2015**, *57*, 410–419. [CrossRef] [PubMed]
28. Marlow, N.; Wolke, D.; Bracewell, M.; Samara, M.; EPICure Study Group. Neurologic and developmental disability at six years of age after extremely preterm birth. *N. Engl. J. Med.* **2005**, *352*, 9–19. [CrossRef] [PubMed]
29. Morse, S.B.; Zheng, H.; Tang, Y.; Roth, J. Early school age outcomes of late preterm infants. *Pediatrics* **2009**, *123*, 622–629. [CrossRef]
30. Roberts, G.; Lim, J.; Doyle, L.W.; Anderson, P.J. High rates of school readiness difficulties at 5 years of age in very preterm infants compared with term controls. *J. Dev. Behav. Pediatr.* **2011**, *32*, 117–123. [CrossRef]
31. Wolke, D.; Samara, M.; Bracewell, M.; Marlow, N.; EPICure Study Group. Specific language difficulties and school achievement in children born at 25 weeks of gestation or less. *J. Pediatr.* **2008**, *152*, 256–262. [CrossRef]
32. Holm, A.; Crosbie, S. Literacy skills of children born preterm. *Aust. J. Learn. Diffic.* **2010**, *15*, 51–69. [CrossRef]
33. Russell, R.B.; Petrini, J.R.; Damus, K.; Mattison, D.R.; Schwarz, R.H. The changing epidemiology of multiple births in the United States. *Obstetrics Gynecol.* **2003**, *101*, 129–135. [CrossRef]
34. Cleary-Goldman, J.; Malone, F.D.; Vidaver, J.; Ball, R.H.; Nyberg, D.A.; Comstock, C.H.; Saade, G.R.; Eddleman, K.A.; Klugman, S.; Dugoff, L.; et al. Impact of maternal age on obstetric outcome. *Obstet. Gynecol.* **2005**, *105 Pt 1*, 983–990. [CrossRef] [PubMed]
35. McLennan, A.S.; Gyamfi-Bannerman, C.; Ananth, C.V.; Wright, J.D.; Siddiq, Z.; D'Alton, M.E.; Friedman, A.M. The role of maternal age in twin pregnancy outcomes. *Am. J. Obstet. Gynecol.* **2017**, *217*. [CrossRef]
36. Bogner, G.; Wallner, V.; Fazelnia, C.; Strobl, M.; Volgger, B.; Fischer, T.; Jacobs, V.R. Delivery of the second twin: Influence of presentation on neonatal outcome, a case controlled study. *BMC Pregnancy Childbirth* **2018**, *18*, 176. [CrossRef]
37. Seelbach-Goebel, B. Twin Birth Considering the Current Results of the "Twin Birth Study". *Geburtshilfe Frauenheilkd.* **2014**, *74*, 838–844. [CrossRef] [PubMed]
38. Barrett, J.F.; Hannah, M.E.; Hutton, E.K.; Willan, A.R.; Allen, A.C.; Armson, B.A.; Gafni, A.; Joseph, K.S.; Mason, D.; Ohlsson, A.; et al. A randomized trial of planned cesarean or vaginal delivery for twin pregnancy. *N. Engl. J. Med.* **2013**, *369*, 1295–1305. [CrossRef] [PubMed]
39. Pasupathy, D.; Smith, G.C. Neonatal outcomes with caesarean delivery al term. *Arch. Dis. Child Fetal Neonatal Ed.* **2008**, *93*, 174–175. [CrossRef]
40. Robbins, L.S.; Blanchard, C.T.; Biasini, F.J.; Powell, M.F.; Casey, B.M.; Tita, A.T.; Harper, L.M. General anesthesia for cesarean delivery and childhood neurodevelopmental and perinatal outcomes: A secondary analysis of a randomized controlled trial. *Int. J. Obstet. Anesth.* **2020**, *45*, 34–40. [CrossRef]
41. González-Mesa, E.; Cazorla-Granados, O.; González-Valenzuela, M.J. The influence of obstetric variables on school achievement, intelligence and neuropsychological development in a sample of Spanish twins at the age of six: A retrospective study. *J. Matern. Fetal Neonatal Med.* **2016**, *29*, 1595–1602. [CrossRef]
42. González-Valenzuela, M.J.; González-Mesa, E.; Cazorla-Granados, O.; López-Montiel, D. Type of Delivery, Neuropsychological Development and Intelligence in Twin Births. *Front. Psychol.* **2019**, *10*, 972. [CrossRef]
43. González-Valenzuela, M.J.; López-Montiel, D.; Cazorla-Granados, O.; González-Mesa, E. Type of delivery and reading, writing, and arithmetic learning in twin births. *Dev. Psychobiol.* **2019**, *64*, 484–495. [CrossRef]
44. Fletcher, J.M.; Coulter, W.A.; Reschly, D.J.; Vaughn, S. Alternative approaches to definition and identification on learning disabilities: Some questions and answers. *Ann. Dyslexia.* **2004**, *54*, 304–331. [CrossRef] [PubMed]
45. García-Vidal, J.; González-Manjón, D.; Ortiz, B.G. *Batería Psicopedagógica EVALÚA-1*; Psychopedagogical Battery EVALUA-1; EOS: Madrid, Spain, 2011.
46. González-Valenzuela, M.J.; Martín-Ruiz, I. Effects on Reading of an Early Intervention Program for Spanish Children at Risk of Learning Difficulties. *RASE* **2017**, *38*, 67–75. [CrossRef]
47. Welsch, R.G. Using experimental analysis to determine interventions for reading fluency and recalls of students with learning disabilities. *Learn. Disabil. Q.* **2007**, *30*, 115–129.
48. Kaufman, A.S.; Kaufman, N.L. *Test Breve de Inteligencia de Kaufman (K. BIT)*; TEA Ediciones: Madrid, Spain, 2000.
49. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed.; Lawrence Erlbaum: Hillsdale, NJ, USA, 1998.
50. Tomczak, A.; Tomczak, E. The need to report effect size estimates revisited. An overview of some recommended measures of effect size. *Trends Sport Sci.* **2014**, *1*, 19–25.
51. Hosmer, D.W.; Lemeshow, S. *Applied Logistic Regression*, 2nd ed.; Wiley: New York, NY, USA, 2000.
52. Kleinbaum, D.G.; Klein, M. *Logistic Regression: A Self-Learning Text*; Springer: New York, NY, USA, 2001.
53. Chen, H.; Cohen, P.; Chen, S. How big is a big odds ratio? Interpreting the magnitudes of odds ratios in epidemiological studies. *Commun. Stat. Simul. Comput.* **2010**, *39*, 860–864. [CrossRef]

54. Efron, B.; Tibshirani, R.J. *An Introduction to the Bootstrap*; Chapman and Hall: New York, NY, USA, 1993. [CrossRef]
55. Simón-Areces, J.; Dietrich, M.; Hermes, G.; García-Segura, L.; Arévalo, M.; Horvath, T. Ucp2 induced by natural birth regulates neuronal differentiation of the hippocampus and related adult behavior. *PloS ONE* **2012**, e42911. [CrossRef]
56. Saccone, G.; Berghella, V. Antenatal corticosteroids for maturity of term or near term fetuses: Systematic review and meta-analysis of randomized controlled trials. *BMJ* **2016**, *12*, 355. [CrossRef]
57. El-Khodor, B.F.; Boksa, P. Long-term reciprocal changes in dopamine levels in prefrontal cortex versus nucleus accumbens in rats born by Caesarean section compared to vaginal birth. *Exp Neurol.* **1997**, *145*, 118–129. [CrossRef]
58. Vaillancourt, C.; Boksa, P. Birth insult alters dopamine-mediated behavior in a precocial species, the guinea pig. Implications for schizophrenia. *Neuropsychopharmacology* **2000**, *23*, 654–666. [CrossRef]
59. El-Khodor, B.F.; Boksa, P. Differential vulnerability of male versus female rats to long-term effects of birth insult on brain catecholamine levels. *Exp. Neurol.* **2003**, *182*, 208–219. [CrossRef] [PubMed]
60. Van Handel, M.; Swaab, H.; de Vries, L.S.; Jongmans, M.J. Long-term cognitive and behavioral consequences of neonatal encephalopathy following perinatal asphyxia: A review. *Eur. J. Pediatr.* **2007**, *166*, 645–654. [CrossRef] [PubMed]
61. Hokkanen, L.; Launes, J.; Michelsson, K. Adult neurobehavioral outcome of hyperbilirubinemia in full term neonates-a 30-year prospective follow-up study. *PeerJ* **2014**, *2*, e294. [CrossRef]
62. Cho, K.; Frijters, J.C.; Zhang, H.; Miller, L.L.; Gruen, J.R. Prenatal exposure to nicotine and impaired reading performance. *J. Pediatr.* **2013**, *162*, 713–718. [CrossRef]
63. Mascheretti, S.; Bureau, A.; Battaglia, M.; Simone, D.; Quadrelli, E.; Croteau, J.; Cellino, M.R.; Giorda, R.; Beri, S.; Maziade, M.; et al. An assessment of gene-by-environment interactions in developmental dyslexia-related phenotypes. *Genes Brain Behav.* **2013**, *12*, 47–55. [CrossRef] [PubMed]

children

MDPI

Article

Gaze Following and Pupil Dilation as Early Diagnostic Markers of Autism in Toddlers

Raquel Camero [1], Verónica Martínez [1,*] and Carlos Gallego [2]

1 Department of Psychology, University of Oviedo, 33003 Oviedo, Spain; uo217941@uniovi.es
2 Department of Experimental Psychology, Cognitive Processes and Speech Therapy, Complutense University of Madrid, 28223 Madrid, Spain; carlosgallego@ucm.es
* Correspondence: martinezveronica@uniovi.es

Abstract: *Background*: Children with autism spectrum disorder (ASD) show certain characteristics in visual attention. These may generate differences with non-autistic children in the integration of relevant social information to set the basis of communication. Reliable and objective measurement of these characteristics in a language learning context could contribute to a more accurate early diagnosis of ASD. Gaze following and pupil dilation are being studied as possible reliable measures of visual attention for the early detection of ASD. The eye-tracking methodology allows objective measurement of these biomarkers. The aim of this study is to determine whether measurements of gaze following and pupillary dilation in a linguistic interaction task are potential objective biomarkers for the early diagnosis of ASD. *Method*: A group of 20 children between 17 and 24 months of age, made up of 10 neurotypical children (NT) and 10 children with an increased likelihood of developing ASD were paired together according to chronological age. A human face on a monitor pronounced pseudowords associated with pseudo-objects. Gaze following and pupil dilation were registered during the task These measurements were captured using eye-tracking methodology. *Results*: Significant statistical differences were found in the time of gaze fixation on the human face and on the object, as well as in the number of gazes. Children with an increased possibility of developing ASD showed a slightly higher pupil dilation than NT children. However, this difference was not statistically significant. Nevertheless, their pupil dilation was uniform throughout the different periods of the task while NT participants showed greater dilation on hearing the pseudoword. *Conclusions*: The fixing and the duration of gaze, objectively measured by a Tobii eye-tracking system, could be considered as potential biomarkers for early detection of ASD. Additionally, pupil dilation measurement could reflect differential activation patterns during word processing in possible ASD toddlers and NT toddlers.

Keywords: language acquisition; autism; eye-tracker; pseudowords; pupillometry; gaze fixation

Citation: Camero, R.; Martínez, V.; Gallego, C. Gaze Following and Pupil Dilation as Early Diagnostic Markers of Autism in Toddlers. *Children* **2021**, *8*, 113. https://doi.org/10.3390/children8020113

Academic Editor: Eva Aguilar Mediavilla

Received: 13 December 2020
Accepted: 1 February 2021
Published: 5 February 2021

1. Introduction

From an early age, babies show a preference for maintaining visual contact with their parents and directing their attention towards the human voice and relevant social stimuli [1]. Moreover, it is stated that, at first, typically developing infants pay more attention to the eyes of their interlocutors [2]. Then, from 6–9 months of age, they go on to spend more time paying attention to the mouth, as they begin to specialize in language and build phonological system. After that, between 12 and 15 months, they direct more attention to the eyes [3]. This allows them to learn and develop the keys to social learning. They also fix their attention on the mouth when they do not know the word, or when they hear a foreign word [2], presumably to get support in visual information [4]. The fact that they pay attention to the eyes and the mouth of the people who interact with them is key to learning phonological and lexical references. It allows babies to construct social knowledge, fundamental to their neurological development and the learning of language [5]. Evidence exists that contact and gaze following act as a precursor to the acquisition of overall attention abilities, the imitation and the acquisition of new knowledge and cognitive abilities which are fundamental in language development [6].

Children with ASD show differences with neurotypical children (NT) when attending to socially relevant areas of the face. As a consequence, they do not analyse gestures and social information from others, or social situations. They appear to perform differently when acquiring basic social knowledge that neurotypical children learn easily [1]. This specific impairment for paying spontaneous attention to that which is socially relevant and to the activities of others is present in these children from the first year of life. Children with ASD show a different pattern of social attention which influences their acquisition of language [7]. In fact, one of the first indications of possible ASD is delayed speech development. Therefore, visual attention could serve as a phenotypical characteristic for the identification and diagnosis of children with high ASD likelihood, before reaching one year of age [8].

Recent research has suggested that eye movements and the reactions to verbal/visual stimuli used in eye-tracking methodology could be used as signs or biomarkers of early diagnosis of ASD [9–11]. Eye-tracking is a non-invasive and relatively economical methodology that might potentially be used to detect early biomarkers of autism in children of very early ages (even less than 12 months) [12]. However, only when these markers are reliably established and, consequently, early intervention is initiated, will this be translated into a better quality of life for the parents and ASD children. The use of the eye-tracking methodology for the diagnosis of ASD is widely documented [8] even though this methodology is not consistently used to diagnose ASD in the clinical context.

The eye-tracking methodology allows measurement and objectivization of which zones a person directs their attention to during a certain task (gaze following). Studies carried out with this methodology have found that children and adolescents with ASD, in comparison with those who are neurotypical, spend a lot less time paying attention to those areas relevant to social communication, such as the eyes and the mouth [13–15]. Furthermore, it has been observed that eye movements in children with a high likelihood of autism, between six and nine months, show significantly lower gaze fixation in comparison with the neurotypical group. Babies that carry out shorter gaze fixations were afterward diagnosed with autism at 36 months of age [16].

Also, it has been observed that two-year-olds diagnosed with ASD show a greater preference for fixing their attention on geometric figures than on human faces [5]. Equally, significant differences have been found in children with ASD, with respect to neurotypical children, in changes in gaze during a word processing task, given that the former do not move their gaze towards an object when they hear the word [17].

Apart from a decrease in gaze following, there exists a different regulation in the autonomic nervous system (ANS) in children with ASD, which may also be contributing to the differences that they show in social processing. Children with ASD seem to have a higher level of activation of the ANS and show an attentional preference for objects rather than people. According to Porges [18], the ANS plays a central role in communication, a key domain that requires substantial support in the ASD phenotype. From a cognitive point of view, an appropriate level in the situation of arousal facilitates the processing of social information. In this sense, it has been observed that an appropriate autonomic state could be associated with social abilities, but when arousal increases, social behaviour is compromised [19,20]. A reliable measure for studying this ANS regulation would be pupil dilation, given that babies are capable of controlling eye movements from four months of age [21]. Pupillometry has been found to be an adequate measure for testing ANS in paediatric and clinical populations, such as individuals with ASD, because it is less invasive and easy to perform [11].

Anderson et al. [22] studied pupil response to images of faces and non-faces in children with ASD and found that these showed pupil constriction as a response to images of children's faces. However, neurotypical children showed pupil dilation in response to the same stimuli. Years later, the same authors [23], using the same methodology, included a baseline measure, and observed that the group of children with ASD showed greater pupil dilation at that moment in comparison to the neurotypical group. These results are in line with the theory of the existence of a high level of arousal in children with ASD [11] and it has been speculated that acetylcholine, a neurotransmitter in the ANS, is dysregulated in people with ASD [24]. In this same direction, Martineau et al. [25] observed different behaviour patterns in a group of children with ASD compared to neurotypical children on visualising slides of faces, avatars and objects. While the neurotypical group had a significant decrease in pupil size when they had already been shown the stimuli, in the children with ASD high pupil dilation was observed during the entire experimental situation. This seems to indicate atypical functioning of ANS. Other studies also support this hypothesis [12,16,26,27]; a high level of arousal could be atypical, giving rise to more invariant patterns of gaze and visual movements. Furthermore, this variable seems to be related to frequent sleep disruptions which are suffered by children with ASD [28].

This study aims to test the use of eye-tracking methodology as a measure for early detection of ASD in a communicative interaction task. Currently, as far as we know, measures do not exist which allow objectifying and making an early diagnosis of these neurodevelopmental disorders from a linguistic processing task in Spanish. The objective may be considered of great relevance due to the prevalence estimates of ASD in Spain, which is similar to the international rate of 1% [29]. So far, these measures have not been studied in a linguistic interaction task with toddlers. In a study with adults [30], it was found that when neurotypical subjects hear a new word, their pupil dilates significantly compared to the baseline. Therefore, children with typical development, interested in learning language and with a balanced arousal level, are expected to show greater pupillary dilation after listening to a pseudoword compared to children with ASD.

Therefore, the first aim will be to compare the gaze following of children with ASD and neurotypical children (NT) when they hear a pseudoword emitted by a human face using an eye-tracking methodology. This aim is about confirming that children with NT will fix their gaze on the human face a greater number of times, specifically on the eyes, except on hearing the pseudoword, when visual attention will be fixed on the mouth. On the other hand, patterns of visual attention in children with a high likelihood of autism will be more inconsistent, fixing their visual attention a greater number of times on the object and ignoring or paying very little attention to the human face.

A second objective will be to compare the size of pupil dilatation in both groups when they hear the pseudoword. The hypothesis, supported by other studies previously cited, is that the pupil dilation in children with NT will increase when the pseudoword is presented. This is thought to be due to the fact that attention, and cognitive activity in general, increases when having to process a linguistic element, especially if it is unknown. In contrast, in the case of children at risk of having ASD, the greatest dilation will be seen outside the linguistically relevant, because they tend to show preferential gazing towards non-social information as opposed to social information. It has been argued that pupilar dilations in response to cognitive tasks depend on attentional control and they seem to be independent of those produced by emotional arousal [31,32]. It has been assumed that this also happens also when processing the word [30].

2. Materials and Methods

2.1. Participants

The sample was made up of 20 Spanish toddlers. Of these, 10 participants were identified with a high likelihood of ASD (1 girl and 9 boys), with an age range of between 17 and 24 months (M = 21 and SD = 2.357) and 10 NT individuals (3 girls and 7 boys), with an age range of between 17 and 24 months (M = 20 and SD = 1.944). All of the participants were attending Preschool in Asturias, Spain.

The children with ASD were referred by the Early Attention Unit service, in the location where the referral was made to this specific Unit for the treatment of autism "ADANSI" (Association of people with autism "Silent Children"). The criteria for selection of the children with a high likelihood of autism was: children aged between 17 and 24 months with diagnostic reports of autism from the Neuropediatric Service, and in accordance with the following criteria: significant language delay, scarce visual contact, lack of response when called by name, without hearing or vision problems, low communicative intention and scarcity or lack of capacity to imitate. Furthermore, a protocol of previous evaluation was applied to the entire sample to confirm inclusion in the ASD group. This consisted of three tests: the Revised M-CHAT questionnaire (M-CHAT-R/F) for the detection of autism in small children with a follow-up interview [33], the Brunet–Lezine Scale (PY.BL.R) of psychomotor development in early infancy [34], and the Autism Diagnostic Observation Schedule-2 (ADOS-2)—Toddler Module and Module 1 (Spanish version) [35]. The M-CHAT was answered by the caregivers at home before the interview, while ADOS-2 was carried out by one of the authors who has wide experience with this scale. The entire sample of ASD children met the established diagnostic criteria.

The neurotypical sample was taken from the first year of Preschool at an Early Education School in the same location. Inclusion criteria for the comparison group were to score at least below 10 in the ADOS-2 schedule and below 2 in the M-CHAT questionnaire in the absence of any neurological, social, intellectual, sensorial or motor disorder as well as having no first-degree relatives with a previous ASD diagnosis. At the beginning of the school year, the centre sent an information letter to all the families of the children in the course for children of 2–3 years of age. All parents or legal guardians gave their consent to participation in the study.

Table 1 shows the scores of the ASD group and the NT group on the three scales that make up the evaluation protocol for the confirmation of the diagnosis. Both the chronological age of the participants and the global development age on the Brunette–Lezine Scale is in months and all the participants a high possibility of ASD show a global age under that for their chronological age. Furthermore, these show a score for the diagnosis of autism of between 7 and 10 (CSS) in ADOS-2, which indicates a high likelihood of ASD, while the NT participants show a score of between 0 and 4 (CSS). Finally, in the M-CHAT questionnaire, the total scores of the ASD group range between 8–20, which indicates a high possibility of ASD. This ranges from 0–1 in the NT group.

Table 1. Scores of participants with autism spectrum disorder (ASD) and neurotypical children (NT) in the three diagnostic tests.

Group	Participant	Brunette–Lezine		ADOS (CSS)	M-CHAT
		Chronological Age	Global Age of Development		
ASD	1	17	10	22 (10)	15
	2	21	18	14 (7)	8
	3	20	12	21 (10)	10
	4	18	9	22 (10)	12
	5	22	18	22 (10)	15
	6	24	18	20 (10)	20
	7	24	16	22 (10)	8
	8	20	18	15 (9)	9
	9	21	12	21 (10)	10
	10	23	12	20 (9)	11
NT	1	18	18	2 (2)	0
	2	20	21	0 (0)	0
	3	20	20	6 (3)	0
	4	17	18	9 (4)	1
	5	22	24	8 (2)	0
	6	24	24	2 (2)	0
	7	20	21	0 (0)	0
	8	20	20	0 (0)	0
	9	19	18	4 (2)	1
	10	20	20	6 (2)	0

The research design was approved by the Ethics Committee for Research of the University of Oviedo. The study was developed in accordance with the code of ethics of the World Medical Association (Declaration of Helsinki) for experiments involving human subjects in research and the Spanish Law for Personal Data Protection (15/1999 and 3/2018) principles.

2.2. Procedure

Gaze following (fixing and duration of gaze) and pupil dilation of the participants, measured through the pupil diameter were registered during the task. This emulated a communicative situation of acquisition of language in which the emission of words by a human face in association with objects was observed. In a video projected on a screen, a real face was presented which said a pseudoword at the same time as a drawing of a pseudo-object (non-existent invented object) appeared.

The task consisted of nine trials, where the first two were for training purposes. Each of these consisted of a video that began with a blue screen, a neutral colour that does not influence the child's pupil dilation, and a fixation point to direct the child's attention to the centre of the screen. This point, which was maintained for two seconds, corresponds to the baseline of the task. Next, a pseudo-object appeared and emitted an attention-getting sound while remaining in the centre of the screen. When the object remained still, a female face appeared which asked the question: "*What is that*?" with happy and surprised intonation. The face was the only visible part of the body. Immediately following this, the face said the name of the pseudo-object (a pseudoword) with adult-directed natural speech. After hearing the pseudoword, the image of the pseudo-object remained on the screen for two seconds. This was supposed to be the fading and processing time of the pseudoword. After this, the drawing of the pseudo-object disappeared and only the face remained, saying "*It's gone! And what is it called*?" The face was maintained for another two seconds.

To record information, an eye-tracker apparatus, Tobii Spectrum 600 Hz, was used. The participants sat in the laps of their parents in front of a 16" monitor with a panoramic aspect ratio of 16.9 in a dark soundproof room. Their central vision was lined up with the centre of the monitor, at a distance of 60 cm between the eye and the monitor. Once the participant was in place, a calibration of 5 points was carried out through colourful and attractive cartoons. This way, the luminosity was controlled to ensure that changes in pupil dilatation were due to the task itself and not due to changes in the light. To do this, a photometer MASTECH MS6612 was used, with the criterion that luminosity did not pass 110 lumens.

A group of nine pseudowords was selected from a list of test items MEMOFON [36]. Of these, two were for training purposes (*muz and norba*). The pseudowords were differentiated by their complexity both in number and in the type of component syllables. Therefore, two monosyllabic pseudowords were selected, one phonologically simpler with a consonant + vowel + consonant pattern (CVC) (*sel*) and another more complex one with a closed syllable (*tron*). Two pseudowords with two syllables were selected, one more simple (*sina*) and another more complex, since it contains an inverse syllable (*pamul*). Another three pseudowords of three syllables were also selected, two easier (*bésica* and *gapata*) and another more phonologically complex one (*calcemar*). Each pseudoword was presented in association with a drawing of the pseudo-object, an invented object. The pseudo-objects were designed specifically for the experiment and were randomly associated with the pseudowords. In Figure 1, an example of a pseudoword associated with a pseudo-object can be seen.

Once the pseudowords were associated with the pseudo-objects, the order in which the task stimuli were shown to the different participants was random. Figure 2 represents the sequence of one of the trials. The first moment of the sequence corresponded to the baseline (BL) register; the second and third ones corresponded to the moment of the presentation of the pseudoword (PW); the fourth moment was the period of time in the fading of the pseudoword (FPW) and the last sequence of the video was when the pseudo-object first, and later the human face, disappeared from the screen (PO).

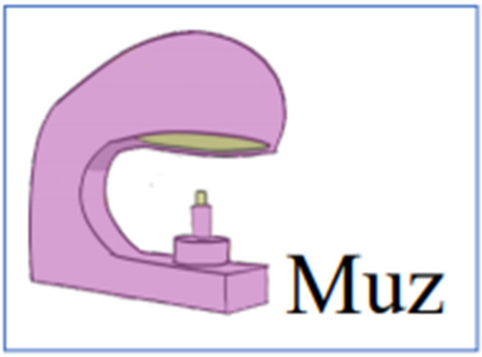


Figure 1. Pseudoword associated with its pseudo-object.

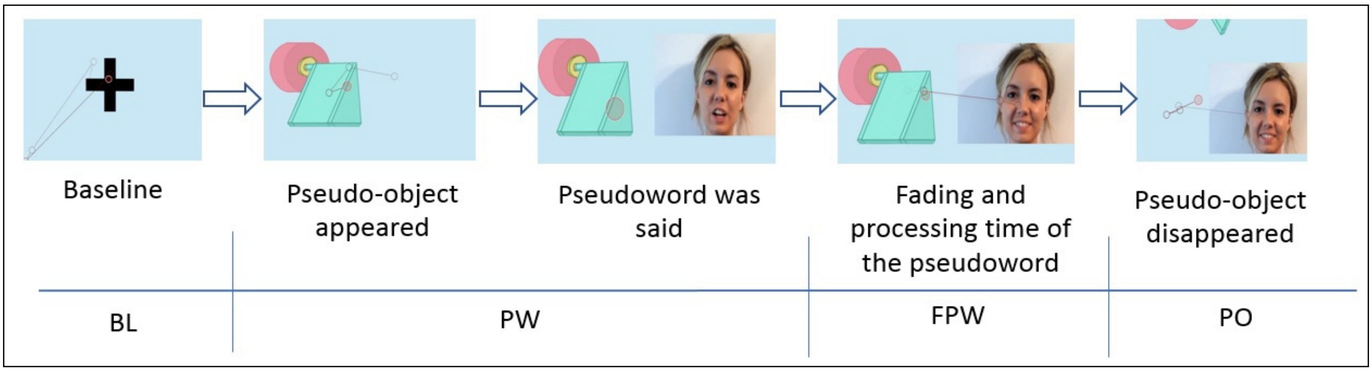


Figure 2. Sequence of a trial.

The appearance time of the pseudo-objects and the waiting time between the continuation of the image of the object and the production of the pseudoword was determined based on a pilot study of NT toddlers aged between 18 and 30 months. Here, the same pseudo-objects and pseudowords were used. The times obtained were taken as reference criteria since, in the scientific literature, times are not clearly established for a linguistic processing task using eye-tracking methodology at such an early age.

During the entire task, pupil dilation was registered and measured in millimetres every two milliseconds, as well as gaze and the areas of interest on which gaze was fixed: the pseudo-object (AOI 1), the eyes of the face (AOI 2) and/or the area of the mouth pertaining to the face (AOI 3). Data were obtained through the system's software "Tobii Pro Lab" and included the number and time of gazes at the previously defined areas of interest (AOI), and pupil dilation during the whole task.

2.3. Data Analysis

The data obtained were analysed with the programme IBM SPSS Statistics—version 22.0 for Windows. The indices for asymmetry and kurtosis were carried out and a descriptive analysis of the dependent variables (gaze following and pupil measurement), as well as of the variable of classification by chronological age (CA), were carried out. Due to the size of the groups in the study and to the violation of normality and homogeneity variance assumptions, the data were analysed using nonparametric statistics.

In order to confirm whether differences existed between both groups in the gaze following measurement, pairwise Mann–Whitney U tests were used for between-group comparisons. Cliff's Delta (δ) statistic was chosen as the effect size estimator because it is more appropriate when the homogeneity of variance or normality assumptions are violated. Based on Cohen norms [37], we consider an effect size of 0.2 as a small effect, 0.5 as a medium effect, and 0.8 and upwards as a large effect. These analyses were carried out based on the total number of gaze fixations and the total time of fixations of the participants when looking at the object, the mouth and the eyes. Friedman tests were also carried out to establish within-group differences between AO, and Wilcoxon signed-ranks tests were used for pairwise post hoc comparisons. Bonferroni corrections were applied to adjust the *p*-values for multiple post hoc comparisons.

With regard to the analysis of pupil dilation data, a model used by López-Ornat et al. [30] was followed, thus establishing four periods of measurement. In the first period, during the 400 ms that preceded the start of the trial, the baseline (BL) of pupil dilation of each participant was set, this being a measurement of pupil diameter. At the same time, the point of gaze fixation was determined [38]. After that, pupil dilation was considered during the time of presentation of the pseudoword (PW). This period was of variable duration due to the different lengths of the words, between 650 ms and 1200 ms. The third period corresponded to the following 2000 ms where measurements were registered of the period following the fading of the pseudoword (FPW). The fourth period corresponded to that section of the video when the pseudo-object (PO) and the human face disappeared from the screen, with a duration of two seconds before the next trial began. Only the period of presentation of the pseudoword was of variable duration depending on the length of the pseudoword. Each of these periods included a set of observations taken every two ms. This structure was used for each of the nine trials.

In order to test if there were differences between groups in pupillary diameter throughout the task and in the different periods, Mann–Whitney *U* tests were used. Cliff's Delta (δ) statistic was used to estimate the effect size. Friedman tests were also carried out to establish within-group differences between periods, and Wilcoxon signed-ranks tests were used for pairwise post hoc comparisons. Bonferroni corrections were applied to adjust the *p*-values for multiple post hoc comparisons.

3. Results

No statistically significant differences between the two groups were found with regard to chronological age ($U = 34.50$; $Z = -1.201$; $p = 0.23$; Cliff's $\delta = 0.443$). As expected, there were statistically significant differences with regard to developmental age ($U = 6.00$; $Z = -3.412$; $p < 0.001$; Cliff's $\delta = -0.88$).

The skewness index (A) in the number of object fixations (NOF), mouth fixations (NMF), eye fixations (NEF) and in the eye fixation time (EFT) indicated that the distribution data were asymmetrical in the NT group (A = 1.617; 1.138; −2.207; and −3.180, respectively) as well as NMF (A = 1.548) in the ASD group. On the other hand, the kurtosis index (K) indicated non-normal distribution in the NOF (K = 3.515) and the eye fixation time (EFT) (K = 5.362) in the NT group.

Regarding differences between groups, there were statistically significant differences between the ASD group and the NT group in the total number of gaze fixations ($U = 20.00$; $Z = -2.27$; $p = 0.023$; Cliff's $\delta = -0.60$) and the total time of fixations ($U = 0.00$; $Z = -3.78$; $p < 0.001$; Cliff's $\delta = -0.60$).

As shown in Table 2, the performance of the groups in gaze following showed significant differences. The NT participants looked at the eyes a greater number of times ($U = 0.00$; $Z = -3.79$; $p < 0.001$; Cliff's $\delta = -1$) and for a longer time than the ASD participants ($U = 0.00$; $Z = -3.78$; $p < 0.001$; Cliff's $\delta = -1$). On the contrary, the ASD group looked at the mouth a greater number of times than the NT group ($U = 0.00$; $Z = -3.80$; $p < 0.001$; Cliff's $\delta = 1$) and for a longer time ($U = 0.00$; $Z = -3.78$; $p < 0.001$; Cliff's $\delta = 1$). Similarly, the ASD participants looked at the pseudo-object a greater number of times ($U = 0.00$; $Z = -3.78$; $p < 0.001$; Cliff's $\delta = 1$) and for a longer time than the NT participants ($U = 0.00$; $Z = -3.78$; $p < 0.001$; Cliff's $\delta = 1$).

Table 2. Median (MDN), Interquartile range (IQR), differences between groups, the *p* values, and effect size in number and time of gaze fixation on the areas of interest between ASD and NT.

		ASD		NT				
		MDN	**IQR**	**MDN**	**IQR**	***Z***	***p***	**Cliff's δ**
Eyes	NEF	34.50	16.00	210.50	10.00	−3.79	0.000	−1
	EFT	16.00	19.71	157.73	13.28	−3.78	0.000	−1
Mouth	NMF	41.00	45.00	3.00	4.00	−3.80	0.000	1
	MFT	19.17	36.76	1.44	4.17	−3.78	0.000	1
Object	NOF	129.50	45.00	35.00	13.00	−3.78	0.000	1
	OFT	53.54	14.23	26.37	4.51	−3.78	0.000	1

NEF = Number of eye fixations, EFT = Eyes fixation time, NMF = Number of mouth fixations, MFT = Mouth fixation time, NOF = Number of object fixations, OFT = Object fixation time.

Within-group differences showed statistically significant differences in the number of fixations in the different areas in the NT group ($\chi^2 = 20$; $p < 0.001$) and also in the ASD group ($\chi^2 = 15$; $p = 0.001$), but not in the same way. Pairwise post hoc comparisons showed that NT participants made more fixations on the eyes rather than on the object (Z = 2.80; $p = 0.005$; Cliff's δ = −1), on the eyes rather than on the mouth (Z = 2.81; $p = 0.005$; Cliff's δ = −1), and on the object rather than the mouth (Z = 2.80; $p = 0.005$; Cliff's δ = 1). Toddlers with a possible autism diagnosis made more fixations on the object rather than the eyes (Z = 2.80; $p = 0.005$; Cliff's δ = 1) and on the object rather than the mouth (Z = 2.80; $p = 0.005$; Cliff's δ = 1). However, there were no significant differences between the number of fixations on the mouth as compared to fixations on the eyes in this group (Z = −0.97; $p = 0.330$; Cliff's δ = 0.34).

Within-group comparisons also showed statistically significant differences in the time of gaze fixation in the same direction in both groups, NT group ($\chi^2 = 20$; $p < 0.001$) and ASD group ($\chi^2 = 12.20$; $p = 0.002$). Pairwise post hoc comparisons showed NT toddlers spent more time looking at the eyes rather than the object (Z = 2.80; $p = 0.005$; Cliff's δ = −1), at the eyes rather than the mouth (Z = 2.80; $p = 0.005$; Cliff's δ = −1), and at the object rather than the mouth (Z = 2.80; $p = 0.005$; Cliff's δ = 1). Toddlers with a possible autism diagnosis spent more time looking at the object rather than the eyes (Z = 2.80; $p = 0.005$; Cliff's δ = 1) and at the object rather than the mouth (Z = 2.5; $p = 0.013$; Cliff's δ = 0.610). However, there were no significant differences between the time of gaze fixation on the mouth rather than on the eyes in this group (Z = −0.87; $p = 0.386$; Cliff's δ = 0.30).

For the calculation of pupil dilation, firstly, a prior pruning of the data was carried out in order to exclude missing data or blinking. Afterward, the mean of pupil dilation was calculated in sections (BL, PW, FPW and PO) for each pseudoword and the total mean of the group of pseudowords was calculated for each section (BL, PW, FPW and PO). Thus, the mean for dilation for the two groups in each section was obtained (Table 3). It can be seen how this was higher in children with possible ASD. In particular, the greater mean for this group was given in section PO, which corresponds to the moment in which the pseudo-object disappears. With respect to the NT group, greater dilation was observed in section PW, the moment when the pseudoword was heard. However, no statistically significant differences between the mean of both groups for pupil dilation ($U = 32.00$; Z = −1.36; $p = 0.173$; δ = 0.20) were observed.

Table 3. Total mean of global pupil dilation for each of the periods (ms).

Group	Total Mean	BL	PW	FPW	PO
ASD	3.702	3.685	3.711	3.659	3.753
NT	3.583	3.525	3.661	3.578	3.567

BL = Baseline; Number, PW = Time of presentation of the pseudoword, FPW = Period following the fading of the pseudoword, PO = Pseudo-object and the human face disappeared.

Nevertheless, the within-group comparisons in pupil dilation by periods showed statistically significant differences in the NT group ($\chi^2 = 10.80$; $p = 0.013$), although there were no differences in the ASD group ($\chi^2 = 6.96$; $p = 0.073$). Pairwise post hoc comparisons between the four periods in the NT group revealed that pupil dilation was larger during the time of presentation of the pseudoword as compared to that of the pseudo-object and the disappearance of the human face ($Z = -2.84$; $p = 0.004$; Cliff's $\delta = 0.38$). Differences in pupil dilation were also close to significance between the time of presentation of the pseudoword as compared with the baseline ($Z = -2.30$; $p = 0.021$; Cliff's $\delta = 0.16$), and the time of presentation of the pseudoword as compared with the period following the fading of the pseudoword ($Z = -2.32$; $p = 0.020$; Cliff's $\delta = 0.16$), after having been applied Bonferroni correction to adjust the p-values for multiple post hoc comparisons ($p = 0.17$). In both cases, pupil dilatation was greater at the time of presentation of the pseudoword.

Figure 3 shows the mean (in millimetres (mm), y-axis) of pupil dilation across groups by observation periods during the entire task. The period between BL and PW, marked with dots in Figure 3, has not been analysed because it was not part of the object of study. The x-axis shows the trial sequence through observations registered every 20 ms. It can be seen that the ASD participants showed an activation level above that of the NT participants during all periods, although as previously seen, the difference was non-statistically significant. However, the similarity in the shape of the curve indicates that, although there was a slightly higher level of activation, the ASD participants behaved in a similar way to the NT participants. Both groups presented a lower activation level during the baseline (BL) register, at the moment preceding the beginning of the task. Activation increased notably at the beginning of the presentation of the pseudowords (PW) and continued to increase during the task. However, differences may be observed, first, at the end of the time of presentation of the pseudowords, and second, at the maximum peaks of the curve. When the longest pseudowords end, the NT participants showed a drop in their activation that seemed to become stable, while in the ASD group, it continued to increase. On the other hand, the maximum peak of the NT participants was produced when they were processing the pseudoword and the waiting time was going to commence (FPW). In contrast, in the ASD group this was produced almost at the same moment (PW) but also, again, at the moment of the object's disappearance (PO). Finally, the period of fading of the pseudoword (FPW) lowered the activation of both groups. Then, when the object disappeared and the face said the pseudoword again (PO), activation once again increased in both groups but only in the ASD group did it again reach the peak of maximum activation.

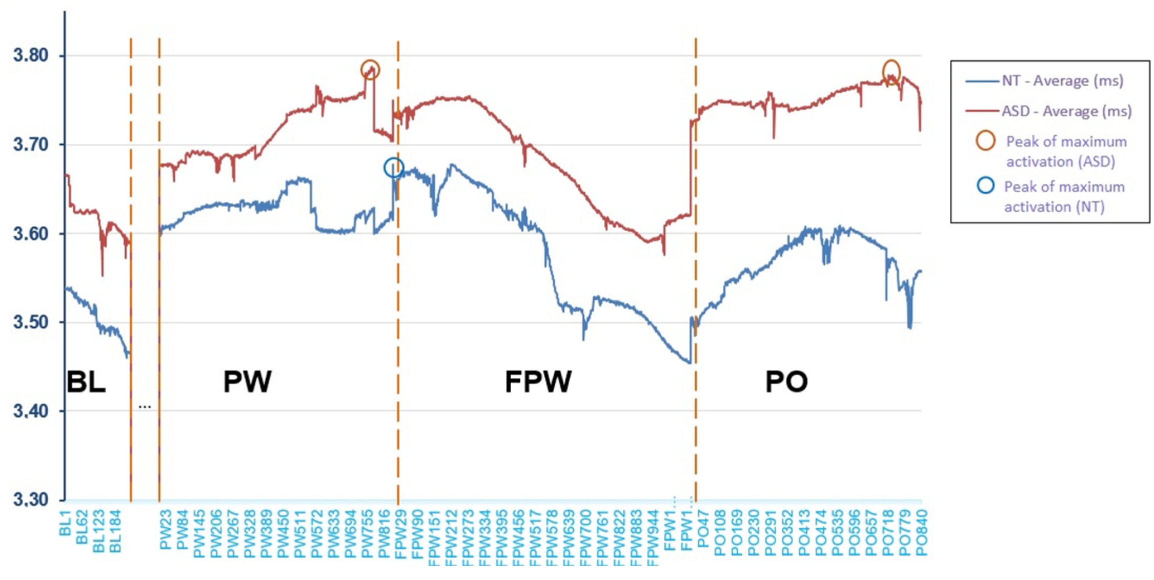


Figure 3. Mean of pupillary dilation throughout the different periods in the entire task (mm).

4. Discussion

Gaze following and pupil dilation in a linguistic processing task were tested as possible biomarkers for early diagnoses of autism spectrum disorder. Regarding gaze following, the present study has objectively corroborated, using the eye-tracking methodology, that while NT toddlers displayed more numerous and longer fixations on the eye regions than children with a high possibility of developing autism, these displayed more and longer fixations on objects and the mouth regions than NT toddlers. Additionally, ASD toddlers look more and for a longer time at objects rather than eyes and mouth. However, NT toddlers fix their attention on the eyes rather than objects and they pay little attention to the mouth. Thus, while in communicative interaction, NT children spend most of the time looking at the eyes, the children with possible ASD show preference for objects, which could translate into difficulties when integrating social information [5,15]. Therefore, the results obtained are in line with the conclusions already made in other studies [12,15–17], suggesting the presence of a non-typical control of attention in ASD, reduced general attention to the eyes and greater attention to non-social elements.

This different use of gaze following as a clue to social reference for learning words could interfere in the acquisition of language, given that eye contact is essential for labelling a referent with a certain word [39]. This different pattern could be attributed to the fusiform gyrus hypoactivation in ASD [13], and it could have later consequences, paying more attention to phonological information than semantic information and social cues [40] and failing to form more robust lexical representations of words [17]. Additionally, it was observed that this atypical pattern in the initiation of joint attention and gaze alternation in ASD, could make the caregiver respond less to a child who does not initiate joint attention [39].

With the results found here, it may be claimed that the measurement of the visual following and attentional preference could be sensitive when differentiating an ASD gaze pattern with a neurotypical gaze pattern. Thus, gaze following measurements related to social attention are good candidates for use as early biomarkers. This may allow us to objectively establish a suspicion or a high likelihood of autism at an early age. This could be quite useful because detection or diagnoses of autism by the Early Attention Unit service are now carried out too late (usually at 30 months) and perhaps it could contribute to distinguishing children with ASD from late talkers or those misdiagnosed with maturational delays [41,42]. With these eye-tracking measurements the diagnoses would no longer depend on subjective clinical judgments, but rather, would provide us with an objective and reliable measure to make solid autism diagnoses.

Regarding the measurement of pupil dilation, the results are not so clear. The ASD group shows a slightly larger average pupil size throughout the task than the NT group. This could suggest that children with ASD show hyper-arousal in the tasks which they must face, which would be in accordance with previous research [26–28,43]. A rising level of activation during the task would translate into attention level difficulties that could form the basis of the characteristics that these children show when processing social information in different contexts. Nevertheless, this difference was not large enough to reach a significance level in the complete task. So, no conclusion can be drawn. Other researchers also found no differences in arousal between children with ASD and NT children [44]. This could be attributed to the early age of the participants [27,44–46]. Dinalankara et al. [45] observed that the baseline pupil size increased with age, up to four years in NT children, but this pattern was not observed in children with ASD. However, from the age of four, children with ASD had a larger mean baseline pupil size than NT children. These changes with age appear to be due to the increased acceleration of white matter maturation in ASD [45]. Another possible explanation for our results could be the level of possible autism in our participants because it was observed that toddlers with a high risk of ASD presented larger base pupil size in resting than toddlers with a low possibility of ASD [46].

However, an interesting issue arises from within-group pupillary dilation results. There were no significant differences in activation measured through pupillary dilation between periods in the ASD toddlers. Nevertheless, in the NT toddlers, a higher level of activation took place during the hearing (and processing) of the pseudoword, compared to the other periods. This suggests a higher level of active attention in this period. This may confirm the previously formulated hypothesis that pupil dilation in children with NT will increase when the pseudoword is presented, because they are attending to language. While in children with ASD, no larger dilation will be seen in the linguistically relevant, because they would show low selective attention to relevant information for communication. Indeed, in this group higher activation is observed in the final period. It was observed that cognitively relevant pupil dilations are caused by the inhibition of the parasympathetic nervous system and by acetylcholine, which plays an important role in the regulation of attention control [24].

In addition, the NT group's maximum peak of dilation was found at the end of the processing of the pseudoword, indicating that these children are paying attention and retaining the phonological representation of the pseudoword [30] in working phonological memory, and that they are making a greater cognitive effort at this point. They are ready to learn language and to concentrate their interest on this. In the children with possible ASD, the maximum average value is produced during the disappearance of the pseudo-object (PO), and the maximum peak of dilation occurs at the end of the pseudoword presentation (PW) as in the NT group. However, this also occurs in the period during which the pseudo-object disappears. Two maximum peaks were considered, since the variation between both of them is practically null. These results are not in accordance with what Anderson and Colombo set out in their study [23], since these authors found the maximum peak in the baseline section.

To sum up, in the present study, the eye-tracking methodology was used in an innovative way in a linguistic processing task in children of an early age. It was shown that these types of tests could provide evidence when measuring attention bases in the development of the process of acquisition of language in children, not only after 24 months of age [9] but also before that age. In addition, in comparing NT toddlers with possible ASD toddlers, differences are observed in the development of the pattern of gaze during the acquisition of linguistic abilities which appear to have great diagnostic potential.

These findings should be interpreted from a neuropsychological perspective, since alterations in visual attention are indicative of a state of anomalous neural activation. The results found indicate that indirect, objective measurements of the level of activation, such as number and time of gaze fixations (registered through eye-tracking) are potential candidate biomarkers for diagnostic indicators of the presence of ASD [43]. Even so, it would be necessary to carry out a larger future study of these measurements to refine this technique for non-invasive diagnostic screening. It is easy to administer and economical for the detection of anomalous gaze patterns in children who could have an autism spectrum disorder.

Regarding pupil dilation measurement results, these are not conclusive. Its use as a biomarker diagnostic indicator to identify children with a high likelihood of autism at an early age is not clear. However, it appears to be a hopeful candidate for investigating the differential processing of new words in NT toddlers and possible ASD toddlers. In any case, further research is required and the number and type of stimuli to be processed must be increased.

Finally, despite the encouraging results obtained, some limitations must be mentioned. First, the size of the sample, since these results could not be extrapolated to the entire autism population. Secondly, it is clear that the applied measurements do not allow the establishment of a definitive diagnosis of ASD and, at the moment, the participants are not being longitudinally monitored to ascertain a final diagnostic outcome. Finally, it should be pointed out that the stimuli in the linguistic interaction task were presented in a video and not in a live social situation.

Author Contributions: R.C., V.M. and C.G. conceived the idea. R.C., V.M. and C.G. conducted the investigation. R.C. and V.M. carried out data curation. R.C. and C.G. analysed data. R.C. and V.M. wrote the first draft. All authors have read and agreed to the published version of the manuscript.

Funding: This research did not receive any specific grant from funding agencies. Nevertheless, the work has been supported by Early Attention Unit service from Gijón (Spain), who provides the Tobii Spectrum System and the locations where the research was carried out. The ASD Association "ADANSI" (Association of people with autism 'Silent Children' of the Principality of Asturias (Spain)) facilitated the recruitment of participants with ASD who did not receive any financial compensation for their participation but were they informed of the individual results obtained. The publication fees are partially funded by the annual aid provided by the Complutense University of Madrid to the Research Group UCM EQUIAL 940733 "Equipo de Investigación sobre Adquisición del Lenguaje".

Institutional Review Board Statement: The research design was approved by the Ethics Committee for Research of the University of Oviedo on 20 February 2019. The study was developed in accordance with the code of ethics of the World Medical Association (Declaration of Helsinki) for experiments involving human subjects in research and the Spanish Law for Personal Data Protection (15/1999 and 3/2018) principles.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to ethical principles.

Acknowledgments: We appreciate the families who participated in this study. We also acknowledge the Children with ASD Association "ADANSI" of the Principality of Asturias, for their collaboration and involvement. We would also like to thank Guillermo Gallego for the design and drawing of the visual pseudo-objects used in the experiment.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Karmiloff, K.; Karmiloff-Smith, A. *Pathways to Language. From Fetus to Adolescent*; Harvard University Press: Cambridge, MA, USA, 2001. [CrossRef]
2. Lewkowicz, D.J.; Hansen-Tift, A.M. Infants deploy selective attention to the mouth of a talking face when learning speech. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 1431–1436. [CrossRef]
3. Tenenbaum, E.; Sobel, D.; Sheinkopf, S.; Malle, B.; Morgan, J. Attention to the mouth and gaze following in infancy predict language development. *J. Child Lang.* **2015**, *42*, 1173–1190. [CrossRef]
4. Young, G.S.; Merin, N.; Rogers, S.J.; Ozonoff, S. Gaze behavior and affect at 6 months: Predicting clinical outcomes and language development in typically developing infants and infants at risk for autism. *Dev. Sci.* **2009**, *12*, 798–814. [CrossRef]
5. Shic, F.; Macari, S.; Chawarsk, K. Speech disturbs face scanning in 6-month-olds who develop autism spectrum disorder. *Biol. Psychiatry* **2014**, *75*, 231–237. [CrossRef]
6. D'Souza, D.; D'Souza, H.; Johnson, M.H.; Karmiloff-Smith, A. Concurrent relations between face scanning and language: A cross syndrome infant study. *PLoS ONE* **2015**, *10*, e0139319. [CrossRef]
7. Howard, P.L.; Zhang, L.; Benson, V. What can eye movements tell us about subtle cognitive processing differences in autism? *Vision* **2019**, *3*, 22. [CrossRef]
8. Chawarska, K.; Macari, S.; Shic, F. Decreased spontaneous attention to social scenes in 6-month-old infants later diagnosed with autism spectrum disorders. *Biol. Psychiatry* **2013**, *74*, 195–203. [CrossRef]
9. Murias, M.; Mayor, S.; Davlantis, K.; Franz, L.; Harris, A.; Rardin, B.; Sabatos-DeVito, M.; Dawson, G. Validation of eye—Tracking measures of social attention as a potential biomarker for autism clinical trials. *Autism Res.* **2018**, *11*, 166–174. [CrossRef]
10. Frazier, T.W.; Klingemier, E.W.; Parikh, S.; Speer, L.; Strauss, M.S.; Eng, C.; Hardan, A.Y.; Youngstrom, E.A. Development and validation of objective and quantitative eye tracking-based measures of autism risk and symptom levels. *J. Am. Acad. Child Adolesc. Psychiatry* **2018**, *57*, 858–866. [CrossRef]
11. De Vries, L.; Fouquaet, I.; Boets, B.; Naulaers, G.; Steyaert, J. Autism spectrum disorder and pupillometry: A systematic review and meta-analysis. *Neurosci. Biobehav. Rev.* **2020**. [CrossRef]
12. Nyström, P.; Gliga, T.; Jobs, E.N.; Gredebäck, G.; Charman, T.; Johnson, M.H.; Bölte, S.; Falck-Ytter, T. Enhanced pupillary light reflex in infancy is associated with autism diagnosis in toddlerhood. *Nat. Commun.* **2018**, *9*, 1678. [CrossRef] [PubMed]
13. Dalton, K.M.; Nacewicz, B.M.; Johnstone, T.; Schaefer, H.S.; Gernsbacher, M.A.; Goldsmith, H.H.; Alexander, A.L.; Davidson, R.J. Gaze fixation and the neural circuitry of face processing in autism. *Nat. Neurosci.* **2005**, *8*, 519–526. [CrossRef] [PubMed]
14. Klin, A.; Jones, W.; Schultz, R.T.; Volkmar, F.R.; Cohen, D.J. Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Arch. Gen. Psychiatry* **2002**, *59*, 809–816. [CrossRef]
15. de Wit, T.C.J.; Falck-Ytter, T.; von Hofsten, C. Young children with Autism Spectrum Disorder look differently at positive versus negative emotional faces. *Res. Autism Spectr. Disord.* **2008**, *2*, 651–659. [CrossRef]

16. Wass, S.V.; Jones, J.H.; Gliga, T.; Smith, T.J.; Charman, T.; Johnson, M.H. Shorter spontaneous fixation durations in infants with later emerging autism. *Sci. Rep.* **2015**, *5*, 1–8. [CrossRef]
17. Chita-Tegmark, M.; Arunachalam, S.; Nelson, C.A.; Tager-Flusberg, H. Eye-tracking measurements of language processing: Developmental differences in children at high risk for ASD. *J. Autism Dev. Disord.* **2015**, *45*, 3327–3338. [CrossRef]
18. Porges, S.W. The Vagus: A mediator of behavioral and physiological features associated with autism. In *The Neurobiology of Autism*, 2nd ed.; Bauman, M.L., Kemper, T.L., Eds.; The Johns Hopkins University Press: Baltimore, MD, USA, 2005; pp. 65–78.
19. Patriquin, M.A.; Hartwig, E.M.; Friedman, B.H.; Porges, S.W.; Scarpa, A. Autonomic response in autism spectrum disorder: Relationship to social and cognitive functioning. *Biol. Psychol.* **2019**, *145*, 185–197. [CrossRef]
20. Benevides, T.W.; Lane, S.J. A review of cardiac autonomic measures: Considerations for examination of physiological response in children with autism spectrum disorder. *J. Autism Dev. Disord.* **2015**, *45*, 560–575. [CrossRef]
21. Verschoor, S.A.; Spapé, M.; Biro, S.; Hommel, B. From outcome prediction to action selection: Developmental change in the role of action-effect bindings. *Dev. Sci.* **2013**, *16*, 801–814. [CrossRef]
22. Anderson, C.J.; Colombo, J.; Jill, S.D. Visual scanning and pupillary responses in young children with Autism Spectrum Disorder. *J. Clin. Exp. Neuropsychol.* **2006**, *28*, 1238–1256. [CrossRef] [PubMed]
23. Anderson, C.J.; Colombo, J. Larger tonic pupil size in young children with autism spectrum disorder. *Dev. Psychobiol.* **2009**, *51*, 207–211. [CrossRef]
24. Hellmer, K.; Nyström, P. Infant acetylcholine, dopamine, and melatonin dysregulation: Neonatal biomarkers and causal factors for ASD and ADHD phenotypes. *Med. Hypotheses.* **2017**, *100*, 64–66. [CrossRef]
25. Martineau, J.; Hernandez, N.; Hiebel, L.; Roché, L.; Metzger, A.; Bonnet-Brilhault, F. Can pupil size and pupil responses during visual scanning contribute to the diagnosis of autism spectrum disorder in children? *J. Psychiatr. Res.* **2011**, *45*, 1077–1082. [CrossRef]
26. Wagner, A.E.; Toffanin, P.; Bas, K.D. The timing and effort of lexical access in natural and degraded speech. *Front. Psychol.* **2016**, *7*, 398. [CrossRef]
27. Segers, M.; Bebko, J.M.; Zapparoli, B.L.; Stevenson, R.A. A pupillometry study of multisensory social and linguistic processing in autism and typical development. *Dev. Psychol.* **2020**, *56*, 2080–2094. [CrossRef]
28. Moore, M.; Evans, V.; Hanvey, G.; Johnson, C. Assessment of sleep in children with Autism Spectrum Disorder. *Children* **2017**, *4*, 72. [CrossRef]
29. Morales-Hidalgo, P.; Roigé-Castellví, J.; Hernández-Martínez, C.; Voltas, N.; Canals, J. Prevalence and characteristics of autism spectrum disorder among Spanish school-age children. *J. Autism Dev. Disord.* **2018**, *48*, 3176–3190. [CrossRef] [PubMed]
30. López-Ornat, S.; Karousou, A.; Gallego, C.; Martín, L.; Camero, R. Pupillary measures of the cognitive effort in auditory novel word processing and short-term retention. *Front. Psychol.* **2018**, *9*, 2248. [CrossRef] [PubMed]
31. Gilzenrat, M.S.; Nieuwenhuis, S.; Jepma, M.; Cohen, J.D. Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function. *Cognit. Affect. Behav. Neurosci.* **2010**, *10*, 252–269. [CrossRef]
32. Karatekin, C.; Couperus, J.; Marcus, D. Attention allocation in the dual-task paradigm as measured through behavioral and psychophysiological responses. *Psychophysiology* **2004**, *41*, 175–185. [CrossRef]
33. Robins, D.L.; Fein, D.; Barton, M. *The Modified Checklist for Autism in Toddlers, Revised, with Follow-Up (M-CHAT-R/F)*; Self-Published: Washingtonville, NY, USA, 2009.
34. Josse, D.; Pereda, S. *Brunet Lézine Revisado: Escala de Desarrollo Psicomotor de la Primera Infancia (Brunet Lézine Reviewed: A Psychomothor Development Scale for Infants)*; Symtec.: Madrid, Spain, 1997.
35. Lord, C.; Rutter, M.; DiLavore, P.C.; Risi, S.; Gotham, K.; Bishop, S.L. *Autism Diagnostic Observation Schedule, Second Edition (ADOS-2) Manual (Part I): Modules 1–4*; Western Psychological Services: Torrance, CA, USA, 2012.
36. Mariscal, S.; Gallego, C. La imitación como herramienta para investigar y evaluar el desarrollo lingüístico temprano: Un estudio piloto de repetición de palabras y pseudopalabras (Imitation as a tool for research and assessment of early language development: A pilot study of word and non-word repetition). *Rev. Investig. Logop.* **2013**, *3*, 53–75.
37. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed.; Lawrence Earlbaum Associates: Hillsdale, NJ, USA, 1988.
38. Guasch, M.; Ferré, P.; Haro, J. Pupil dilation is sensitive to the cognate status of words: Further evidence for non-selectivity in bilingual lexical access. *Bilingualism* **2017**, *20*, 49–54. [CrossRef]
39. Nyström, P.; Thorup, E.; Bölte, S.; Falck-Ytter, T. Joint attention in infancy and the emergence of autism. *Biol. Psychiatry* **2019**, *86*, 631–638. [CrossRef]
40. Norbury, C.F.; Griths, H.; Nation, K. Sound before meaning: Word learning in autistic disorders. *Neuropsychologia* **2010**, *48*, 4012–4019. [CrossRef]
41. Brett, D.; Warnell, F.; McConachie, H.; Parr, J.R. Factors affecting age at ASD diagnosis in UK: No evidence that diagnosis age has decreased between 2004 and 2014. *J. Autism Dev. Disord.* **2016**, *46*, 1974–1984. [CrossRef]
42. Fernald, A.; Marchman, V.A. Individual differences in lexical processing at 18 months predict vocabulary growth in typically developing and late-talking toddlers. *Child Dev.* **2012**, *83*, 203–222. [CrossRef]
43. Del Valle Rubido, M.; Hollander, E.; McCracken, J.T.; Shic, F.; Noeldeke, J.; Boak, L.; Khwaja, O.; Sadikhov, S.; Fontoura, P.; Umbricht, D. Exploring social biomarkers in high-functioning adults with autism and Asperger's versus healthy controls: A cross-sectional analysis. *J. Autism Dev. Disord.* **2020**, *50*, 4412–4430. [CrossRef]

44. Fan, X.; Miles, J.H.; Takahashi, N.; Yao, G. Abnormal transient pupillary light reflex in individuals with autism spectrum disorders. *J. Autism Dev. Disord.* **2009**, *39*, 1499–1508. [CrossRef]
45. Dinalankara, D.M.R.; Miles, J.H.; Takahashi, T.N.; Yao, G. Atypical pupillary light reflex in 2–6-year-old children with autism spectrum disorders. *Autism Res.* **2017**, *10*, 829–838. [CrossRef]
46. Kercher, C.; Azinfar, L.; Dinalankara, D.M.R.; Takahashi, T.N.; Miles, J.H.; Yao, G. A longitudinal study of pupillary light reflex in 6- to 24-month children. *Sci. Rep.* **2020**, *10*, 1205. [CrossRef]

children

MDPI

Article

Language and Communication in Preschool Children with Autism and Other Developmental Disorders

Vicenç Torrens * and Cristina Ruiz

Department of Developmental Psychology, National University of Distance Learning, 28040 Madrid, Spain; cristina.ruiz@email.com

* Correspondence: vtorrens@psi.uned.es; Tel.: +34-913988650

Abstract: In this research, we studied the language and communication skills of preschool children with a diagnosis of autistic syndrome disorder (ASD) (n = 51) compared to children with other developmental disorders (DD) (n = 42), using direct measures and parental reports when assessing the development of language and communication. As a novelty, this research studied a sample of children with low language and communication skills. We found a high correlation between direct measures and parental reports for both populations. Therefore, we propose that combining the information supplied by direct measures together with that supplied by parental reports would be a suitable strategy for language assessment in these populations. In addition, the results show a delay in language comprehension with respect to language production in children with ASD, along with many difficulties with non-verbal communication, compared to children with other developmental disorders (DD). We also found significant differences between both groups with respect to lexical categories. The differences in language and communication profiles of children with ASD compared to children with other DD might have some implications for diagnoses and language intervention in these populations.

Keywords: atypical language acquisition; autism; developmental disorders; assessment

Citation: Torrens, V.; Ruiz, C. Language and Communication in Preschool Children with Autism and Other Developmental Disorders. *Children* **2021**, *8*, 192. https://doi.org/10.3390/children8030192

Academic Editor: Eva Aguilar Mediavilla

Received: 7 January 2021
Accepted: 28 February 2021
Published: 4 March 2021

1. Introduction

Language difficulties are a crucial symptom in defining the Autism Disorder Spectrum (ASD) since children with this disorder show a deficit in the development of social and communicational interaction [1]. Previous studies have identified differences in the pace of general development, language development, and individual differences between children with ASD and children with typical development [2]. In addition, previous studies on children with ASD have found patterns of language development and gestures in language production and comprehension that are atypical [3–5]. The language skills found in older children with ASD are predicted by the early use of gestures [6,7] and early language performance [5,8].

The study of communicative skills in children with ASD presents some difficulties. One such difficulty is the choice of appropriate tools to assess language when children have very limited cognitive and communicative skills. In previous studies on language skills in children with developmental disorders, authors have used direct standardized measures and parental reports. Direct measures might not be suitable when children have very low levels of language comprehension and production [9]. In addition, children who lack motivation or have attentional deficits might show a performance that does not match their real capabilities [10]. Other additional difficulties could be a lack of pragmatic comprehension of the situations to be evaluated, a lack of empathy with the evaluator, environmental distractions, a lack of familiarity with the context in which the assessment takes place, a lack of the ability to point, having to repeatedly answer the same question, low tolerance for frustration, and anxiety in assessment situations [11].

Another way to evaluate language performance in these populations is the use of parental reports in order to complement or replace direct measures. Parental reports also have some limitations, basically because parents have a tendency to overestimate the skills of their own children [12]. In spite of the limitations, though, some researchers have studied the relationship between direct measures and parental reports in early language development, and they have found data in favor of the use of reports by parents when studying communicational skills in children with ASD [4,6]. For instance, strong correlations have been found between parental reports such as Vineland-II (Vineland Adaptive Behavior Scale, VABS-II) [13] and the Mullen Scales of Early Learning (MSEL) [14], with direct measures such as the sequenced inventory of communication development (SICD) [15] when they compared samples of children 2.0 and 3.0 years old diagnosed with ASD, with children diagnosed with other developmental disorders [1]. The MacArthur communicative development inventories (MCDI) [16] are the most common parental reports on early language in research on children with typical or atypical development [17] and are the most common for children with ASD [18]. These reports are surveys on early communicative skills in children with typical development between the ages of 8 and 30 months. In the English version, the surveys have two forms: Words and Gestures, and Words and Sentences, which evaluate early skills for the comprehension and production of words, sentences, grammatical development, imitation, labeling, and gestures. It has also been found that in children with ASD, there are significant correlations between the scores obtained by the MacArthur communicative development inventories and the scores obtained with direct measures and other parental reports. In another study with children with ASD aged between 18 and 33 months, Luyster et al. studied the relationship between MCDI, the VABS-II survey, and the direct measure of MSEL test [6]. Comparing the scores obtained and the equivalent ages, these authors found that all measures correlated significantly, with the production component having the highest correlation. In addition, similar results have been found using these three measures in children with ASD [4,19].

However, many of the previous studies on early communicative skills in children with ASD have not included control groups and/or they have used standardized data in order to compare the results among these populations [2]. The choice of control groups supplies a framework to better understand whether the performance is due to a typical pattern of development or whether this pattern can be classified into a specific group [20]. The evidence available so far suggests that the communicative development in children with ASD differs from the development observed in children with typical development, showing a delayed development in language comprehension and production [2–5,21]. Further, it has been found that children with ASD have weaker communicative and language skills than children with other DD when matched for chronological and mental age [21–26].

Due to the difficulties in evaluating children with no or very limited language skills, children with ASD with very reduced verbal skills have been excluded from many studies [27,28]. One of the most studied aspects in the literature is the fact that children with ASD have a specific language and communicative pattern compared to children with other DD. Some authors have found a delay in language comprehension with respect to language production in children with ASD compared to preschool children with typical development [4,24], or compared to children diagnosed with a non-specified ASD [2], or compared to children with a developmental delay that does not meet the criteria for ASD [2,8], or compared to children with mental retardation [8]. Therefore, finding a delay in language comprehension compared to language production has been considered a sign for diagnosing children with ASD [8].

Another aspect that has been studied in the communicative development of children with ASD is the profile of vocabulary based on semantic and grammatical categories. The results are not conclusive because it is difficult to compare all studies due to differences in the designs and methods, the levels of development of participants, and the diagnosis criteria of the samples [29]. When authors compared the lexicon of children with ASD and children with typical development, the results were contradictory: Bruckner et al.

found differences in the lexicon for vocabulary in comprehension in children with ASD compared to children with typical development when analyzing the items of MCDI [30]. However, when comparing children with ASD and children with typical development, Charman et al. found that the vocabulary pattern of preschool children was similar [3]; in the same vein, when comparing the lexicon of children aged 2.0 and 3.0 years, Weismer et al. found that the expressive vocabulary of children with ASD was delayed but similar to the vocabulary found in late talkers, which is evidence that differences were quantitative more than qualitative [29]. Luyster et al. could not find any differences when comparing the vocabulary of children with ASD with children with typical development and mentally retarded children, with respect to some categories of MCDI such as nouns, predicates, and pivot words [8]. However, Tager-Flusberg et al. found that children with ASD used more nouns, whereas children with Down's syndrome used more pivot words (especially pronouns or determiners) when comparing children with Down's syndrome and children with ASD in a longitudinal study [31]. With respect to semantic development in children with ASD and children with typical development, many differences have been found in semantic categorizing and integration between these populations [32,33].

The use of gestures is another aspect under study with respect to communicative development in children with ASD. It has been found that children with ASD experience a delay in non-verbal communication compared to children with typical development [3,5,21,34]. Also, differences have been found when comparing the use of gestures in children with ASD and children with mental retardation. For example, Sigman and Ungerer found that the ability to imitate gestures is lower in children with ASD compared to children with mental retardation and children with typical development. In this vein, it has been found that children with Down's syndrome are more advanced in their development of gestures compared to the development of language comprehension in children with typical development [35]. Toret and Acarlar compared the development of gestures in children with ASD, children with Down's syndrome, and children with typical development, and they found differences in the frequency of gestures: children with typical development used gestures more frequently than the rest of the groups; children with ASD used gestures the least [34].

Based on the facts presented in this section, this research tries to study the communicative skills of preschool children aged between 2.0 and 6.0 years who have been diagnosed with ASD and who have low language skills. We have compared the communicative development of children with ASD and children with other DD. Previous studies have only compared communicative skills with children younger than 3.0 years old [2,8], or they have used samples with an age range that is too broad, such as a sample aged between 1.0 and 11.0 years [24]. The comparison between both groups can allow us to conclude whether all developmental disorders have the same language and communication pattern or whether these have different profiles.

Another goal was to explore the relationship between the scores of parental surveys and the scores of standardized tests, which have a direct measure, in order to check their external validity. Our prediction is that children with ASD will perform less well in linguistic and communicative skills than children with other DD and that children with ASD will perform less well in language comprehension with respect to production, which is the opposite pattern than that found in children with other DD or with typical development. With respect to the use of gestures, we expect to find a lower performance in children with ASD compared to children with other DD. Finally, we expect to find different language profiles with respect to vocabulary and semantic profiles within these populations.

2. The Present Study

2.1. Participants

The sample in this study is based on the database of two studies developed by the team of Research on Autism and Developmental Disorders, which is based at Stanford University in California, which aimed to measure the effectiveness of the pivotal response treatment

(PRT) on language skills in children with developmental disorders [36]. Participants were preschool children aged between 2.0 and 5.11 years who lived in the San Francisco Bay Area, California. The first group (ASD, n = 51, mean age = 47.33 months, boys = 89%, girls = 11%) were children with ASD. The second group (DD, n = 42, mean age = 42.07 months, boys = 58%, girls = 42%) were children with other DD who did not meet the criteria to be diagnosed with ASD. The second group was very heterogeneous (unspecified developmental delay (n = 12), developmental language disorder (n = 6), Down's syndrome (n = 6), cerebral palsy (n = 6), cri-du-chat syndrome (n = 6), Klinefelter syndrome (n = 3), and fragile X syndrome (n = 3)). This study has been approved by the Ethics Committee of the National University of Distance Learning, with the reference COEDU_FECORA. The protocol of the Ethics Committee of the National University of Distance Learning was approved on 7th May 2018. With respect to the ethnicity of participants, 70% of participants were White, 14% were Hispanic, 10% were African American, and 6% were Asian. With respect to languages spoken, 80% were monolingual English speakers, 14% were bilingual Spanish/English speakers, and 6% were Chinese/English bilingual speakers. All participants were middle class; ethnicity and language dominance was balanced among the research groups. All parents of the participants included in the sample provided written informed consent at the beginning of the study, and pertinent measures have been followed to maintain their anonymity.

2.2. Procedure

In the autism group (ASD), the participants were recruited with the following criteria: they had to have a diagnosis of ASD based on the revised version of the Autism Diagnostic Interview (ADI-R) [37], Autism Diagnostic Observation Schedule (ADOS) [38], and the opinion of a clinical expert. In addition to this, participants had to present a delay in the acquisition of language of at least one standard deviation under the mean for language production of the Preschool Language Scales 5th edition (PLS-5) [39]. In the second group of other DD, the criteria for inclusion were a diagnosis of mental retardation or language impairment based on DSM-IV-TR, CIE-10, and the evaluation of a clinical expert. In addition to this, participants needed to have a delay in the acquisition of language of at least one standard deviation under the mean for language production of the Preschool Language Scales 4th edition (PLS-4) [40]. For both groups, parents had to complete different surveys, such as Word and Gestures from MCDI and VABS-II scales. In addition, when parents visited the lab, the following tests were supplied: MSEL scales and Preschool Language Scales (PLS-4) [39,40]. In this study, we used the results obtained in the baseline for each of the researches. We did not collect any qualitative information from the parents aside from survey responses to MCDI and VABS-II scales.

2.3. Materials

With respect to cognitive development, we ran the MSEL scales [14]. The score for non-verbal IQ was obtained through the subtests of visual organization and motor skills. With respect to language development, we ran the MCDI parental report, which has two sections: the survey Words and Gestures, and the survey Words and Sentences [16]. The survey Words and Gestures has two sections: the first section measures language comprehension, labeling of objects, and imitation. Words is organized into 19 categories, which consist of nouns, sounds of animals, words for actions, words for timing, descriptive words, pronouns, interrogative words, prepositions, places, and quantifiers. Words in these categories can be classified as closed words or open words. Open words can be nouns, verbs, and adjectives; on the other hand, closed words can be pronouns, determiners, conjunctions, prepositions, and some adverbs, and it is a category to which it is difficult to add new terms; different styles of development are characterized by a different pattern in the development of these categories. Open words mean that new words can be added to the class as the need arises. The classes of nouns, verbs, and adjectives are potentially infinite since they can continually increase in the process of lexical acquisition. Open words can appear as lone words in a sentence, and they can combine with other open or closed words. On the other hand, closed words is a category to which it is difficult to add new terms (i.e., pronouns, determiners, conjunctions, prepositions, and some adverbs) since they are made up of finite sets of words. Closed words never appear as lone words in a sentence, and they are never combined with other closed words. Closed words are usually more difficult to learn since they usually have syntactic functions, such as specifying the sex and number of nouns, defining the function of a complement in a sentence, or gathering different sentences. In the parental report MCDI, the category of prepositions and places includes prepositions and adverbs of place. The category of Words for time includes basically temporal adverbs. In English, adverbs can be both open and closed words; more precisely, some adverbs of place and time, such as the ones included in MCDI, are closed class; this issue will be discussed in the Discussion section. The second section evaluates the use of gestures and consists of a list of 63 gestures organized into two sections: early gestures (e.g., Communicative Gestures, Games, and Routines) and late gestures (e.g., Actions with Objects, Pretending to be a Parent, and Imitating Adult Actions). As a second evaluation, we ran VABS-II [13], MSEL [14], and PLS scales [39,40], which are standardized tests.

2.4. Design

We used data from all surveys and tests described for both groups of participants. We tested the relation between direct measures on language production and comprehension for all tests for each sample using a Spearman correlation. In order to analyze the relations of the samples, we excluded the analysis of the Preschool Language Scale because we used a different version for each group. Then, we analyzed the language profiles within each group using equivalent ages for language comprehension and production for every group in the MSEL, VABS-II, and PLS tests in order to compare these scores with the chronological age and to compare the performance in language comprehension and production within each group. Because of the reduced number of participants, we used the non-parametric Wilcoxon test.

Afterward, we compared the direct measures between the groups for language comprehension and production. We used a non-parametric test for independent measures (i.e., the Mann–Whitney U test). Then, we studied the differences in communicative skills for MCDI between both ASD and DD groups. The analyses were applied on seven variables: the first variable was communicative skills before speech, which was calculated from five items of subtests, first signs of understanding, and starting to talk; this grouping was first used by Luyster et al. [5]. The second variable was the number of sentences (up to 28) that parents indicated their children could understand. The third and fourth variables consisted of the number of words understood and produced by children (up to 396). The fifth and sixth variables were the total number of early and late gestures following the distinction proposed by the authors of MCDI [16]. Finally, we analyzed the differences between groups with respect to the kind of vocabulary for language comprehension and production reported by parents. Nine participants in the ASD group and three of the DD group did not have any language production, and therefore we did not apply any sort of analysis for it. We explored the differences between the grammatical categories with respect to the total vocabulary of children in MCDI and then analyzed the semantic categories of nouns used compared to the proportions of each of the total of nouns for lexical comprehension and production.

3. Results

3.1. Relations between Language Measures

First of all, we compared the performance of the scores of MCDI for vocabulary and the direct measures of VABS-II and MSEL tests for both language comprehension and production, collapsing both groups: for MCDI and VABS surveys, the correlations scored the values $\rho = 0.608$, $p < 0.002$ for language comprehension and the values $\rho = 0.795$,

$p < 0.002$ for language production; for MCDI and Mullen test, the correlations scored the values $\rho = 0.462$, $p < 0.02$ for language comprehension and the values $\rho = 0.872$, $p < 0.002$ for language production; for VABS and MSEL test, the correlations scored the values $\rho = 0.57$, $p < 0.02$ for language comprehension and the values $\rho = 0.705$, $p < 0.02$ for language production. For all the analyses mentioned above, we applied the Bonferroni correction for inflated alpha levels. When we analyzed the samples within each group separately, we included the scores of PLS-4 and PLS-5. However, this test was not included in the analysis of the total sample, as mentioned before, since we used different versions for each group. In the ASD group ($n = 51$), we found significant correlations between all measures from moderate to high. However, for the other DD group ($n = 42$), the pattern was different: with respect to language production, there were correlations for all measures except for PLS-4 and VABS-II; with respect to language comprehension, we did not find significant correlations between parental reports and direct measures, which was perhaps because the sample was very heterogeneous. However, we found correlations within the parental reports (VABS-II and MCDI) and within direct measures (MSEL and PLS4). Results are shown in Table 1.

Table 1. Spearman correlations between measures of language.

ASD	ρ	p	DD	ρ	p
Comprehension					
MCDI-VABS	0.656	$p < 0.001$	MCDI-VABS	0.765	$p < 0.001$
MCDI-MSEL	0.512	$p < 0.001$	MCDI-MSEL	0.097	NS
MCDI-PLS	0.511	$p < 0.001$	MCDI-PLS	0.054	NS
PLS-VABS	0.645	$p < 0.001$	PLS-VABS	0.243	NS
MSEL-VABS	0.636	$p < 0.001$	MSEL-VABS	0.037	NS
MSEL-PLS	0.712	$p < 0.001$	MSEL-PLS	0.709	$p < 0.001$
Production					
MCDI-VABS	0.831	$p < 0.001$	MCDI-VABS	0.798	$p < 0.001$
MCDI-MSEL	0.889	$p < 0.001$	MCDI-MSEL	0.840	$p < 0.001$
MCDI-PLS	0.782	$p < 0.001$	MCDI-PLS	0.544	$p < 0.001$
PLS-VABS	0.724	$p < 0.001$	PLS-VABS	0.496	NS
MSEL-VABS	0.698	$p < 0.001$	MSEL-VABS	0.777	$p < 0.001$
MSEL-PLS	0.751	$p < 0.001$	MSEL-PLS	0.716	$p < 0.001$

Note: MCDI: MacArthur Communicative Development Inventories [16]. VABS: Vineland Adaptive Behavior Scale [13]. MSEL: Mullen Scales of Early Learning [14]. PLS: Preschool Language Scales [39,40]. p: statistical significance.

3.2. Language Profiles within Each Group

We applied the non-parametric Mann–Whitney U test to check whether there were differences between groups with respect to chronological age and non-verbal IQ. We did not find any significant difference for any of the variables and therefore did not include any of these in the analysis between groups. However, we compared the equivalent ages obtained for language comprehension and production for MSEL, PLS, and VABS-II, with respect to chronological age (see Table 2), and the data showed a lower performance to be expected for chronological age for all measures and for all groups, with significant differences ($p < 0.001$).

Table 3 shows the descriptive statistics corresponding to equivalent ages for language comprehension and production for MSEL, VABS-II, and PLS, with standard deviations, showing a high variability for all areas evaluated. In order to study all communicative profiles for each group, we analyzed the differences between the scores for equivalent ages in language comprehension and production for the three measures. With respect to the DD group, the performance average was higher in language comprehension than in language production, although when we applied the non-parametric test, we found no significant differences between language comprehension and production for any of the areas explored (see Table 1). With respect to the ASD group, the average of equivalent ages was higher for language production than language comprehension in the MSEL and PLS-5 tests, whereas in the VABS-II parental report, the age average was higher for language comprehension. In this group, we did not find any significant difference in equivalent ages between language comprehension and production. When we compare the size of the effect, we can see in Table 3 that the difference of means between language comprehension and production is higher for the DD group than the ASD group.

Table 2. Differences for equivalent ages for MSEL, PLS, and VABS-II in language comprehension and production compared to chronological age.

ASD	*U*	*p*	DD	*U*	*p*
MSEL (*n* = 51)			**MSEL** (*n* = 42)		
Comprehension	17. 77	$p < 0.001$	Comprehension	7.72	$p < 0.001$
Production	15.77	$p < 0.001$	Production	10.81	$p < 0.001$
PLS-5 (*n* = 51)			**PLS-4** (*n* = 42)		
Comprehension	19.04	$p < 0.001$	Comprehension	8.15	$p < 0.001$
Production	19.52	$p < 0.001$	Production	8.28	$p < 0.001$
VABS-II (*n* = 51)			**VABS-II** (*n* = 42)		
Comprehension	19.34	$p < 0.001$	Comprehension	24.04	$p < 0.001$
Production	18.71	$p < 0.001$	Production	8.22	$p < 0.001$

Note: MSEL: Mullen Scales of Early Learning [14]. PLS: Preschool Language Scales (PLS-4) [39,40]. VABS-II: Vineland Adaptive Behavior Scale [13]. *p*: statistical significance.

Table 3. Descriptive statistics for equivalent age on language comprehension and production for all measures of language MSEL, VABS-II, and PLS.

			Comprehension			Production			Size of Effect	Contrast Comprehension and Production	
	n		Median	M	SD	Median	M	SD	*d*	Z	*p*
ASD	51	MSEL	14.5	17.33	7.73	16.5	17.41	8.98	−0.01	0.396	$p > 0.05$
	51	PLS-5	16	17.7	6.32	18	17.98	6.16	−0.04	−0.45	$p > 0.05$
	50	VABS-II	15	15.96	6.79	15.5	15.72	7.03	0.04	−0.245	$p > 0.05$
DD	42	MSEL	22	21.46	9.14	16	16.69	6.42	0.6	−2.125	$p < 0.05$
	42	PLS-4	25	23.77	8.77	21	20.23	3.68	0.53	−1.06	$p > 0.05$
	42	VABS-II	23	25.92	12.98	19	19	5.15	0.7	−1.69	$p > 0.05$

Note. MSEL: Mullen Scales of Early Learning [14]. VABS-II: Vineland Adaptive Behavior Scale [13]. PLS: Preschool Language Scales [39,40]. M: mean. SD: standard deviation. *d*: *d* Cohen statistics for size of effect. Z: statistics for contrast U Mann–Whitney. *p*: statistical significance.

3.3. Differences between Groups

Table 4 shows the descriptive statistics (mean and standard deviation) for language comprehension and production, the effect of size for the differences between groups, and the significance obtained for each variable using the Mann–Whitney U test.

In order to analyze the differences in language comprehension and production between groups, we compared the direct measures for MSEL and VABS-II. We also compared performance in word comprehension and production for MCDI and sentence comprehension for each group. With respect to language comprehension, we found significant differences for all variables, with the highest score average in the DD group for the two standardized tests (MSEL ($z = -2.102$, $p = 0.04$) and VABS-II ($z = -3.259$, $p = 0.001$)) and the survey MCDI: vocabulary comprehension total ($z = -1.061$, $p = 0.289$) and sentence comprehension ($z = -2.222$, $p = 0.026$). However, with respect to language production, none of the analyses showed any significant differences for any variable analyzed.

Afterward, we analyzed the differences between groups with respect to pre-speech skills: the analysis showed no significant differences between groups with respect to skills previous to speech for MCDI ($z = -1.061$, $p = 0.289$). With respect to the differences in language comprehension and production in the MCDI categories, we compared the proportion of the number of words for each category with respect to total vocabulary for each participant. In vocabulary comprehension (see Table 5), we observed significant differences for prepositions, where the highest proportion was in the DD group (z = −2.866, *p* = 0.004). We could not find any significant differences in the rest of the categories. With respect to language production (Table 6), we found significant differences in the proportions of the following three categories, with the highest scores in the DD group: words related to time (z = −3, 03, *p* = 0.002), pronouns (z = −2.193, *p* = 0.028) and prepositions (z = −2.928, *p* = 0.003). Further, we found significant differences between groups with respect to the proportions of adjectives in the total sample lexical production and, this time, the ASD group obtained the highest proportions (z = 2.284, *p* = 0.022). With respect to the analysis of the semantic categories of nouns, we could not find any significant differences in the proportions of any of the categories with respect to the total number of nouns that are part of lexical comprehension and production in children, following the information supplied by parents.

Table 4. Descriptive statistics corresponding to direct measures in language comprehension and production in MSEL, VABS-II, PLS, and MCDI and differences between groups.

Test	ASD		DD		ASD vs. DD	
	M	**SD**	**M**	**SD**	***d***	**Z**
MSEL						
Comprehension	18.10	6.63	22.77	6.66	−0.7	(z = −2.102, *p* = 0.04)
Production	16.96	7.52	17.54	4.86	−0.09	(z = −0.401, *p* = 0.69)
PLS						
Comprehension	22.18	5.58	27.50	6.89	−0.85	(z = −2.608, *p* = 0.009)
Production	22.73	5.19	25.57	3.30	−0.67	(z = −1.905, *p* = 0.057)
VABS-II						
Comprehension	15.18	5.98	22.00	5.91	−1.15	(z = −3.259, *p* = 0.001)
Production	24.75	11.71	30.15	10.87	−0.48	(z = −1.721, *p* = 0.085)
MCDI						
Comprehension	181.12	114.25	223.93	99.83	−0.4	(z = −1.061, *p* = 0.289)
Production	94.45	99.03	62.99	97.84	0.32	(z = 0.823, *p* = 0.410)
Pre-speech skills	3.55	1.40	4.29	0.99	−0.62	(z = −1.783, *p* = 0.075)
Sentence comprehension	16.20	8.22	21.57	5.12	−0.8	(z = −2.222, *p* = 0.026)

Note. MSEL: Mullen Scales of Early Learning [13]. VABS-II: Vineland Adaptive Behavior Scale [12]. PLS: Preschool Language Scales (PLS-4: [38,39]). M: mean. SD: standard deviation. *d*: *d* Cohen statistics for size of effect. Z: statistics for contrast U Mann–Whitney. *p*: statistical significance.

With respect to the use of gestures (see Table 7), the results show significant differences between groups, with higher scores for the DD group regarding total score (z = −3.001, *p* = 0.003), in early gestures (z = −3.41, *p* = 0.001), and late gestures (z = −3.001, *p* = 0.003).

Table 5. Descriptive statistics corresponding to the proportions for each MCDI category with respect to lexical comprehension.

Test	ASD		DD			ASD vs. DD
MCDI	**M**	**SD**	**M**	**SD**	***d***	**Z, *p***
Comprehension						
Categories of MCDI						
Nouns	0.69	0.08	0.68	0.04	0.17	(z = 0.527, *p* = 0.598)
Verbs	0.17	0.05	0.15	0.02	0.57	(z = 0.997, *p* = 0.319)
Time	0.01	0.01	0.004	0.01	0.6	(z = 0.578, *p* = 0.563)
Adjectives	0.05	0.03	0.06	0.02	−0.4	(z = −0.950, *p* = 0.342)
Pronouns	0.01	0.01	0.01	0.01	0	(z = −1.94, *p* = 0.052)
Interrogatives	0.01	0.01	0.004	0.005	0.8	(z = −0.327, *p* = 0.744)
Prepositions	0.02	0.01	0.03	0.02	−0.67	(z = −2.866, *p* = 0.004)
Quantifiers	0.01	0.01	0.01	0.01	0	(z = −0.68, *p* = 0.496)
Sounds	0.05	0.03	0.05	0.02	0	(z = −0.743, *p* = 0.458)

Note: MacArthur Communicative Development Inventories [16]. M: mean. SD: standard deviation. *d*: *d* Cohen statistics for size of effect. Z: statistics for contrast U Mann–Whitney. *p*: statistical significance.

Table 6. Descriptive statistics corresponding to the proportions for each category of MCDI with respect to lexical production.

Test	ASD			DD		ASD vs. DD
MCDI Production	**M**	**SD**	**M**	**SD**	***d***	**Z, *p***
Categories of MCDI						
Nouns	0.19	0.19	0.15	0.17	0.22	(z = 0.144, *p* = 0.886)
Verbs	0.11	0.08	0.07	0.05	0.61	(z = 1.687, *p* = 0.092)
Time	0.001	0.004	0.04	0.05	−1.44	(z = −3.03, *p* = 0.002)
Adjectives	0.03	0.04	0.01	0.02	0.67	(z = 2.284, *p* = 0.022)
Pronouns	0.01	0.02	0.02	0.02	−0.5	(z = −2.193, *p* = 0.028)
Interrogatives	0.002	0.004	0.0003	0.001	0.65	(z = 1.389, *p* = 0.165)
Prepositions	0.02	0.03	0.06	0.07	−0.8	(z = −2.928, *p* = 0.003)
Quantifiers	0.01	0.02	0.02	0.02	−0.5	(z = −0.833, *p* = 0.405)
Sounds	0.08	0.16	0.09	0.05	−0.09	(z = −1.933, *p* = 0.053)

Note: MacArthur Communicative Development Inventories [16]. SD: standard deviation. *d*: *d* Cohen statistics for size of effect. Z: statistics for contrast U Mann–Whitney. *p*: statistical significance.

Table 7. Descriptive statistics corresponding to direct measures for early gestures, late gestures, and the total number for MCDI.

Test	ASD		DD			ASD vs. DD
	M	**SD**	**M**	**SD**	***d***	**Z, *p***
MCDI Gestures						
Early	9.47	3.97	13.71	3.2	−1.17	(z = −3.41, *p* = 0.001)
Late	21.14	9.63	30.07	11.42	−0.84	(z = −2.659, *p* = 0.008)
Total	30.47	12.38	43.79	13.92	−1.01	(z = −3.001, *p* = 0.003)

Note: MacArthur Communicative Development Inventories [16]. SD: standard deviation. *d*: *d* Cohen statistics for size of effect. Z: statistics for contrast U Mann–Whitney. *p*: statistical significance.

4. Discussion

4.1. Relations between Language Measures

After fully analyzing the data of the sample, including the children with ASD and DD, we found significant correlations between all measures with respect to the scores in language comprehension and production. The highest correlations were in language production. We observed the same pattern of correlations when analyzing children with

ASD and children with other DD. These results replicate the relations found in previous studies [2,6,19]; as a novelty, this research studied these correlations in a sample of children with low verbal skills. Children with ASD showed high significant correlations both in direct measures and parental reports. These results show evidence in favor of the use of parental reports in the study of communicative development in children with ASD with low verbal skills since these reports are significantly related to direct measures and standardized tests.

With respect to the group of children with other DD, the significant correlations between parental measures and language proficiency tests were restricted to all correlations except for PLS-4 and VABS-II on language production and restricted to direct measures (MSEL and PLS) and parental reports (MCDI and VABS-II) on language comprehension. Following these results, we can conclude that direct measures and parental reports offer differentiated information, depending on when we study language comprehension in children with other DD. Previous studies found that it is very difficult to assess language comprehension in children with communicative difficulties since the conditions where the assessment takes place [11] or the motivational and attentional aspects [10] make it difficult to observe the capacities of language comprehension. Other studies found no weaker agreement when assessing language comprehension, and they found that parent reports of language skills were equivalent to scores on direct testing in language comprehension [41]. Miller et al. argue that it might be due to their reliance on Vineland, which is a semi-structured parent review, instead of a parent report checklist; this outcome suggests to these authors that parents are usually reliable reporters. Taking into account parental reports, it has also been observed in previous studies that parents overestimate the skills in language comprehension of their own children [12]. Some authors suggest that parents usually report higher fine motor skills compared to direct assessment; this could be due to the fact that parents assume that children can perform a motor task without having observed it; in addition, children might not want to perform some tasks during assessment because they are not interested in the testing materials or because they have difficulties comprehending testing demands [41]. Another reason for the discrepancy is the fact that a child might not perform an item during a direct assessment, but she might perform that item at home. The discrepancy between these language measures could reflect the fact that the assessment by parents of their own children differs from the assessment by expert evaluators, although we should be careful because of the heterogeneity and the size of the samples. However, based on the results of this study, we can conclude that a suitable strategy for language assessment would be to combine the information supplied by direct measures together with the information supplied from parental reports [9].

4.2. Language Profiles within Each Group

With regard to the language profiles for each group, after comparing the scores expected for their age on the standardized measures, the results show a delay in language skills for both groups. This is not a surprise since communicative and language difficulties are basic symptoms in ASD and developmental language disorder [1]. In addition, one of the conditions to be part of the samples in this study was to have a language delay.

In order to find out whether there are different language profiles for both groups, we compared the scores for typical children with equivalent ages for language comprehension and production. In the DD group, we found a typical language pattern, where language comprehension skills were more developed than language production skills. With respect to children with ASD, we expected to find the opposite pattern since previous studies have found differences in this direction [24]. In fact, we found that in the parental report of VABS-II, the mean of the scores was higher for language production than language comprehension. With respect to the rest of the language measures (MSEL and PLS-5), we found the same pattern of higher scores in language production but with no significant differences. In any case, we found a delay in language comprehension with respect to language production for children with ASD if we compare the scores with the group of children with other DD; this delay could be due to the difficulties of children with ASD to generalize skills across contexts [42]. However, some previous studies did not find a significant interaction between language measures and the DD group [41]. If we take a look at the size of the effect of the differences between language comprehension and production, we find that the size of the effect is higher for children with other DD. In addition, when we compared the direct measures for language comprehension in different measures, we observed significant differences in both groups for all variables in language comprehension, with higher performance in children with other DD; however, we found no significant differences in language production for none of the measures. Therefore, both groups have a similar performance in language production, but difficulties in language comprehension were greater for children with ASD [3–5]. Another aspect that could have been taken into account is the fact that some skills could have a reporter bias: it has been found in previous studies that there is a lower correlation for the assessment of language comprehension compared to language production [12,43]. Even though we did not test this fact in this study, previous studies have found mixed results: while there is a high correlation of items measuring basic skills, it is not the case for more demanding language skills [41].

4.3. Differences between Groups in Language Skills

When we compare language skills between both groups, we find that the levels of language comprehension differ significantly between groups, whereas the levels of language production are similar between both groups. With respect to language comprehension, we found significant differences for all the analyzed variables, that is, the performance in language comprehension in direct measures, Mullen and PLS with the parental report VABS-II, and with sentence comprehension and vocabulary size of MCDI.

With respect to the properties of vocabulary based on the categories proposed for MCDI, the results show that language comprehension is very similar for both groups. In fact, when we compare the proportions for each category of word with respect to the total number of words understood, we found no difference, except for prepositions, since the group of children with other DD had higher proportions of these categories. With respect to language production, the distribution of words based on categories for both groups is similar, except for the fact that children with other DD have a higher proportion of language production for prepositions, pronouns, and words for time. Therefore, the results obtained when we compare these samples supply more evidence that favors the results found by Tager-Flusberg et al. [31] about the tendency for children with Down's syndrome to use more closed words than do children with ASD, since we found that the group of children with other DD had higher proportions for prepositions and places than the group of children with ASD. Finally, in language production, we found that children with ASD used more adjectives proportionally than children with other DD, which provides more evidence for the existence of different patterns of lexical categories in language production in children with ASD, with respect to children with other DD regarding the difference between open and closed words. This different pattern could be due to different cognitive styles among these populations: Nelson found that expressive children use a higher rate of pronouns than referential children, who are more focused on the learning of full noun phrases and words; this difference is higher when children have a low MLU (i.e., in the initial stages of language acquisition) [44]. She also found that referential children produce more subjects with the thematic role of agent, whereas referential children use more subjects with the thematic role of experiencer. Nelson found that referential children use more qualifying adjectives, whereas referential children use more possessive adjectives. According to Nelson, referential children start to learn lexical items and afterward learn the parameters concerning phrase structure, whereas expressive children learn patterns of word order and then they increase their lexical repertoire. Lieven found that referential children used sentences with less variability than expressive children [45,46]. Also, she found that referential children are more analytical since they learn lexical items with fewer complements and specifiers than expressive children, whereas expressive children include more components in their phrases, and these forms are less analyzed. Bates et al. also found that expressive children had more social skills and better memory storage than referential children, whereas referential children were more analytical [47]. She proposes that the individual differences found in language development depend on faculties such as analytical processing versus memory skills and the performance on language production versus performance on language comprehension. Therefore, it could be the case that children with ASD could have a referential cognitive style compared to children with other DD, who could have a more expressive style.

We then compared the semantic categories of nouns from the lexical sample of both groups found in the parental reports. However, in this case, we did not find any significant difference in the proportions observed for either lexical comprehension or language production between both children with ASD and children with other DD.

4.4. Differences between Groups on Non-Verbal Communication

Finally, we analyzed the differences in non-verbal communication between both groups, and our results provide evidence in favor of the higher performance of children with other DD, because we found significant differences in the use of gestures with respect to children with ASD, based on parental reports. It could be the case that the lower use of gestures in ASD is related to their communication difficulties in this population and with better communication abilities in Down's syndrome. Even though it only applies to a subgroup of the children with DD, previous studies found that children with Down's syndrome are more advanced in their development of gestures compared to children with typical development [35]. Caselli et al. studied the development of language and communication in children with Down's syndrome. The goal of this research was to examine the relations between language comprehension, language production, and the development of gestures in children with Down's syndrome compared to typically developing children. They found that children with Down's syndrome had a lower performance compared to typically developing children in language development. They found a similar development between lexical comprehension and the development of gestures. However, they found that children with Down's syndrome had a higher gestural development compared to typically developing children [35]. They found that children with Down's syndrome produce a higher frequency of symbolic communicative gestures, pretending gestures, and actions to perform symbolic transformations. Following Caselli et al., in the initial stages, the gestural and vocal production of children with Down's syndrome are similar to those of typically developing children matched for word comprehension; however, they found that later on, symbolic communicative gestures and actions increase and are more developed in children with Down's syndrome, based on their level of development of word comprehension and production. This fact could explain the data that we have obtained in our study so far.

5. Conclusions

The results found in this research might have implications for the assessment of children with low language and communication skills: the consistency between different measures supports the use of direct measures and parental reports for therapists working with children with ASD and other DD. The specific patterns found in the delay in the development of language comprehension, the properties of vocabulary, and the low use of gestures of children with ASD compared to children with other DD could help practitioners with a differential diagnosis after a deeper exploration from a clinical perspective. The results found in this research underline the importance of including improvements in verbal and non-verbal communication in children with ASD as important goals on intervention. However, we should be cautious because we did not collect any qualitative information from the parents aside from survey responses to MCDI and VABS-II scales, also because there was a high heterogeneity of the participants, and because the size of the sample was small in this study, which could make it difficult to generalize the results to other studies and to provide a complete profile of the properties of language and communication in these populations.

Author Contributions: Conceptualization, V.T.; methodology, V.T.; software, C.R.; validation, V.T., and C.R.; formal analysis, C.R.; investigation, V.T.; resources, C.R.; data curation, C.R.; writing—original draft preparation, C.R.; writing—review and editing, V.T.; visualization, C.R.; supervision, V.T.; project administration, V.T. Both authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board (or Ethics Committee) of National University of Distance Learning (protocol code COEDU_FECORA, 7th May 2018l).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Acknowledgments: We would like to thank Stanford University and the audience of the Conference on Developmental Language Disorders (DevO). We would also like to especially thank all children in this project for their willingness to participate in this research.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. *Diagnostic and Statistical Manual of Mental Disorders*, 5th ed.; American Psychiatric Association: Washington, DC, USA, 2013.
2. Weismer, S.E.; Lord, C.; Esler, A. Early Language Patterns of Toddlers on the Autism Spectrum Compared to Toddlers with Developmental Delay. *J. Autism Dev. Disord.* **2010**, *40*, 1259–1273. [CrossRef]
3. Charman, T.; Drew, A.; Baird, C.; Baird, G. Measuring early language development in preschool children with autism spectrum disorder using the MacArthur Communicative Development Inventory (Infant Form). *J. Child Lang.* **2003**, *30*, 213–236. [CrossRef]
4. Hudry, K.; Leadbitter, K.; Temple, K.; Slonims, V.; McConachie, H.; Aldred, C.; Howlin, P.; Charman, T.; PACT Corporation. Preschoolers with autism show greater impairment in receptive compared with expressive language abilities. *Int. J. Lang. Dis.* **2010**, *45*, 681–690. [CrossRef]
5. Luyster, R.; Lopez, K.; Lord, C. Characterizing communicative development in children referred for Autism Spectrum Disorders using the MacArthur-Bates Communicative Development Inventory (CDI). *J. Child Lang.* **2007**, *34*, 623–654. [CrossRef] [PubMed]
6. Luyster, R.; Kadlec, M.B.; Carter, A.; Tager-Flusberg, H. Language assessment and development in toddlers with autism spectrum disorders. *J. Autism Dev. Disord.* **2008**, *38*, 1426–1438. [CrossRef] [PubMed]
7. Luyster, R.; Seery, A.M.; Thompson, M.R.; Tager-Flusberg, H. Identifying early risk markers and developmental trajectories for language impairment in neurodevelopmental disorders. *Dev. Disabil. Res. Rev.* **2011**, *17*, 151–159. [CrossRef] [PubMed]
8. Barbaro, J.; Dissanayake, C. Developmental profiles of infants and toddlers with autism spectrum disorders identified prospectively in a community-based setting. *J. Autism Dev. Disord.* **2012**, *42*, 1939–1948. [CrossRef] [PubMed]
9. Charman, T. Matching preschool children with autism spectrum disorders and comparison children for language ability: Methodological challenges. *J. Autism Dev. Disord.* **2004**, *34*, 59–64. [CrossRef] [PubMed]
10. Koegel, L.K.; Koegel, R.L.; Smith, A. Variables related to differences in standardized test outcomes for children with autism. *J. Autism Dev. Disord.* **1997**, *27*, 233–243. [CrossRef]
11. Tager-Flusberg, H. The challenge of studying language development in autism. In *Methods for Studying Language Production*; Menn, L., Bernstein-Ratner, N., Eds.; Lawrence Erlbaum Associates: Mahwah, NJ, USA, 1999; pp. 313–332.
12. Tomasello, M.; Mervis, C.B. The instrument is great, but measuring comprehension is still a problem. *Monogr. Soc. Res. Child Dev.* **1994**, *59*, 174–179. [CrossRef]
13. Sparrow, S.S.; Cicchetti, D.V.; Balla, D.A. *Vineland Adaptive Behavior Scales*, 2nd ed.; Pearson: San Antonio, TX, USA, 2005.
14. Mullen, E.M. *Mullen Scales of Early Learning: AGS Edition*; American Guidance Service: Circle Pines, MN, USA, 1995.
15. Hendrick, D.; Prather, E.; Tobin, A. *Sequenced Inventory of communication Development*, Revised Edition; University of Washington Press: Seattle, WA, USA, 1984.
16. Fenson, L.; Marchman, V.; Thal, D.; Dale, P.; Reznick, S.; Bates, E. *The MacArthur Communicative Development Inventories: User's Guide and Technical Manual*, 2nd ed.; Brookes: Baltimore, ML, USA, 2006.
17. De Bildt, A.; Serra, M.; Luteijn, E.; Kraijer, D.; Sytema, S.; Minderaa, R. Social skills in children with intellectual disabilities with and without autism. *J. Intellect. Disabil. Res.* **2005**, *49*, 317–328. [CrossRef]
18. Tager-Flusberg, H.; Rogers, S.; Cooper, J.; Landa, R.; Lord, C.; Paul, R. Defining spoken language benchmarks and selecting measures of expressive language development for young children with autism spectrum disorders. *J. Speech Lang. Hear. Res.* **2009**, *52*, 643–652. [CrossRef]
19. Paul, R.; Chawarska, K.; Chicchetti, D.; Volkmar, F. Language outcomes of toddlers with autism spectrum disorders: A two-year follow-up. *Autism* **2008**, *1*, 99–107. [CrossRef]

20. Burack, J.A.; Iarocci, G.; Bowler, D.; Mottron, L. Benefits and pitfalls in the merging of disciplines: The example of developmental psychopathology and the study of persons with autism. *Dev. Psychopathol.* **2002**, *14*, 225–237. [CrossRef] [PubMed]
21. Sigman, M.; Ungerer, J. A Cognitive and language skills in autistic, mentally retarded, and normal children. *Dev. Psychol.* **1984**, *20*, 293–302. [CrossRef]
22. Carpentieri, S.; Morgan, S.B. Adaptive and intellectual functioning autistic and nonautistic retarded children. *J. Autism Dev. Disord.* **1996**, *26*, 611–620. [CrossRef] [PubMed]
23. Loveland, K.A.; Kelley, M.L. Development of Adaptive Behavior in Preschoolers with Autism or Down Syndrome. *Am. J. Ment. Ret.* **1991**, *96*, 13–20.
24. Maljaars, J.; Noens, I.; Scholte, E.; van Berckelaer-Onnes, I. Language in Low-Functioning Children with Autistic Disorder: Differences between Receptive and Expressive Skills and Concurrent Predictors of Language. *J. Autism Dev. Disord.* **2012**, *42*, 2181–2191. [CrossRef] [PubMed]
25. Perry, A.; Flanagan, H.E.; Dunn Geier, J.; Freeman, N.L. Brief Report: The Vineland Adaptive Behavior Scales in Young Children with Autism Spectrum Disorders at Different Cognitive Levels. *J. Autism Dev. Disord.* **2009**, *39*, 1066–1078. [CrossRef]
26. Stone, W.L.; Ousley, O.Y.; Hepburn, S.L.; Hogan, K.L.; Brown, C.S. Patterns of adaptive behavior in very young children with autism. *Am. J. Ment. Retard.* **1999**, *104*, 187–199. [CrossRef]
27. Plesa-Skwerer, D.; Jordan, S.E.; Brukilacchio, B.H.; Tager-Flusberg, H. Comparing methods for assessing receptive language skills in minimally verbal children and adolescents with autism spectrum disorders. *Autism* **2016**, *20*, 591–604. [CrossRef] [PubMed]
28. Tager-Flusberg, H.; Kasari, C. Minimally verbal school-aged children with autism spectrum disorder: The neglected end of the spectrum. *Autism Res.* **2013**, *6*, 468–478. [CrossRef] [PubMed]
29. Weismer, S.; Gernsbacher, M.A.; Stronach, S.; Karasinki, C.; Eernisse., E.R.; Venker, C.E.; Sindberg, H. Lexical and grammatical skills in toddlers on the autism spectrum compared to late talking toddlers. *J. Autism Dev. Disord.* **2011**, *41*, 1065–1075. [CrossRef] [PubMed]
30. Bruckner, C.; Yoder, P.; Stone, W.; Saylor, M. Construct validity of the MCDI-I receptive vocabulary scale can be improved: Differential item functioning between toddlers with autism spectrum disorders and typically developing infants. *J. Speech Lang. Hear. Res.* **2007**, *50*, 1631–1638. [CrossRef]
31. Tager-Flusberg, H.; Calkins, S.; Nolin, T.; Baumberger, T.; Anderson, M.; Chadwick-Dias, A. A longitudinal study of language acquisition in autistic and Down syndrome children. *J. Autism Dev. Disord.* **1990**, *20*, 1–21. [CrossRef]
32. Naigles, L.R.; Tek, S. 'Form is easy, meaning is hard' revisited: (re) characterizing the strengths and weaknesses of language in children with autism spectrum disorder. *WIREs Cog. Sci.* **2017**, e1438. [CrossRef]
33. Portrzeba, E.R.; Fein, D.; Naigles, L. Investigating the shape bias in typically developing children and children with autism spectrum disorders. *Front. Psychol.* **2015**, *6*, 446.
34. Toret, G.; Acarlar, F. Gestures in Prelinguistic Turkish Children with Autism, Down Syndrome and Typically Developing Children. *Educ. Sci. Theory Pr.* **2011**, *11*, 1471–1478.
35. Caselli, M.; Vicari, S.; Longobardi, E.; Lami, L.; Pizzoli, C.; Stella, G. Gestures and words in early development of children with Down syndrome. *J. Speech Lang. Hear. Res.* **1998**, *41*, 1125–1135. [CrossRef]
36. Mohammadzaheri, F.; Koegel, L.K.; Rezaee, M.; Rafiee, S.M. A randomized clinical trial comparison between Pivotal Response Treatment (PRT) and structured Applied Behavior Analysis (ABA) intervention for children with autism. *J. Autism Dev. Disord.* **2014**, *44*, 2769–2777. [CrossRef]
37. Lord, C.; Rutter, M.; Le Couteur, A. Autism diagnostic interview-revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J. Autism Dev. Disord.* **1994**, *24*, 659–685. [CrossRef]
38. Lord, C.; Risi, S.; Lambrecht, L.; Cook, E.H., Jr.; Leventhal, B.L.; Di Lavore, P.C.; Rutter, M. The autism diagnostic observation schedule-generic: A standard measure of social and communication deficits associated with the spectrum of autism. *J. Autism Dev. Disord.* **2000**, *30*, 205–223. [CrossRef] [PubMed]
39. Zimmerman, I.L. *PLS-5: Preschool Language Scales*; Pearson: London, UK, 2011.
40. Zimmerman, I.L.; Steiner, V.G.; Pond, R.E. *Preschool Language Scale*, 4th ed.; The Psychological Corporation: San Antonio, TX, USA, 2002.
41. Miller, L.E.; Perkins, K.A.; Dai, Y.G.; Fein, D.A. Comparison of Parent Report and Direct Assessment of Child Skills in Toddlers. *Res. Autism Spectr. Disord.* **2017**, *41*, 57–65. [CrossRef] [PubMed]
42. Nordahl-Hansen, A.; Kaale, A.; Ulvund, S.E. Language assessment in children with autism spectrum disorder: Current validity between report-based assessments and direct tests. *Res. Autism Spectr. Disord.* **2014**, *8*, 1100–1106. [CrossRef]
43. Sachse, S.; Suchodoletz, W. Early identification of language delay by direct language assessment or parent report? *J. Dev. Behav. Pediatrics* **2008**, *29*, 34–41. [CrossRef]
44. Nelson, K. Individual differences in language development: Implications for development and language. *Dev. Psychol.* **1981**, *17*, 170–187. [CrossRef]
45. Lieven, E. Conversations between mothers and young children: Individual differences and their possible implications for the study of language learning. In *The Development of Communication: Social and Pragmatic Factors in Language Acquisition*; Waterson, N., Snow, C., Eds.; Wiley: New York, NY, USA, 1978.

46. Lieven, E.; Pine, J.; Barnes, H.D. Individual differences in early vocabulary development: Redefining the referential-expressive distinction. *J. Child Lang.* **1992**, *19*, 287–310. [CrossRef]
47. Bates, E.; Bretherton, I.; Snyder, L. *From First Words to Grammar*; Cambridge University Press: Cambridge, UK, 1988.

children

MDPI

Article

Prevalence of Language Delay among Healthy Preterm Children, Language Outcomes and Predictive Factors

Miguel Pérez-Pereira

Department of Developmental and Educational Psychology, Faculty of Psychology, University of Santiago de Compostela, 15705 Santiago de Compostela, Spain; miguel.perez.pereira@usc.es; Tel.: +34-600940125

Abstract: Language delay (LD) and its relationship with later language impairment in preterm children is a topic of major concern. Previous studies comparing LD in preterm (PT) and full-term (FT) children were mainly carried out with samples of extremely preterm and very preterm children (sometimes with additional medical problems). Very few of them were longitudinal studies, which is essential to understand developmental relationships between LD and later language impairment. In this study, we compare the prevalence of LD in low-risk preterm children to that of FT children in a longitudinal design ranging from 10 to 60 months of age. We also analyze which variables are related to a higher risk of LD at 22, 30 and 60 months of age. Different language tests were administered to three groups of preterm children of different gestational ages and to one group of full-term children from the ages of 10 to 60 months. ANOVA comparisons between groups and logistic regression analyses to identify possible predictors of language delay at 22, 30 and 60 months of age were performed. The results found indicate that there were practically no differences between gestational age groups. Healthy PT children, therefore, do not have, in general terms, a higher risk of language delay than FT children. Previous language delay and cognitive delay are the strongest and longest-lasting predictors of later language impairment. Other factors, such as a scarce use of gestures at 10 months or male gender, affect early LD at 22 months of age, although their effect disappears as children grow older. Low maternal education appears to have a late effect. Gestational age does not have any significant effect on the appearance of LD.

Keywords: preterm children; language delay; predictive factors; language development

Citation: Pérez-Pereira, M. Prevalence of Language Delay among Healthy Preterm Children, Language Outcomes and Predictive Factors. *Children* **2021**, *8*, 282. https://doi.org/10.3390/children8040282

Academic Editor: Anneloes van Baar

Received: 22 February 2021
Accepted: 2 April 2021
Published: 6 April 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

Preterm children are considered to be an at-risk population, though not all of them share the same percentage of risk. Important differences exist among preterm children in relation to different biomedical factors. One of them, gestational age (GA), also determines whether other factors co-exist. Usually, birth weight (BW) is strongly associated with GA, in such a way that the shorter the GA the lower the BW (with the exception of those children small for GA). Preterm children are classified according to GA into 4 groups [1–3]: late preterm children (LPT), who have a GA of 34–36 weeks; moderately preterm (MPT) children, with a GA between 32–33 weeks; very preterm (VPT) children, with a GA between 28–31 weeks; and extremely preterm (EPT) children, with a GA below 28 weeks.

The risk of suffering medical complications increases as GA and BW are lower. EPT and VPT children have a greater probability of being affected by them than LPT and MPT children [4]. The most common medical complications affect lungs (bronchopulmonary dysplasia (BPD), respiratory distress syndrome) and cerebrum (intraventricular hemorrhage (IVH), periventricular leukomalacia (PLM)), with important consequences for children's development [5].

The study of language development in preterm children has offered controversial results. The majority of studies, as well as a few meta-analyses, have reported that preterm children of different ages tend to show lower results than full-term peers in a diversity of language measures [6–8]. Not only do PT children show a smaller vocabulary size than their FT counterparts, but they also show a lower level of grammatical skills than their FT peers in their first years [9–12].

These studies were mostly carried out with samples of EPT or VPT (or very and extremely low birth weight) children, and some of them did not report and/or use clear exclusion criteria. This means that children with serious biomedical complications or sensory, motor and cognitive handicaps may be unidentified, producing a confounding effect with prematurity on the outcomes. In contrast, a few studies, mostly carried out with healthy preterm children with a variety of GAs, have not found significant differences between PT and FT children in different measures of language development taken at different ages [13–15].

The research of the prevalence of late talkers (LT) or language delayed (LD) children in PT as compared to FT children, the language outcomes of these children a few years later and the predictors of LD is the main focus of this study. This research is related to the study of the differences in language development between PT and FT children, but of a slightly different kind.

Late talkers or language delayed children are those children between 18 and 36 months of age who show limited language development as compared to their FT peers, in the absence of neurological damage, environmental deprivation, sensory impairment or cognitive delay. The cut-off criterion used to establish LD varies in different studies, and it depends on the reference age. However, quite commonly, those children below the 10th percentile in language tests are considered to be LD or LT [16]. This criterion is usable for children of different ages in the above-mentioned range, while, in contrast, other criteria, such as to have a productive vocabulary below 50 words and no word combinations is only well suited for the age of 24 months [17,18]. Late talkers may have combined expressive and receptive delays or only expressive delays. The estimates of prevalence of LT or LD children oscillate between 9% and 20% of the population of children aged 24–36 months [15,19]. Some of these children (around 50–70%), called *late bloomers*, catch up to their typically developing (TD) peers by 4–5 years of age [20]. This fact points to the difficulty of predicting later language impairment from early language delay. Early prediction of language impairment or developmental language disorder after 5 years of age is a major focus of concern and developmental follow-up.

Table 1 displays the main findings of those studies which have investigated the prevalence of LD in PT children as compared to FT. In all of them, age correction for prematurity has been applied, with the exception of Lee and Lee [21] in which comparisons were performed using chronological age. Only studies which have identified LD with language specific and appropriate tests were included. For this reason, two studies were not included in Table 1, because the test used is not considered a language ability test, but of verbal intelligence (Wechsler Preschool and Primary Scale of Intelligence) [22,23]. Another two studies which used only partial versions (selection of only a few items) of developmental scales [24,25] were also excluded. Those studies which did not use normative scores (percentile, standard deviation (SD)) to establish cut off criteria were not included either [26].

Table 1. Summary of the investigations which compared the prevalence of LD in preterm (PT) and full-term (FT) children.

References (Ordered by Publication Year)	FT and PT Group Characteristics	Age of Assessment	% of Delayed PT Children	% of Delayed FT Children	Language Measure and Classification Criteria
Singer et al. [27]	98 VLBW with BPD Mean GA 27 (SD:2) weeks.	36 months			BDI communication subscale
	70 VLBW without BPD. Mean GA: 30 (2)				DQ < 85 (equivalent to <−1 SD)
	95 FT children. Groups did not differ in gender, maternal education SES, race. Exclusion: serious neurological problems, socially disadvantaged.		49%BPD/34%	30% $	Receptive
			44%BPD/25%	25% $	Expressive
			43%BDP/31%	28%	Overall
Wolke et al. [28]	241 EPT children (GA ≤25 weeks)	72 months			PLS-3 < −2 SD
			15.6%	1.9% *	Total score
			9.5%	1.3%	Receptive
			12.1%	1.3%	Expressive
Woodward et al. [5]	105 EPT/VPT Mean GA: 28 weeks. 16%: cerebral palsy	48 months			CELF-P < −1 SD
			30%	15.2%*	Receptive
	107 FT with higher SES than the PT		25%	12.4%*	Expressive
			31%	15.2%*	Overall
Foster Cohen et al. [29]	105 VPT (stay in NICU): 19 with moderate or severe and 60 with mild white matter abnormality. Mean GA: 27.8 (2.3)	48 months			CELF-P
	107 FT children				Overall
	Exclusion: congenital		16%	8.6%	Mild LD: >−1 SD below the mean of the FT group
	abnormalities and non-English speaking parents.		15%	6.7%*	Moderate or severe LD: <−1,5 SD
	SES and maternal education of FT significantly higher.		31%	15.3%	Total LD
Sansavini et al. [30]	64 very preterm. Mean GA: 30.4 (2.1) without serious complications longitudinally followed.	30 months	24.1% (CDI)	13.6%	Italian CDI total word prod. 10th percentile
			16.1% (MLU)	9.1%	PRF MLU < −1.25 SD
	FT group: 22 / 40 (at age 30 and 42 months)	42 months	34.4% (MLU)	7.5% *	PRF MLU < −1.25 SD

Table 1. *Cont.*

References (Ordered by Publication Year)	FT and PT Group Characteristics	Age of Assessment	% of Delayed PT Children	% of Delayed FT Children	Language Measure and Classification Criteria
Sansavini et. al. [31]	104 VPT children. Mean GA: 29.5 (2.1) weeks 20 FT children	24 months	20%	5% *	Italian CDI short form word product. <10th percentile
Charollais et al. [32]	117 VPT children GA range: 25–32 weeks. Stay in NICU. No medical or demographic information. Control: Normative sample, N = 385	24 months	41%	10% *	French CDI short form. Word production. <10th percentile.
Stolt et al. [33]	141 VPT. Mean GA: 28 (3) weeks. 146 FT longitudinally followed	24 months			Finnish CDI
			18%	9%*	Word prod.
			18%	8%*	MLU3
			16%	10%	Bayley-II expressive language
	120–131 VPT 123–137 FT Similar maternal education	60 months	27%	10% *	Nepsy lang.
			20%	10% *	5 to 15 lang scores.
					<10th percentile of the FT group
Lee and Lee [21]	86 PT Korean children. Mean GA = 30.45 weeks. Median stay at NICU: 34 days.	10 to 30 months			Korean CDI word prod.
			46.5%	0% *	word prod
					SELSI:
			43%	0% *	Combined
			50%	0% *	Expressive
			34.9%	0%	Receptive
					10th percentile
Do et al. [34]	184 Vietnamese PT children. Mean GA = 31.6 (2.5).	24 months	8%	0% *	BSID-III: Language Composite Score.
	Mean stay in NICU of 34 days.				<−2 SD
	Control PT: normative sample of 78 children (no sociodemographic information)				

Notes: * Significant differences between delay percentages of the two groups. $ Significant differences between delay percentages of the PT group with BPD and the other two (FT and PT without BPD). BDI: Batelle Developmental Inventories. BSID-III: Bayley Scales of Infant Development-III. CELF-P: Clinical Evaluation of Language Fundamentals- Preschool. CDI: Communication Development Inventories-MacArthur-Bates scales. MLU3: Mean Length of Utterances of the three longest utterances (obtained with the CDI). NICU: Neonatal Intensive Care Unit. PLS-3: Preschool Language Scale-3 (UK), which comprises Auditory Comprehension and Expressive Communication scales. PRF MLU: Prova di Repetizione di Frasi (Italian Test of sentence repetition) Mean Length of Utterances. SELSI: Sequenced Language Scale for Infant. 5 to 15 language: Five to Fifteen Language.

From Table 1, it is clear that there is a great disparity in the estimates of the percentage of children with language delay. The range of prevalence of LD among PT children goes from 8% to 49%, while that for FT children goes from 0% to 30%. This great variation occurs even though the PT children were of similar GA. Practically all (not that of Do and collaborators [34]) of the samples were of VPT and EPT children, with a mean GA of in between 28 and 30 weeks, with the exception of the study by Wolke and his collaborators [28], in which the EPT children were born below 26 weeks of GA. All the samples were of VPT and EPT children, although in some cases this was termed as extremely low birth weight children (ELBW).

The important variation in the percentages found of children with language delay or language impairment could be related to different factors, such as the age of assessment, the instruments used in each study, the cut-off criteria and the characteristic of the participants, particularly the PT children.

The age of assessment varies greatly among the studies, ranging from below 24 months to 72 months of age, although there are many data for similar ages (namely, at 24 and 48 months of age) which have provided diverging results. It is logical to think that the proportion of LD children may change with age, thus introducing differences in the results of the investigations.

The instruments used are not the same in all the studies, which is comprehensible given the differences in the ages of assessment. In many of them different adaptations of the MacArthur-Bates Communicative Development Inventories (CDI) to different languages were used to assess children of 30 months of age or younger [21,30–33]. In these cases, lack of agreement in the results indicates that other factors must be responsible for the discrepancy.

The cut-off criteria to identify LD vary among the different studies. In many cases the 10th percentile of the normative sample of the instrument as reference has been used as criterium to determine the limit of LD both for the PT and the FT groups. In other cases, the limit was a certain point of SD in relation to the normative sample. Two investigations, however, when establishing the cut-off point did not use the normative sample of the test as a reference but used the FT control group instead [29,33]. This fact shed some doubts upon the adequacy of the comparisons just in case the participants chosen for the FT group might have had a higher performance not coincident with the norms. In other cases, other cut-off criteria were used (scores below −1 SD, below −1.25 SD, below −1.5 SD, below −2 SD, or below developmental quotient (DQ) 85), which creates classification criteria that are more or less stringent, resulting in lower or higher percentages of LD children, respectively.

Finally, the characteristics of the participants in the different studies may also be a source of variability in the results found. It is certainly true that the PT participants of most of the studies were VPT and EPT children around 28–30 weeks of mean GA. However, the selection criteria changed a lot among the revised studies. Some studies established clear exclusion criteria, which are quite strict, and children with serious biomedical complications were not included in the premature group [30]. In these cases, children with major cerebral damage, such as IVH higher than II or PLM, hydrocephalus, BPD, retinopathy of prematurity, visual or hearing impairment or congenital malformations were excluded. In other cases, the criteria were less strict, and only children with some of these criteria were excluded: congenital abnormalities, chromosomal anomalies, coming from homes where the language of the community was not spoken, admission to the Neonatal Intensive Care Unit (NICU), or mother's use of alcohol or drugs during pregnancy [29,33] were excluded. Other studies do not offer information on the exclusion criteria (Charolais et al., 2014; Lee and Lee, 2016), which does not guaranty (all the contrary) that the sample of PT children are free of these biomedical hazards (IVH, PLM, BPD, etc.). Two studies directly chose participants who were in the NICU for a long stay [32,34]. Another study [27], which used strict exclusion criteria, however, also included children with BPD in the PT group because one aim of the study was to test the effect of this disease on the risk of suffering language delay. It seems reasonable to think that these differences in the inclusion/exclusion criteria may have important consequences in the differences found regarding the prevalence of language delay. It is enlightening that when Stolt et al. [33] compared only VPT children without neurological damage with FT children, the differences in percentages of LD children are lower, and no significant differences in LD were found in the tasks administered, with the exception of the Nepsy language score at 5 years of age. Complementarily, when VPR children have additional handicaps (such as BPD) they show a higher incidence of language impairment, which may rise to 43% [27,30].

The comparability of the FT and PT groups in certain critical characteristics, such as similarity of maternal education, Socio Economic Status (SES), or balanced gender distribution, is a key point, which is fulfilled by several investigations [27,30,31]. A few studies do not provide information in this regard or not enough information [21,34], and others clearly do not fulfil these requirements, and the PT group is composed of children whose mothers have lower education and/or SES than those of the FT group [5,29]. All these introduce serious doubts on the interpretation of the results found, because the FT and PT groups were not comparable, which introduces a threat to the internal validity of the investigations. On some occasions there was no control group of FT children; instead, the normative sample of a given test was used as the comparison group [32,34].

Only two studies adopted a longitudinal perspective, with repeated measures for the participants [30,33], although in one of them [30] one longitudinal sample of PT children is compared to two different cross-sectional FT samples at different ages (30 and 42 months). This increases the variability of the two samples of FT children, which may differ in their characteristics. Therefore, intraindividual patterns of change cannot be observed for FT children, which is a limit for the accurate longitudinal comparisons of the PT and FT groups.

The results of the control groups are really very unusual in some of the studies. There is always a certain percentage of FT children who are below the cut-off criteria to define language delay (or language impairment for older children), which is usually over 7% [35,36]. However, there are two studies in which the percentage of FT children below percentile 10 or below −2 SD is 0% [21,34]. In the case of Lee and Lee [21] chronological age (not the corrected age) was used for comparisons, and this fact can explain the unusual gap between PT and FT children. In addition, as mentioned before, in these two studies [21,34] no information is provided on the similarities of the PT and FT samples in sociodemographic characteristics (e.g., parental educational level or SES) or gender, which are important characteristics to ascertain that both samples are comparable.

Most of the reviewed studies were conducted with VPR or EPR children, who are considered to be at higher risk of suffering developmental problems than other populations of preterm children, such as moderately or late preterm. In addition, in an important number of the studies carried out, the VPT/EPT children have other associated medical problems (neurological damage, BPD) or risk situations (stay in the NICU for a relatively long time). For this reason, it is not a surprise that the incidence of language delay or language impairment of VPT and EPT children clearly exceeds that of FT children. This population represents around 20% of the total population of PT children [2], which provides a reason to extend studies on the prevalence of LD in PT children to other segments of the total population of PT children in order to get a wider panorama of what happens with the PT population.

Therefore, and this is a purpose of the present study, there is a need to study a sample of PT children with a relatively wide range of GA, and with no serious biomedical complications. On the other hand, there is a dearth of studies carried out with a longitudinal design. One important advantage of longitudinal studies, apart from the description of intraindividual change, is that they allow us to investigate the predictive effect of different factors on the determination of language delay. In this research a longitudinal follow up of 3 groups of PT children with different GAs and one group of FT children will be carried out.

In relation to the most relevant predictive factors of language delay or language impairment, previous research has highlighted a variety of biomedical, environmental and psychological factors, which will be briefly reported on as follows.

Among biomedical factors, gestational age or birth weight were found to have a predictive effect on language delay, and the risk of suffering language impairments [9,10,23,37,38]. Other authors, however, have suggested that neurobehavioral outcomes at an early school age can be predicted based on IVH incidence as opposed to birth weight or GA [39]. Neurological impairment (IVH, PLM), on its own, or in association with other factors has also had an important effect on language delay [5,29,40]. IVH higher than II, but not lower, has been found to have negative effects on cognitive and language measures [41]. Bronchopulmonary Dysplasia seems to have a very detrimental effect on the possibility of PT children having language delay [27,30,31,42–44]. Male gender has been found to increase the risk of language delay [28,31,35,36], and family history of language or learning disorders predicted lower language development [15].

Several studies found an influence of environmental factors on language delay such as the level of maternal or parental education [19,20,31,34,36,45–48], the SES [49], a combination of these two factors [40], and the quality of home environment [50].

Finally, among the psychological or personal factors, previous cognitive development [15,29,31,51,52], previous use of gestures [31,53–55], and previous language abilities [30,33,56,57] are good predictors of later linguistic development.

There is a lack of information, however, on the prevalence of language delay in low-risk PR children, and on whether this rate increases as children grow older.

One major strength of the present research is that there is a longitudinal follow up of four groups of children with different gestational ages, covering a range from extremely preterm to full term children (GA 26–41 weeks). Another strong point is that the PT children do not have major medical complications, being considered healthy or low-risk children, a group which paradoxically has been scarcely studied despite that they constitute the majority of newly born PT children.

The aims of the present research are the following:

(1) To compare the prevalence of language delay in healthy preterm children (PR) with different GAs to that of full-term children (FT) in a longitudinal design ranging from 10 to 60 months of age.
(2) To analyze which variables are related to a higher risk of language delay at 22, 30 and 60 months of age.

2. Materials and Methods

2.1. Participants

One group of FT and another of PT children were recruited at birth in four hospitals in Galicia (Spain) and longitudinally followed and assessed at different points in time.

The initial participants of the PT group were 151 PT children (with GA range between 26 and 36 weeks), and those children with the following characteristics were excluded: cerebral palsy (as diagnosed up until 9 months of age), periventricular leukomalacia (PLM), intraventricular hemorrhage (IVH) greater than grade II, hydrocephalus, encephalopathy, genetic malformations, chromosomal syndromes and metabolic syndromes associated with mental retardation, important motor or sensorial impairments, and Apgar scores below 6 at 5 min. The initial participants of the FT group were 49 children with standard GA and no evidence of impairment. The children were assessed at 10, 22, 30, 48 and 60 months of age.

The number of participants and their distribution by GA groups at every assessment point is displayed in Table 2. The participants were distributed into four groups according to their GA.

The PT and FT groups did not differ in terms of mother's education (X^2 (1) = 8.66, p = 0.194), gender (X^2 (1) = 0.000, p = 0.997) or Apgar score (t (197) = -0.909, p= 0.365).

The characteristics of the samples remained similar throughout time.

Table 2. Composition of the sample throughout time.

Age	GA ≥ 37 (%)	GA 36–34 (%)	GA 33–32 (%)	GA ≤ 31 (%)
15 days	49	65	37	49
10 months	49	65	37	49
22 months	43	58	36	43
30 months	37	48	32	37
48 months	34	42	33 *	36
60 months	33	42	31	34

Note: * One child not tested at 30 months was tested at 48 months of age.

2.2. Instruments

To assess language development the Inventario do Desenvolvemento de Habilidades Comunicativas (IDHC) Palabras e xestos (Words and Gestures) and Palabras e Oracións (Words and sentences) [58,59] were filled in by the children's mothers when the children were 10, 22 and 30 months of age (see below). The IDHC is the Galician version of the MacArthur-Bates Communicative Development Inventories (CDI) [60]. The IDHC Words and Gestures for children between 8 and 15 months of age, was applied when the children were 10 months old. The form Words and Sentences, for children aged between 16 and 30 months was applied when the children were 22 and 30 months of age. The following measures were taken into consideration for the present study. At 10 months of age, word understanding and first communicative gestures, which will be considered as predictive factors in regression analyses. At 22 and 30 months of age, word production, which is considered the central feature identifying language delayed children [51].

When the children were 48 months of age, they were assessed through the Reynell Developmental Language Scales [61]. The RDLS is comprised of two scales: expressive and comprehension language scales. Because of the deficient adaptation of the RDLS into Spanish (no Spanish norms exist, and no adaptation to the characteristics of Spanish language acquisition has been made), only the total raw score in comprehension was used in the analyses performed.

The children's language development was also assessed when the children were 60 months of age through the Peaboby Picture Vocabulary Test 3rd edition (PPVT-III) [62], the test Comprensión de Estructuras Gramaticales (CEG) [63], and the production scale of the Test de Sintaxis de Aguado (TSA) [64].

The widely known PPVT-III was used to assess vocabulary comprehension. The child is required to point to the image that best matches the word pronounced by the researcher out of the 4 pictures on the page. The words that are tested are arranged in order of increasing difficulty.

The CEG was used to assess the comprehension of syntactic structures. The CEG is a Spanish test that is very similar to the well-known Test of Reception of Grammar (TROG-2) [65]. The CEG consists of 80 pages that include four pictures on each page. In each item, the researcher pronounces a sentence (e.g., "El niño que mira a la niña está comiendo": The boy who looks at the girl is eating) and the child points to the image that matches the target sentence. The other three images act as (lexical or grammatical) distractors. The CEG explores 20 different syntactic structures organized into blocks with 4 items each. The CEG can be administered to children from 4 to 12 years of age.

The production subscale of the TSA [64] was used to assess morphological and syntactic production skills. In this test, the child has to imitate a sentence previously produced by the researcher looking at a drawing related to the sentence "what did I say about this drawing?". Thirty items follow this pattern, and in another four items the child has to complete the last part of a sentence given a conversational context created by the researcher ("cuando hace frío . . . .me pongo el abrigo" "when it is cold . . . .I put on my coat"/"si hiciera calor . . . .no me pondría el abrigo" "if it were warm . . . I would not put on my coat"). The TSA explores the production of different morphosyntactic abilities: interrogative sentences, negative sentences, passives, use of possessive, relative, interrogative, possessive and demonstrative pronouns, complex sentences, comparisons, use of prepositions, use of different persons and times in verbs, etc.). The TSA can be administered to children from 3 to 7 years of age.

The cognitive development of the children was assessed through the Batelle Developmental Inventory (BDI) [66] when they were 22 months of age. This scale measures a child's progress in development and in discrete skill sets. The skills assessed by the Batelle scale are adaptive, personal-social, communication, motor, and cognitive. The cognitive score was used for the present analyses.

The mothers of the children completed an interview at the beginning of the study that included socio-demographic information of the family, information on pregnancy, Apgar scores, feeding and health habits, educational level of the parents, etc.

The children lived in a bilingual Spanish-Galician community context which makes it possible to use Spanish or Galician tests. The Galician tests (IDHC) were administered to the mothers of the children. The rest of the tests in Spanish were administered to the children. No adaptations of these tests exist for Galician.

2.3. Procedure

Previous consent from the mothers was obtained, as well as the acceptance by the Comité Ético de Investigación Clínica de Galicia.

The children's communicative and linguistic development was assessed at 10, 22, 30, 48 and 60 months of age (±15 days), with corrected age for the PT group up until 30 months of age but not later.

The parent reports (IDHC) were filled in by the mothers. The remaining tests were administered by a trained psychologist at the specified ages in the children's homes.

The following measures were taken into consideration for the present study. At 10 months of age, word understanding, word production, and first communicative gestures were considered. At 22 and 30 months of age, word production scores were taken into account. We used this measure because at this age word production is the most reliable indicator of language development.

2.4. Analyses Performed

The following analyses were performed.

1. ANOVA for mean comparisons between the results of the PR and FT children at different ages in different measurements.
2. Chi square comparisons between the four GA groups of children and also between the PT and the FT groups regarding the relative percentages of children with and without language delay. Those children with raw scores lower than percentile 10 were considered to have language delay. This criterion, however, was changed in the case of cognitive development [66] at 22 months of age. In this case we have adopted the threshold of percentile 15, because the norms offer percentiles for a range between 18 and 23 months of age, and the children were at the upper limit of the age range.
3. Five logistic regression analyses (enter method) were performed in order to test the predictors of language delay as measured through the different instruments at different ages (dependent variables DV). Previously, the effects of many different variables were tested, and only those which had an effect on the DVs were incorporated in the final analyses, as well as 3 variables of theoretical relevance: gestational age (numerical), gender and maternal education level (three groups: low, medium and high). Among those variables which did not have any effect on any DV were: Apgar score in the 1st minute (risk/no risk = ≥7)), stay in the NICU (1 = no stay, 2 = 1–15 days, 3 = >15 days), family antecedents of language problems (yes/no), mother's age at birth (risk/no risk), risk of maternal depression (yes/no), parental stress (risk/no risk), HOME score (quality of home environment). In addition, the absence of effect of some of them on language risk/delay has been demonstrated in a

preliminary research [67]. These variables were not included in the regression models, and, therefore, no information is offered on them for brevity's sake.

The logistic regression analyses were carried out with all the participants, because the number of FT children was not large enough to perform separate analyses for PT and FT children, and for the sake of the strength of the tests.

In the first logistic regression model, the dependent variable (DV) was children with or without lexical delay (word production) at 22 months of age. The predictive variables introduced were: gestational age in weeks, gender, maternal education, total score in first communicative gestures at 10 months (IDHC) and total score in vocabulary understanding at 10 months of age (IDHC).

In the second logistic regression model, the dependent variable (DV) was children with or without lexical delay at 30 months and the Predictors were those factors which in previous logistic regression analyses had had a significant effect on the DV or theoretical relevance: gestational age, gender, maternal education, risk of cognitive delay at 22 months (BDI), and risk of vocabulary delay (word production) at 22 months of age (IDHC).

In the following three logistic regression analyses the Dependent variables were children with or without language delay at 60 months of age. The threshold was percentile 10 in the the PPVT-III (vocabulary comprehension), CEG (grammar understanding), and the TSA (morphosyntactic production) in each analysis. The predictive variables were always the same for these three analyses. The Predictors were those factors which in previous logistic regression analyses had had a significant effect on the DV or theoretical relevance: gestational age, gender, maternal education, risk of cognitive delay at 22 months, risk of vocabulary delay at 30 months, and total score in the comprehensions scale of the RDLS at 48 months of age.

3. Results

3.1. Descriptive Results and Comparisons between Groups

The results of the one factor ANOVA to compare the GA groups in different measures is offered in Table 3. As can be observed, there were no significant differences between the groups in any measure, with the exception of the CEG at 60 months of age ($p < 0.05$). In this case, the differences were due to the differences between the GA $\leq$ 37 week group and the GA 36–34 week group (Bonferroni post hoc $p < 0.05$).

Table 3. Mean (SD) scores and ANOVA comparisons between the four gestational age (GA) groups.

Task (Age)	GA ≥ 37 Mean (SD)	GA36–34 Mean (SD)	GA33–32 Mean (SD)	GA ≤ 31 Mean (SD)	F	df	*p*
First Gestures (10 months)	7.5 (2.5)	7.2 (2.4)	7.4 (2.7)	6.5 (2.7)	1.262	190	0.289
Comprehension of words (10 months)	71.8 (58.8)	88 (77.2)	71.5 (70)	73.3 (73.3)	0.694	190	0.557
Cognition BDI (22 months)	27.5 (4)	26.7 (3.7)	26.8 (3.2)	26.5 (2.9)	0.743	180	0.528
Word Production (22 months)	173.7 (137.1)	174.5 (163.8)	154.2 (130.1)	140.9 (137.8)	0.573	179	0.633
Word Production (30 months)	411.4 (171.3)	412.58 (189.7)	431.00 (149.2)	408.05 (181.5)	0.116	153	0.951
Comprehension RDLS (48 months)	46.5 (5.2)	43.1 (8.6)	43 (4.7)	44.2 (5.9)	2.230	144	0.087
PPVT (60 months)	62 (12.5)	57.8 (11.4)	57 (12.1)	56.2 (13)	1.460	141	0.228
CEG (60 months)	52 (7.3)	44.4 (13.8)	47.1 (9.1)	48.6 (13.4)	2.804	139	0.042
TSA Production (60 months)	43.5 (8.1)	38.4 (15.8)	41.5 (11.2)	39.7 (15.9)	1.034 *	116,539	0.380

Note: * Brown–Forsythe test; F = value of F-statistic; df = degrees of freedom; *p* = significance value.

3.2. Language Delay Comparisons between Groups

The percentages of children in each GA group with language delay (LD) as assessed through different measures taken at different ages are presented in Table 4. This table indicates the number and relative percentage of children of each GA group who got a score below percentile 10 in the tests applied at different ages. Those children are considered as part of the language delay/language impairment group. In addition, the results of the chi squared test are also presented.

Table 4. Frequency and (percentage) of children with language delay at 22, 30 and 60 months of age (<10th percentile), and GA group comparisons.

Assessment (Age)	GA ≥ 37 (%)	GA 36–34 (%)	GA 33–32 (%)	GA ≤ 31 (%)	X^2	*p*
Word Production (22 months)	8 (18.6)	15 (25.9)	6 (16.7)	14 (32.6)	3.595	0.309
Word Production (30 months)	7 (18.9)	8 (16.7)	4 (12.5)	9 (24.3)	1.720	0.632
PPVT (60 months)	1 (3)	1 (2.4)	2 (6.3)	4 (11.4)	3.490	0.322
CEG (60 months)	5 (15.2)	17 (40.5)	12 (38.7)	6 (17.6)	9.378	0.025
TSA Production (60 months)	7 (16.7)	15 (35.7)	12 (28.6)	8 (19.0)	3.608	0.307

X^2 = Chi square value; *p* = significance value.

In general terms, there are no significant differences in the proportion of children with language delay/impairment among the four groups, with the exception of the results with the CEG (grammar comprehension) ($X^2 = 9.378$, $p < 0.05$), in which the GA groups 36–34 and 33–32 weeks clearly have a very high percentage of children with LD (40.5% and 38.7%, respectively), while the groups with children having a GA of 37 weeks or above and with a GA of 31 weeks or below have lower (and quite similar) proportions of children with LD (15.2% and 17.6%, respectively).

In order to make the results more manageable and to make the comparisons clearer, we have integrated the results of all the GA groups below 37 weeks in a group of preterm infants. These results are presented in Table 5.

Table 5. Frequency and (percentage) of PT and FT children with language delay at 22, 30 and 60 months of age (<10th percentile), and with of cognitive delay (<15th percentile) at 22 months of age and comparisons between groups.

Risk of Delay	FT (%)	PT (%)	X^2	*p*
Word Production (22 months of age)	8 (18.6)	35 (25.5)	0.868	0.352
Word Production (30 months of age)	7 (18.9)	21 (17.9)	0.018	0.894
PPVT (60 months of age)	1 (3.0)	7 (6.4)	0.548	0.459
CEG (60 months of age)	5 (15.2)	35 (32.7)	3.810	0.051
TSA Production (60 months of age)	7 (21.9)	35 (33.0)	1.442	0.230
(BDI) Cognitive delay (22 months of age)	6 (14.0)	22 (15.9)	0.099	0.753

X^2 = Chi square value; *p* = significance value.

In this case, the difference between the PT and the FT groups in the CEG is only marginally significant ($X^2 = 3.810$, $p = 0.051$). The rest of the comparisons do not reach significance. There are no significant differences between FT and PT children in vocabulary production at 22 and 30 months of age. Nor there are differences in receptive vocabulary (PPVT-III) or morphosyntactic production (TSA). The frequency of children with delay in the PPVT-III is clearly lower than in the rest of the tests, although the percentage of PT (6.4%) children with LD is double the percentage of FT (3.0%) children.

Table 5 also displays the percentage of PT and FT children with cognitive delay measured at 22 months of age through the BDI, because this score will be used in logistic regression analyses. No significant difference between FT and PT children is found in this regard.

3.3. Logistic Regression Analyses

The following tables display the results of the logistic regression analyses performed.

In Table 6, the results of the logistic regression for delay/not delay in word production at 22 months of age (IDHC Words and Sentences) as the dependent variable are presented. Out of the predictors introduced in the model, only the total score of first communicative gestures at 10 months of age ($p = 0.013$), gender ($p = 0.032$) and the total score in word comprehension at 10 months of age ($p = 0.046$), in this order, were found to have a significant effect on the variance of having or not having language delay as measured through word production at 22 months of age. The variance explained by the model is moderate (Negalkerkes's R^2 = 0.153). The model reaches significance (Hosmer– Lemeshow's X^2 (8) = 9162, $p > 0.329$; X^2 (6) = 19.348, $p = 0.002$; −2LL = 178.578), and correctly classifies 78.9% of the participants (specificity: 97.8, sensitivity = 18.6).

Table 6. Logistic regression analysis: predictors of language delay (LD) in word production at 22 months (IDHC).

Variables	B	SE	Wald's X^2	*p*	OR	95% CI
Gestational Age	−0.057	0.051	1.244	0.265	0.945	0.856–1.044
Gender	0.838	0.391	4.593	0.032	2.312	1.074–4.975
Maternal education	−0.420	0.238	3.130	0.077	0.657	0.412–1.046
Total first gestures 10 months	−0.198	0.080	6.148	0.013	0.820	0.701–0.959
Total vocabulary understanding 10 months	0.005	0.003	3.988	0.046	1.005	1.000–1.011

B = Unstandardized regression weight; SE = Standard error for the unstandardized B; *p* = Significance value; OR = Odds ratio; 95% CI = Confidence interval of the odds ratio.

In Table 7, the results of the logistic regression for delay in word production at 30 months of age (IDHC Words and Sentences) as the dependent variable are presented. The only significant predictors found are risk of vocabulary delay ($p = 0.000$) and risk of cognitive delay ($p = 0.038$) at 22 months of age. The variance explained by the model is 34% (Negalkerkes's R^2 = 0.344). The model reaches significance (Hosmer–Lemeshow's X^2 (8) = 9005, $p > 0.342$; (X^2 (6) = 36.341, $p = 0.000$; −2LL = 109.292). The model correctly classified 83.7% of the participants (specificity: 95.2, sensitivity = 32.7).

Table 7. Logistic regression analysis: predictors of LD in word production at 30 months of age.

Variables	*B*	SE	Wald's X^2	*p*	OR	95% CI
Gestational Age	0.041	0.064	0.408	0.523	1.042	0.919–1.180
Gender	0.350	0.526	0.444	0.505	1.420	0.506–3.980
Maternal education	−0.449	0.326	1.899	0.168	0.638	0.337–1.209
Cognitive delay 22 months	1.220	0.589	4.284	0.038	3.386	1.067–10.746
Vocabulary delay 22 months	2.165	0.512	17.888	0.000	8.712	3.195–23.754

B = Unstandardized regression weight; SE = Standard error for the unstandardized B; *p* = Significance value; OR = Odds ratio; 95% CI = Confidence interval of the odds ratio.

Table 8 shows the results of the logistic regression analysis for delay in lexical comprehension (PPVT-III) as the dependent variable. In this case, the analysis has to be interpreted with caution, since the frequency of those children with scores under percentile 10 are only 8. The predictive variables which have a significant effect are risk of vocabulary delay at 30 months of age (IDHC-Words and Sentences) ($p = 0.022$) and the total score in language comprehension (RDLS) at 48 months of age ($p = 0.046$). Maternal education has a nearly significant effect ($p = 0.056$). The variance explained by the model reaches 41% (Negalkerke's R^2 = 0.414). The model reaches significance (Hosmer–Lemeshow's X^2 (8) = 9160, $p = 0.329$; X^2 (6) = 18.143, $p = 0.006$; −2LL = 30.765). The model correctly classified 96.2% of the participants (specificity: 99.2, sensitivity = 33.3).

In Table 9, the results of the logistic regression for delay in syntactic understanding (CEG) at 60 months of age as the dependent variable are presented. The only significant predictors found are total score in language comprehension (RDLS) at 48 months of age ($p = 0.009$) and risk of cognitive delay at 22 months of age ($p = 0.012$). The variance explained by the model is 27% (Negalkerkes's R^2 = 0.278). The model reaches significance (Hosmer–Lemeshow's X^2 (8) = 4793, $p = 0.779$; X^2 (6) = 27.853, $p = 0.000$; −2LL = 124.838). The model correctly classified 80.3% of the participants (specificity: 96.9, sensitivity = 34.3).

Table 8. Logistic regression analysis: predictors of vocabulary comprehension delay (PPVT-III) at 60 months of age.

Variables	*B*	SE	Wald's X^2	*p*	OR	95% CI
Gestational Age	−0.231	0.172	1.789	0.181	0.794	0.566–1.113
Gender	−1.344	1.300	1.069	0.301	0.261	0.020–3.332
Maternal education	2.046	1.072	3.642	0.056	7.737	0.946–63.253
Cognitive delay 22 months	0.775	1.622	0.228	0.633	2.171	0.090–52.112
Vocabulary delay 30 m.	2.953	1.289	5.247	0.022	19.172	1.532–239.988
Total comprehension score RDLS	−0.165	0.083	3.969	0.046	0.848	0.721–0.997

B = Unstandardized regression weight; SE = Standard error for the unstandardized B; *p* = Significance value; OR = Odds ratio; 95% CI = Confidence interval of the odds ratio.

Table 9. Logistic regression analysis: predictors of grammar understanding delay (CEG) at 60 months of age.

Variables	*B*	SE	Wald's X^2	*p*	OR	95% CI
estational Age	−0.015	0.063	0.059	0.808	0.985	0.871–1.114
Gender	−0.363	0.467	0.603	0.437	0.696	0.278–1.738
Maternal education	−0.835	0.314	7.061	0.008	0.434	0.234–0.803
Cognitive delay 22 months	1.780	0.705	6.375	0.012	5.929	1.489–23.608
Vocabulary delay 30 months	−0.040	0.633	0.004	0.950	0.961	0.278–3.323
Total comprehension score RDLS	−0.116	0.044	6.905	0.009	0.891	0.817–0.971

B = Unstandardized regression weight; SE = Standard error for the unstandardized B; *p* = Significance value; OR = Odds ratio; 95% CI = Confidence interval of the odds ratio.

Finally, Table 10 shows the results of the logistic regression analysis for delay in morphosyntactic production (TSA) at 60 months of age as the dependent variable. The predictors which reach significance are the total score in language comprehension (RDLS) at 48 months of age ($p = 0.003$) and risk of vocabulary delay at 30 months of age ($p = 0.042$). The model explains 17% of the variance (Negalkerke's $R^2 = 0.176$). The model reaches significance (Hosmer–Lemeshow's X^2 (8) = 4671, $p = 0.792$; X^2 (6) = 17.472, $p = 0.008$; −2LL = 145.351). The model correctly classified 69.5% of the participants (specificity: 91.1, sensitivity = 22.0).

Table 10. Logistic regression analysis: predictors of morphosyntactic production (TSA) delay at 60 months of age.

Variables	*B*	SE	Wald's X^2	*p*	OR	95% CI
Gestational Age	−0.001	0.057	0.000	0.983	0.999	0.892–1.118
Gender	0.249	0.421	0.351	0.554	1.283	0.563–2.925
Maternal education	0.101	0.281	0.130	0.719	1.106	0.638–1.917
Cognitive delay 22 m	−0.209	0.658	0.101	0.750	0.811	0.223–2.946
Vocabulary delay 30 m	1.091	0.536	4.148	0.042	2977	1.042–8.508
Total comprehension score RDLS	−0.128	0.043	9.095	0.003	0.880	0.809–0.956

B = Unstandardized regression weight; SE = Standard error for the unstandardized B; *p* = Significance value; OR = Odds ratio; 95% CI = Confidence interval of the odds ratio.

4. Discussion

In relation to the first aim of the study which was to compare the prevalence of language delay in healthy preterm children (PR) with different GAs to that of full-term children (FT), the results found indicate that there are no significant differences in the percentage of children with language delay among the four GA groups in the following language measures: Word production at 22 and 30 months of age as measured through the Galician CDI, word comprehension at 60 months of age as measured through the PPVT, morphosyntactic production at 60 months of age as measured through the TSA. The only significant difference was found in grammatical structures comprehension ($p < 0.025$), measured through the CEG. The greatest differences occurred between the GA groups of 36–34 and 33–32 weeks (with 40.5% and 38.7% of LD respectively) and the other two groups (FT and VPT-EPT, with 15.2% and 17.6%, respectively). This result is coincident with that found in the ANOVA (Table 3), in which the significance was explained by the difference between the groups GA $\geq$ 37 and GA 36–34 weeks. Unexpectedly, the difference was not due to the difference between the most distant groups ($\geq$31 and $\geq$37 weeks), that is to say the VPT-EPT and the FT groups. Therefore, the GA factor does not seem to explain these results, contrary to other authors' claims [9,10,37]. This conclusion will be confirmed later with the regression analyses and should be interpreted taking into consideration the low-risk characteristic of the sample.

When the results of all the PT children (GA < 37) are put together, the comparison is simpler and, again, the results indicate no significant differences in the language measures taken. Even in the test of comprehension of grammatical structures (CEG), administered at the age of 60 months, the difference in this case does not reach significance, although it is really very close (p = 0.051).

In general terms, the percentage of FT children with LD throughout time, using different tests, remains quite stable (with the exception of the PPVT-III results) in a range between 15.2% and 21.9%. In contrast, the percentages of children with LD in the PT group vary much more over time, in a range between 17.9% and 33%, and there is not a clear incremental trend in the percentage of children with language delay from early years to 5 years of age, as several authors have proposed [30,33,37]. It is obvious that using different tests with different norms makes comparisons throughout time difficult to carry out because variability can be caused not only by changes in the participant, but also by variations in the norming process. Therefore, the results must be taken with caution.

One factor that seems to increase the risk of undergoing language delay in PR children is the existence of medical complications (neurological or pulmonary) [27,33]. When these children were excluded, the rate of language delay of the PR children descended. Probably, the fact that our sample was practically free of children with these medical complications may have affected the results found in the PR group. One additional argument in favor of this idea is that those investigations which included relatively high percentages of PR children with neurological or pulmonary medical problems evidenced very high rates of language delay for PR children [27,29,42,43].

In relation with the second aim, which was to identify those variables related to a higher risk of language delay at 22, 30 and 60 months of age, the results found in the logistic regression analyses permit identification of different predictive factors, which vary according to the moment of assessment as well as the different linguistic abilities.

Three factors were found to have an effect on the probability of suffering from language (lexical) delay at 22 months of age (Table 6): Gender, use of first gestures at 10 months, and total vocabulary understanding at 10 months. Gender reached significance ($p < 0.05$, OR = 2.312), with boys having a higher risk of language delay than girls (more than twice as high). This result is in agreement with other studies with PT and FT children of a similar age [11,13,31] and older [28,35], and does not support the results found by other studies [15] which found practically no effect of gender on word production in children of 30 months of age, or at 24 and 60 months of age [33].

The use of first gestures (total score) also had a significant effect on language delay at 22 months of age ($p < 0.05$, OR = 0.820), indicating that those children with a lower number of gestures at 10 months of age have a higher probability of being language delayed at 22 months of age, which agrees with former investigations carried out with children of similar ages [31,53,55]. Therefore, this result confirms that the use of gestures seems to be a possible predictor of language development in the short term.

Finally, the third factor which has been found to have a significant effect on lexical delay at 22 months of age was word comprehension at 10 months of age ($p < 0.05$, OR = 1.005). In any case, this effect was very reduced (OR = 1.005), and apparently paradoxical (see 33) and contrary to expectations.

In general terms, the logistic regression model for word production at 22 months of age correctly classifies 79% of the participants, even though the sensitivity is low; this means the classification of the participants in the delayed group is not good, with a high proportion of false negatives (children who are not classified as language delayed although they are language delayed).

In relation to the prediction of language delay at 30 months of age (Table 7), two risk factors seem to have a significant effect: cognitive delay at 22 months and productive vocabulary delay at 22 months of age. Vocabulary delay at 22 months has an important impact on later language (lexical) delay ($p < 0.001$, OR = 8.712), indicating that those children with lexical delay at 22 months of age have many more possibilities of suffering from language delay at 30 months of age. Cognitive delay at 22 months also has a significant (although somewhat lower) effect on the probability of suffering language delay at 30 months ($p < 0.05$; OR = 3.386). These results are in tune with those found in other studies which have claimed that previous linguistic [30,33,57] and cognitive development [11,15,29,51,52] are good predictors of later language delay.

The model correctly classifies 84% of the participants in the two groups of language delayed and not language delayed, with a high specificity (95.2) even though the sensitivity is low (32.1), indicating that, again, there is a high percentage of false negatives.

Other factors which were found to have a significant effect on early language delay, such as gestational age [9,10,23,37], or maternal education [19,20,33,34,36,45–48] did not reach significance at either 22 or 30 months of age. Other authors [15], however, found that low parental education level quite unexpectedly did not affect child linguistic outcomes at the age of 36 months.

The logistic regression analyses performed when the participants were 60 months old, give interesting results which point to the effect of previous language, cognitive delay and maternal education level.

The results obtained in the regression analysis with vocabulary comprehension (PPVT-III) at 60 months of age as dependent variable must be taken with caution, because of the low number of children who scored below percentile 10 (8 in all). In this case, those children who were language delayed (vocabulary production) at 30 months of age have a much greater probability of being language delayed (receptive vocabulary) 30 months later ($p < 0.05$, OR = 19.172). Those children who got low scores in language comprehension (RDLS) at 48 months of age have also got a greater probability of being in the group of language delayed children (receptive vocabulary) at 60 months of age ($p < 0.05$ OR = 0.848).

The model correctly classifies 96.2% of the participants in the two groups of language delayed and not delayed, with a high specificity (99.2) but a modest sensitivity (33.3), indicating that there is a high percentage of false negatives.

Three predictive factors seem to be involved in grammar understanding at 60 months of age (CEG): maternal education level, cognitive delay at 22 months, and language comprehension at 48 months of age. Low maternal education increases the probability of having children with language impairment (grammar understanding) at 60 months ($p < 0.01$, OR = 0.434). It is interesting to note that maternal education at this point has a significant effect, but this effect did not exist when the participants were younger. This apparently points to a cumulative effect of maternal education level throughout time, which is compatible with Linsell's et al. [68] suggestion that the impact of environmental factors on cognitive development becomes more prominent over time for VPT children. Other authors suggested the same cumulative effect for language development [14,31,69,70]. Another interesting and somewhat surprising result is that cognitive delay measured at 22 months of age still has a predictive effect on grammar understanding impairment at 50 months of age ($p < 0.05$, OR 5.929), demonstrating a long lasting and strong (OR value) effect, which is coincident with other findings for VPT children of the same age [33]. Not surprisingly, low scores in understanding language (RDLS) at 48 months increment the possibility of having delays in grammar understanding at 60 months of age ($p < 0.01$, OR = 0.891), remarking the predictive role of previous language abilities in the same domain (understanding).

The model (Table 9) correctly classified 80.3% of the children into the two categories of language delayed and not delayed, and has a high specificity (96.9), although a relatively low sensitivity (34.3).

Finally, two factors seem to have a significant predictive effect on morphosyntactic production delay at 60 months of age: vocabulary delay at 30 months, and low scores in the language comprehension scale of the RDLS. The fact of having language (lexical) delay at 30 months increases the probability of having morphosyntactic impairments 30 months later ($p < 0.05$, OR = 2.977). Those children with low scores in language understanding (RDLS) also have a higher probability of being delayed in morphosyntactic production one year later ($p < 0.01$, OR = 0.880).

This time the model (Table 10) is less powerful in the process of classifying the children into the two groups (language delayed/not language delayed in morphosyntactic production) since only 69.5% of the children are correctly classified. Although specificity is high (91.1), sensitivity is even lower than in the other regression analyses (22.0), thus indicating the existence of many false negatives.

The use of a longitudinal design, in which the children were followed from 10 to 60 months of age, allows for the revelation of certain findings which would not be patent in a cross-sectional design or a short-term longitudinal design.

First, these results show that certain predictors of early language delay (22 months), such as a low number of gestures produced at 10 months of age or low vocabulary understanding at the same age, lose their effect as children grow older, contradicting the results of other studies [54]. Similarly, gender seems to have an effect on language delay at the beginning (22 months of age), while it seems to lose its effect on later language development [15,33].

Complementarily, low maternal education does not have an effect on language delay during the first stages of language development (22 and 30 months of age) but, however, emerges as a predictive factor of grammar understanding delay at the age of 60 months. This pattern shows that environmental factors have a cumulative or incremental effect on language development over time [14,31,69,70]. In any case, the effect of maternal education is not general throughout all the linguistic domains. Possibly, grammar understanding assessed through the CEG is more demanding on abilities linked to the effect of family activities and cultural practices which are developed in families with mothers who have a medium to high educational level than the other tests are (PPVT-III, or TSA).

Another surprising (and relatively unexpected) result is the long-lasting effect of cognitive delay on language delay measured at different ages (30 and 60 months). This reinforces the idea that cognitive development is one of the most powerful predictors of language development [14,15,31,51,52], particularly for PT children. Again, the effect of cognitive delay is more evident in the case of grammar understanding, probably because this test is more demanding of cognitive resources (including working memory) than the other tests used.

Previous language delay has also been found to have an important effect on later language delay [14,30,33,57]. This occurs particularly if the domains of language measured are linked and if the time spent between the ages of measurement is not very long.

The different models tested in the logistic regression analyses can only explain a relatively modest percentage of the variance in the different linguistic measurements, ranging from Negalkerke's R^2 = 0.153 in the case of word production at 22 months of age to Negalkerke's R^2 = 0.414 in the case of vocabulary comprehension at 60 months of age. This indicates that other factors, whose effects have not been studied in this research, may also be predictors of language delay at different ages. In fact, the low sensitivity values found (ranging from 18.6 to 34.3) gives support to the former idea.

As a final conclusion, but not less important, GA does not seem to have any important effect on the prediction of language delay in the case of healthy preterm children when no serious handicap is associated, coinciding with other research findings [14,15,71]. These results of the regression analyses also agree with the ANOVA comparisons, giving strength to the conclusion. The fact that gestational age did not have any significant effect on the language delay of low-risk PT children needs to be highlighted, since this is a novel result in the literature, and it contrasts with other previous studies carried out with VPR or EPR children [29,30,33]. Again, this conclusion has to be taken with caution and cannot be generalized for PT children with other conditions.

5. Conclusions

Several conclusions can be drawn from this study. First, healthy PT children do not have, in general terms, a higher risk of language delay than FT children, and seem to have a lower risk of language delay/impairment than very preterm or extremely preterm children studied in other investigations. Second, previous language delay and cognitive delay are the strongest and longest-lasting predictors of later language impairment. The effect of certain predictors of early language delay, such as a low number of gestures and low vocabulary understanding at 10 months of age as well as gender, disappears as language development evolves. On the contrary low maternal education affects language delay after a certain point, indicating a cumulative effect over time.

A limitation of this study is that the effect of medical problems on PT children's language delay could not be studied since children with medical problems were excluded.

Funding: This research was funded by the Ministerio Economía Industria y Competitividad of the Spanish Government (Grants PSI2008-03905, PSI2011-23210 and PSI2015-66697-R to the author). Founds for open access publication fees were received from the Consellería de Educación, Universidade e Formación Profesional -Xunta de Galicia ED431C 2018/02.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Ethics Committee of Clinical Investigation of Galicia (Comité Ético de Investigación Clínica de Galicia) (code: 2008/010, 28/02/2008).

Informed Consent Statement: Informed consent was obtained from all the legal tutors of the participants involved in the study.

Data Availability Statement: Data supporting reported results can be asked for to the author.

Acknowledgments: Thanks are given to all the families participating in the study.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Blencowe, H.; Cousens, S.; Chou, D.; Oestergaard, M.; Say, L.; Moller, A.-B.; Kinney, M.; Lawn, J. Born too soon: The global epidemiology of 15 million preterm births. *Reprod. Health* **2013**, *10* (Suppl. S1), S2. [CrossRef]
2. Goldenberg, R.L.; Culhane, J.F.; Iams, J.D.; Romero, R. Epidemiology and causes of preterm birth. *Lancet* **2008**, *371*, 75–84. [CrossRef]
3. March of Dimes; The Partnership for Maternal, Newborn & Child Health; Save the Children; World Health Organization. *Born Too Soon: The Global Action Report On Preterm Birth*; World Health Organization: Geneva, Switzerland, 2012.
4. Johansson, S.; Cnattigius, S. Epidemiology of preterm birth. In *Neurodevelopmental Outcomes of Preterm Birth. From Childhood to Adult Life*; Nosarti, C., Murray, R.M., Hack, M., Eds.; Cambridge University Press: Cambridge, UK, 2010; pp. 1–16.
5. Woodward, L.J.; Moor, S.; Hood, K.M.; Champion, P.R.; Foster-Cohen, S.; Inder, T.E.; Austin, N.C. Very preterm children show impairments across multiple neurodevelopmental domains by age 4 years. *Arch. Dis. Child. Fetal Neonatal Ed.* **2009**, *94*, 339–344. [CrossRef]
6. Barre, N.; Morgan, A.; Doyle, L.W.; Anderson, P.J. Language abilities in children who were very preterm and/or very low birth weight: A meta-analysis. *J. Pediatr.* **2011**, *158*, 766–774.e1. [CrossRef]
7. Nguyen, T.-N.-N.; Spencer-Smith, M.; Zannino, D.; Burnett, A.; Scratch, S.E.; Pascoe, L.; Ellis, R.; Cheong, J.; Thompson, D.; Inder, T.; et al. Developmental trajectory of language from 2 to 13 years in children born very preterm. *Pediatrics* **2018**, *141*, e20172831. [CrossRef]
8. van Noort-van der Spek, I.L.; Franken, M.-C.J.; Weisglas-Kuperus, N. Language functions in preterm-born children: A systematic review and meta-analysis. *Pediatrics* **2012**, *129*, 745–754. [CrossRef] [PubMed]
9. Gayraud, F.; Kern, S. Influence of preterm birth on early lexical and grammatical acquisition. *First Lang.* **2007**, *27*, 159–173. [CrossRef]
10. Sansavini, A.; Savini, S.; Guarini, A.; Broccoli, S.; Alessandroni, R.; Faldella, G. The effect of gestational age on developmental outcomes: A longitudinal study in the first 2 years of life. *Child Care Health Dev.* **2010**, *37*, 26–36. [CrossRef]
11. Sansavini, A.; Guarini, A.; Savini, S. Linguistic and cognitive delays in very preterm infants at 2 years: General or specific delays? *Rev. Logop. Foniatr. Audiol.* **2011**, *31*, 133–147. [CrossRef]

12. Stolt, S.; Matomäki, J.; Haataja, L.; Lapinleimu, H.; Lehtonen, L.; The PIPARI Study Group. The emergence of grammar in very-low-birth-weight Finnish children at two years of age. *J. Child Lang.* **2012**, *40*, 336–357. [CrossRef] [PubMed]
13. Sansavini, A.; Guarini, A.; Alessandroni, R.; Faldella, G.; Giovanelli, G.; Salvioli, G. Early relations between lexical and grammatical development in very immature Italian preterms. *J. Child Lang.* **2006**, *33*, 199–216. [CrossRef] [PubMed]
14. Pérez-Pereira, M.; Fernández, P.; Gómez-Taibo, M.L.; Resches, M. Language development of low risk preterm infants up to the age of 30months. *Early Hum. Dev.* **2014**, *90*, 649–656. [CrossRef] [PubMed]
15. Suttora, C.; Guarini, A.; Zuccarini, M.; Aceti, A.; Corvaglia, L.; Sansavini, A. Speech and language skills of low-risk preterm and full-term late talkers: The role of child factors and parent input. *Int. J. Environ. Res. Public Health* **2020**, *17*, 7684. [CrossRef] [PubMed]
16. Fisher, E.L. A systematic review and meta-analysis of predictors of expressive-language outcomes among late talkers. *J. Speech Lang. Hear. Res.* **2017**, *60*, 2935–2948. [CrossRef] [PubMed]
17. Carson, C.P.; Klee, T.; Carson, D.K.; Hime, L.K. Phonological profiles of 2-year-olds with delayed language development: Predicting clinical outcomes at age 3. *Am. J. Speech Lang. Pathol.* **2003**, *12*, 28–39. [CrossRef]
18. Rescorla, L.; Schwartz, E. Outcome of toddlers with specific expressive language delay. *Appl. Psycholinguist.* **1990**, *11*, 393–407. [CrossRef]
19. Zubrick, S.R.; Taylor, C.L.; Rice, M.L.; Slegers, D.W. Late language emergence at 24 months: An epidemiological study of prevalence, predictors, and covariates. *J. Speech Lang. Hear. Res.* **2007**, *50*, 1562–1592. [CrossRef]
20. Dale, P.S.; Price, T.S.; Bishop, D.V.M.; Plomin, R. Outcomes of early language delay: I. Predicting persistent and transient language difficulties at 3 and 4 years. *J. Speech Lang. Hear Res.* **2003**, *46*, 544–560. [CrossRef]
21. Lee, Y. Prevalence of developmental language delay among toddlers born as preterm. *Commun. Sci. Disord.* **2016**, *21*, 60–68. [CrossRef]
22. Sajaniemi, N.; Hakamies-Blomqvist, L.; Mäkelä, J.; Avellan, A.; Rita, H.; Von Wendt, L. Cognitive development, temperament and behavior at 2 years as indicative of language development at 4 years in pre-term infants. *Child Psychiatry Hum. Dev.* **2001**, *31*, 329–346. [CrossRef] [PubMed]
23. Woods, P.L.; Rieger, I.; Wocadlo, C.; Gordon, A. Predicting the outcome of specific language impairment at five years of age through early developmental assessment in preterm infants. *Early Hum. Dev.* **2014**, *90*, 613–619. [CrossRef]
24. Stene-Larsen, K.; Brandlistuen, R.E.; Lang, A.M.; Landolt, M.A.; Latal, B.; Vollrath, M.E. Communication impairments in early term and late preterm children: A prospective cohort study following children to age 36 months. *J. Pediatr.* **2014**, *165*, 1123–1128. [CrossRef]
25. Zambrana, I.M.; Vollrath, M.E.; Jacobsson, B.; Sengpiel, V.; Ystrom, E. Preterm birth and risk for language delays before school entry: A sibling-control study. *Dev. Psychopathol.* **2021**, *33*, 47–52. [CrossRef] [PubMed]
26. Briscoe, J.; Gathercole, S.E.; Marlow, N. Short-term memory and language outcomes after extreme prematurity at birth. *J. Speech Lang. Hear. Res.* **1998**, *41*, 654–666. [CrossRef]
27. Singer, L.T.; Siegel, A.C.; Lewis, B.; Hawkins, S.; Yamashita, T.; Baley, J. Preschool language outcomes of children with history of bronchopulmonary dysplasia and very low birth weight. *J. Dev. Behav. Pediatr.* **2001**, *22*, 19–26. [CrossRef]
28. Wolke, D.; Samara, M.; Bracewell, M.; Marlow, N. Specific language difficulties and school achievement in children born at 25 weeks of gestation or less. *J. Pediatr.* **2008**, *152*, 256–262.e1. [CrossRef]
29. Foster-Cohen, S.H.; Friesen, M.D.; Champion, P.R.; Woodward, L.J. High prevalence/low severity language delay in preschool children born very preterm. *J. Dev. Behav. Pediatr.* **2010**, *31*, 658–667. [CrossRef] [PubMed]
30. Sansavini, A.; Guarini, A.; Justice, L.M.; Savini, S.; Broccoli, S.; Alessandroni, R.; Faldella, G. Does preterm birth increase a child's risk for language impairment? *Early Hum. Dev.* **2010**, *86*, 765–772. [CrossRef] [PubMed]
31. Sansavini, A.; Guarini, A.; Savini, S.; Broccoli, S.; Justice, L.; Alessandroni, R.; Faldella, G. Longitudinal trajectories of gestural and linguistic abilities in very preterm infants in the second year of life. *Neuropsychologia* **2011**, *49*, 3677–3688. [CrossRef]
32. Charollais, A.; Stumpf, M.-H.; De Quelen, R.; Rondeau, S.; Pasquet, F.; Marret, S. Delayed language development at two years of age in very preterm infants in the Perinatal Network of Haute-Normandie. *Early Hum. Dev.* **2014**, *90*, 891–892. [CrossRef]
33. Stolt, S.; Matomaki, J.; Lind, A.; Lapinleimu, H.; Haataja, L.; Lehtonen, L. The prevalence and predictive value of weak language skills in children with very low birth weight—A longitudinal study. *Acta Paediatr.* **2014**, *103*, 651–658. [CrossRef]
34. Do, C.H.T.; Kruse, A.Y.; Wills, B.; Sabanathan, S.; Clapham, H.; Pedersen, F.K.; Pham, T.N.; Vu, P.M.; Børresen, M.L. Neurodevelopment at 2 years corrected age among Vietnamese preterm infants. *Arch. Dis. Child.* **2019**, *105*, 134–140. [CrossRef]
35. Tomblin, J.B.; Records, N.L.; Buckwalter, P.; Zhang, X.; Smith, E.; O'Brien, M. Prevalence of specific language impairment in kindergarten children. *J. Speech Lang. Hear. Res.* **1997**, *40*, 1245–1260. [CrossRef]
36. Reilly, S.; Wake, M.; Ukoumunne, O.C.; Bavin, E.; Prior, M.; Cini, E.; Conway, L.; Eadie, P.; Bretherton, L. Predicting language outcomes at 4 years of age: Findings from early language in Victoria study. *Pediatrics* **2010**, *126*, e1530–e1537. [CrossRef] [PubMed]
37. Belgin, P.; Abraham, B.; Baburaj, S.; Mohandas, M. Environmental and biological risk factors associated with the prevalence of language delay in children upto 6 years of age from Rural South India. *J. Clin. Diagn. Res.* **2017**, *11*, 11–14. [CrossRef]
38. Chung, E.H.; Chou, J.; Brown, K.A. Neurodevelopmental outcomes of preterm infants: A recent literature review. *Transl. Pediatr.* **2020**, *9*, S3–S8. [CrossRef]
39. Baron, I.S.; Ahronovich, M.D.; Erickson, K.; Larson, J.C.G.; Litman, F.R. Age-appropriate early school age neurobehavioral outcomes of extremely preterm birth without severe intraventricular hemorrhage: A single center experience. *Early Hum. Dev.* **2009**, *85*, 191–196. [CrossRef] [PubMed]

40. Lean, R.E.; Paul, R.A.; Smyser, T.A.; Smyser, C.D.; Rogers, C.E. Social adversity and cognitive, language, and motor development of very preterm children from 2 to 5 years of age. *J. Pediatr.* **2018**, *203*, 177–184.e1. [CrossRef]
41. Payne, A.H.; Hintz, S.R.; Hibbs, A.M.; Walsh, M.C.; Vohr, B.R.; Bann, C.M.; Wilson-Costello, D.E. Eunice Kennedy Shriver National Institute of Child Health and Human Development Neonatal Research Network. Neurodevelopmental outcomes of extremely low-gestational-age neonates with low-grade periventricular-intraventricular hemorrhage. *JAMA Pediatr.* **2013**, *167*, 451–459. [CrossRef] [PubMed]
42. Doyle, L.W.; Anderson, P.J. Long-term outcomes of bronchopulmonary dysplasia. *Semin. Fetal Neonatal Med.* **2009**, *14*, 391–395. [CrossRef]
43. Lewis, B.A.; Singer, L.T.; Fulton, S.; Salvator, A.; Short, E.J.; Klein, N.; Baley, J. Speech and language outcomes of children with bronchopulmonary dysplasia. *J. Commun. Disord.* **2002**, *35*, 393–406. [CrossRef]
44. Lin, F.; Dong, H.; Song, Y.; Zhang, T.; Qi, J.; Xiao, X.; Cai, Y. Effect of bronchopulmonary dysplasia on early intellectual development in preterm infants. *J. Jpn. Pediatr. Soc.* **2017**, *59*, 691–697. [CrossRef]
45. Campbell, T.F.; Dollaghan, C.A.; Rockette, H.E.; Paradise, J.L.; Feldman, H.M.; Shriberg, L.D.; Sabo, D.L.; Kurs-Lasky, M. Risk factors for speech delay of unknown origin in 3-year-old children. *Child Dev.* **2003**, *74*, 346–357. [CrossRef]
46. Spencer-Smith, M.; Pascoe, L.; Treyvaud, K.; Lee, K.J.; Thompson, D.K.; Cheong, J.L.; Doyle, L.W.; Anderson, P.J. Language skills in children born preterm (<30 wks' gestation) throughout childhood: Associations with biological and socioenvironmental factors. *J. Dev. Behav. Pediatr.* **2019**, *40*, 735–742. [CrossRef]
47. Greene, M.M.; Patel, A.L.; Meier, P.; Patra, K. Maternal education level predicts cognitive, language, and motor outcome in preterm infants in the second year of life. *Am. J. Perinatol.* **2016**, *33*, 738–744. [CrossRef] [PubMed]
48. Sentenac, M.; Johnson, S.; Charkaluk, M.-L.; Sëppanen, A.-V.; Aden, U.; Cuttini, M.; Maier, R.; Mannamaa, M.; Zeitlin, J. Maternal education and language development at 2 years corrected age in children born very preterm: Results from a European population-based cohort study. *J. Epidemiol. Community Health* **2020**, *74*, 346–353. [CrossRef] [PubMed]
49. Marchman, V.A.; Adams, K.A.; Loi, E.C.; Fernald, A.; Feldman, H.M. Early language processing efficiency predicts later receptive vocabulary outcomes in children born preterm. *Child Neuropsychol.* **2016**, *22*, 649–665. [CrossRef] [PubMed]
50. Weisglas-Kuperus, N.; Baerts, W.; De Graaf, M.; Van Zanten, G.; Sauer, P. Hearing and language in preschool very low birthweight children. *Int. J. Pediatr. Otorhinolaryngol.* **1993**, *26*, 129–140. [CrossRef]
51. Desmarais, C.; Sylvestre, A.; Meyer, F.; Bairati, I.; Rouleau, N. Systematic review of the literature on characteristics of late-talking toddlers. *Int. J. Lang. Commun. Disord.* **2008**, *43*, 361–389. [CrossRef]
52. Vandormael, C.; Schoenhals, L.; Hüppi, P.S.; Filippa, M.; Tolsa, C.B. Language in preterm born children: Atypical development and effects of early interventions on neuroplasticity. *Neural Plast.* **2019**, *2019*, 6873270. [CrossRef] [PubMed]
53. Stolt, S.; Mäkilä, A.-M.; Matomäki, J.; Lehtonen, L.; Lapinleimu, H.; Haataja, L. The development and predictive value of gestures in very-low-birth-weight children: A longitudinal study. *Int. J. Speech Lang. Pathol.* **2013**, *16*, 121–131. [CrossRef]
54. Stolt, S.; Lind, A.; Matomäki, J.; Haataja, L.; Lapinleimu, H.; Lehtonen, L. Do the early development of gestures and receptive and expressive language predict language skills at 5;0 in prematurely born very-low-birth-weight children? *J. Commun. Disord.* **2016**, *61*, 16–28. [CrossRef] [PubMed]
55. Suttora, C.; Salerni, N. Gestural development and its relation to language acquisition in very preterm children. *Infant Behav. Dev.* **2012**, *35*, 429–438. [CrossRef] [PubMed]
56. Marchman, V.A.; Ashland, M.D.; Loi, E.C.; Adams, K.A.; Fernald, A.; Feldman, H.M. Predictors of early vocabulary growth in children born preterm and full term: A study of processing speed and medical complications. *Child Neuropsychol.* **2019**, *25*, 943–963. [CrossRef] [PubMed]
57. Schults, A.; Tulviste, T.; Haan, E. Early vocabulary in full term and preterm Estonian children. *Early Hum. Dev.* **2013**, *89*, 721–726. [CrossRef]
58. Pérez Pereira, M.; García Soto, X.R. El diagnóstico del desarrollo comunicativo en la primera infancia: Adaptación de las escalas MacArthur al gallego. *Psicothema* **2003**, *15*, 352–361. (In Spanish)
59. Pérez-Pereira, M.; Resches, M. Concurrent and predictive validity of the Galician CDI. *J. Child Lang.* **2010**, *38*, 121–140. [CrossRef] [PubMed]
60. Fenson, L.; Marchman, V.A.; Thal, D.; Dale, P.S.; Reznick, J.S.; Bates, E. *The MacArthur-Bates Communicative Development Inventories: User's Guide and Technical Manual*, 2nd ed.; Paul H. Brookes: Baltimore, MD, USA, 2007.
61. Edwards, S.; Fletcher, P.M.G.; Hughes, A.; Letts, C.; Sinka, I. *Escala de Desarrollo del Lenguaje de Reynell*, 3rd ed.; Psymtec: Madrid, Spain, 1997. (In Spanish)
62. Dunn, L.M.; Dunn, L.M.; Arribas, D. *PEABODY. Test de Vocabulario en Imágenes*; TEA Ediciones: Madrid, Spain, 2005. (In Spanish)
63. Mendoza, E.; Carballo, G.; Muñoz, J.; Fresneda, D. *CEG. Test de Comprensión de Estructuras Gramaticales*; TEA Ediciones: Madrid, Spain, 2005. (In Spanish)
64. Aguado, G. *Desarrollo de la Morfosintaxis en el Niño (TSA)*; CEPE: Madrid, Spain, 1999. (In Spanish)
65. Bishop, D. *Test for Reception of Grammar (TROG-2)*; Pearson: New York, NY, USA, 2003.
66. Newborg, J.; Stock, J.R.; Wnek, L. *Batelle, Inventario de Desarrollo*; TEA Ediciones: Madrid, Spain, 1996. (In Spanish)
67. Resches, M.; Pérez-Pereira, M.; Cruz Guerrero, R.; Fernández Prieto, M. Risk for language delay in healthy preterm and full-term children. In *Atypical Language Development in Romance Languages*; Aguilar-Mediavilla, E., Buil-Legaz, L., López-Penadés, R., Sánchez-Azanza, V., Adrover-Roig, D., Eds.; John Benjamins Publishing: Amsterdam, The Netherlands, 2019; pp. 57–74.

68. Linsell, L.; Malouf, R.; Morris, J.K.; Kurinczuk, J.J.; Marlow, N. Prognostic factors for poor cognitive development in children born very preterm or with very low birth weight. *JAMA Pediatr.* **2015**, *169*, 1162–1172. [CrossRef] [PubMed]
69. Hoff, E. How social contexts support and shape language development. *Dev. Rev.* **2006**, *26*, 55–88. [CrossRef]
70. Molfese, V.J.; Holcomb, L.; Helwig, S. Biomedical and social-environmental influences on cognitive and verbal abilities in children 1 to 3 years of age. *Int. J. Behav. Dev.* **1994**, *17*, 271–287. [CrossRef]
71. Pérez-Pereira, M.; Martínez-López, Z.; Maneiro, L. Longitudinal relationships between reading abilities, phonological awareness, language abilities and executive functions: Comparison of low risk preterm and full-term children. *Front. Psychol.* **2020**, *11*, 468. [CrossRef] [PubMed]

children

Article

Associations between Language at 2 Years and Literacy Skills at 7 Years in Preterm Children Born at Very Early Gestational Age and/or with Very Low Birth Weight

Eveliina Joensuu [1,*], Petriina Munck [1], Sirkku Setänen [2], Jari Lipsanen [1], Mira Huhtala [3], Helena Lapinleimu [2] and Suvi K. J. Stolt [1]

[1] Department of Psychology and Logopedics, Faculty of Medicine, University of Helsinki, 00014 Helsinki, Finland; petriina.munck@gmail.com (P.M.); jari.lipsanen@helsinki.fi (J.L.); suvi.stolt@helsinki.fi (S.K.J.S.)
[2] Department of Pediatrics and Adolescent Medicine, University of Turku and Turku University Hospital, 20014 Turku, Finland; sirkku.setanen@utu.fi (S.S.); lehela@utu.fi (H.L.)
[3] Oncology and Radiotherapy, Turku University Hospital, 20014 Turku, Finland; mivall@utu.fi
* Correspondence: eveliina.a.joensuu@helsinki.fi

Citation: Joensuu, E.; Munck, P.; Setänen, S.; Lipsanen, J.; Huhtala, M.; Lapinleimu, H.; Stolt, S.K.J. Associations between Language at 2 Years and Literacy Skills at 7 Years in Preterm Children Born at Very Early Gestational Age and/or with Very Low Birth Weight. *Children* **2021**, *8*, 510. https://doi.org/10.3390/children8060510

Academic Editors: Eva Aguilar Mediavilla, Miguel Pérez Pereira, Elisabet Serrat-Sellabona and Daniel Adrover-Roig

Received: 30 May 2021
Accepted: 14 June 2021
Published: 16 June 2021

Abstract: Preterm children (born <37 gestational weeks) who are born at very early gestational age (<32 weeks, very preterm, VP) and/or with very low birth weight (≤1500 g, VLBW) are at increased risk for language and literacy deficits. The continuum between very early language development and literacy skills among these children is not clear. Our objective was to investigate the associations between language development at 2 years (corrected age) and literacy skills at 7 years in VP/VLBW children. Participants were 136 VP/VLBW children and 137 term controls (a 6-year regional population cohort, children living in Finnish-speaking families). At 2 years of corrected age, language (lexical development, utterance length) was assessed using the Finnish version of the MacArthur–Bates Communicative Development Inventory and the Expressive Language Scale from Bayley scales of Infant Development, second edition. At 7 years, children's literacy skills (pre-reading skills, reading, and writing) were evaluated. Statistically significant correlations were found in both groups between language development at 2 years and literacy skills at 7 years (r-values varied between 0.29 and 0.43, $p < 0.01$). In the VP/VLBW group, 33% to 74% of the children with early weak language development had weak literacy skills at 7 years relative to those with more advanced early language skills (11% to 44%, $p < 0.001$ to 0.047). Language development at 2 years explained 14% to 28% of the variance in literacy skills 5 years later. Language development at 2 years had fair predictive value for literacy skills at 7 years in the VP/VLBW group (area under the receiver operating characteristic (ROC) curve (AUC) values varied between 0.70 and 0.77, $p < 0.001$). Findings provide support for the continuum between very early language development and later language ability, in the domain of literacy skills in preterm children.

Keywords: early language development; literacy skills; very preterm; very low birth weight; prematurely born children; longitudinal follow-up; regional cohort study; assessment

1. Introduction

Prematurely born (<37 gestational weeks) children born at very early gestational age (<32 gestational weeks, very preterm, VP) and/or with very low birth weight (≤1500 g, VLBW) are at increased risk for developmental impairments and learning deficits such as difficulties in early language development [1–3] and literacy skills [4–7], including pre-reading skills, reading, and writing. The gap to full-term controls in language and literacy skills is evident even in the absence of neurodevelopmental impairment (NDI), including cerebral palsy, hearing impairment, blindness, or severe cognitive impairment (intelligence quotient, IQ < 70) [8,9]. The goal of clinical follow-up is to identify weak development as early as possible to provide targeted intervention to improve developmental outcomes. Findings of recent longitudinal studies, although sparse, suggest that difficulties in language functions persist from early years through late childhood, up to the age of 13 years [3,10,11]. Previous investigations, including a recent study of a large French cohort [12], highlight the usefulness of a validated parent-reported measure, such as the MacArthur–Bates Communicative Development Inventories (CDI) [13], in assessing early language skills of children born VP/VLBW to predict developmental difficulties in language [2,3].

Although earlier studies have provided essential information regarding the continuum between early language skills and later language performance in children born VP/VLBW, far fewer reports have used literacy skills as an outcome measure [14–17]. Furthermore, most of the existing studies examining associations between language and literacy skills have been based on samples of school-aged children, not assessing very early childhood. To date, the earliest age point in a longitudinal setting has been reported by Pritchard et al. [17] who investigated the relations between school readiness domains, including language, at the corrected age (i.e., adjusted age, representing the age of the child from the expected date of delivery) of 4 years and later educational achievement at school age. To the best of our knowledge, the possible longitudinal associations between very early language development at 2 years and literacy skills at 7 years is an open question in this high-risk population.

In early clinical follow-up of prematurely born children, their development is often followed up to the age of 2 years. However, the language development of VP/VLBW children is not always assessed specifically. Clinicians evaluating early language development of children born VP/VLBW would benefit from the knowledge of whether lexical development and utterance length at 2 years of corrected age have predictive value for literacy outcome at 7 years in this vulnerable population, and whether there is a cost-effective way of identifying toddlers at potential risk for literacy deficits. To maximize the effects of early intervention, it is crucial to identify children with weak skills as early as possible.

In the current study, we analyze longitudinal associations between language skills at 2 years of corrected age and literacy skills at 7 years in a Finnish sample of children born VP/VLBW. In Finland, children begin formal schooling in the year in which they reach the age of 7 years. Finnish is a transparent language with a highly regular grapheme-phoneme correspondence, and thus, basic decoding skills are often acquired during the first year of school see e.g., [18]. In addition, more than one-third of Finnish first-graders can already read before entering school. Previous findings from longitudinal studies investigating Finnish children with a hereditary risk for dyslexia and their controls [19,20] suggest that features of early language development, including lexical development and utterance length, have predictive value for later reading acquisition [20,21]. In preterm children, this association has not been analyzed previously.

This study had three aims: (1) to evaluate the associations between language skills at 2 years of corrected age and literacy skills at the beginning of schooling, at 7 years, in a regional cohort of Finnish-speaking children born VP/VLBW and in their full-term controls; (2) to analyze how much early language skills explain the variance of literacy skills at 7 years; and (3) to assess the predictive value of language skills at 2 years for literacy skills at 7 years measured using area under the receiver operating characteristic (ROC) curve (AUC) values in VP/VLBW children and their controls.

2. Materials and Methods

2.1. Participants

This study is part of the multidisciplinary 6-year regional cohort study of prematurely born children called PIPARI (Development and Functioning of Very Low Birth Weight Infants from Infancy to School Age) [22,23]. The participants were children born <32 weeks of gestational age and/or with birth weight ≤1500 g in Turku University Hospital, Finland, in 2001–2006. From 2001 to 2003, the inclusion criteria were birth weight ≤1500 g and prematurity (<37 gestational weeks). From the beginning of 2004, the inclusion criteria were expanded to include all infants born <32 weeks of gestation, regardless of birth weight. At least one of the parents had to speak Finnish or Swedish, the two official languages in Finland. Children with severe congenital anomalies or diagnosed syndromes affecting their development were excluded.

The present study sample consisted of 136 children born VP/VLBW living in Finnish-speaking families. The flow chart of the children born VP/VLBW participating in this study is presented in Figure 1. Neurodevelopmental impairment was determined if one or more of the following factors were present by the corrected age of 2 years: cerebral palsy, hearing impairment (threshold >40 dB), blindness, or severe cognitive impairment (Mental Developmental Index, MDI of the Bayley Scales of Infant Development II [24], BSID-II, <70 standard scores). In the PIPARI study, the age of VP/VLBW children was corrected for prematurity until the age of 2 years.

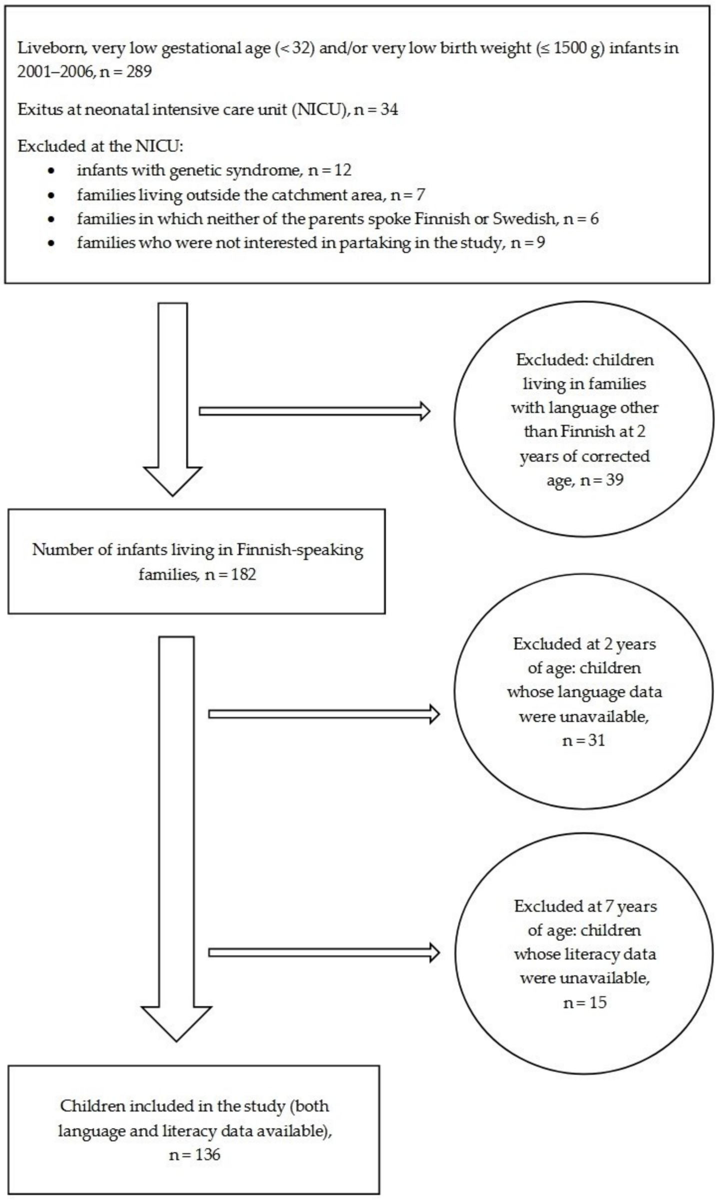


Figure 1. Flow chart of the prematurely born children born at very early gestational age (<32 weeks, very preterm, VP) and/or with very low birth weight (≤1500 g, VLBW) included in the study.

The control group consisted of healthy full-term ($\leq$37 weeks of gestation) infants born in the same hospital between 2001 and 2004. They were recruited by asking the parents of the first boy and the first girl born in each week to take part in the study. If the family was not interested in partaking in the study, the parents of the next boy/girl were invited. The full-term controls were born $\geq$37 weeks of gestation, were not admitted to a neonatal care unit during the first week of life, and had at least one parent speaking either Finnish or Swedish. The exclusion criteria were any major congenital anomalies or chromosomal or genetic syndromes, the mother's known use of illicit drugs or alcohol during pregnancy, and the infant's birth weight being small for gestational age according to age- and gender-specific Finnish growth charts. In the present study, only those 136 children born VP/VLBW and those 137 controls who had data available from both the language assessment at 2 years of corrected age and literacy skills assessment at 7 years were included.

The PIPARI study protocol was approved by the Ethics Review Committee of the Hospital District of Southwest Finland in December 2000 and January 2012. After receiving oral and written information, all parents who agreed to participate provided written informed consent.

2.2. Assessment at 2 Years of Corrected Age

Language skills were assessed with the Finnish long-form version of the MacArthur–Bates Communicative Development Inventory (FinCDI, toddler version) [25], and the Expressive Language Score (ELS) from BSID-II. The FinCDI is a validated, parent-report measure evaluating the development of lexicon and grammar, including inflectional morphological skills. Variables of lexicon size and mean length of the three longest utterances (M3L) were used. Lexicon size is the number of words the parents estimated that their child uses, based on word lists (595 words). M3L is calculated in morphemes (i.e., the smallest units of language creating a difference in meaning) based on the three longest recent utterances the child has made. The ELS consists of 10 pictures and 5 objects that the child was asked to name in the testing situation.

2.3. Assessment at 7 Years

Reading precursors, reading, and writing ability were evaluated during the first weeks of grade 1 of primary school (a 6-week period from August to September during the school entrance year). Reading precursors assessed were phonological awareness and letter knowledge. To evaluate phonological awareness, three- to seven-letter words were presented phoneme by phoneme [26]. Children were instructed to mark one picture out of four alternatives that they thought would best match the word (max 9). To evaluate letter knowledge, the child was asked to name 29 uppercase letters presented in random order (max 29) [27]. In this study, the sum score of the tasks of phonological awareness and letter knowledge was used as the measure of precursors of reading (max 38).

Reading skills were evaluated using a short version of the Finnish reading test ARMI—a tool for assessing reading and writing skills in Grade 1 [27], consisting of a wordlist of two-syllable (seven words), three-syllable (two words), and five-syllable (one word) words. The child was asked to read the words aloud. The score for reading skills was the number of correctly read words (max 10). To evaluate writing skills, the child was asked to write 5 words and 8 pseudowords said aloud one word at a time [19]. The writing skills score was the total number of correctly written items (max 13).

2.4. Statistical Analyses

All analyses were run separately for all children born VP/VLBW, for preterm children without neurodevelopmental impairment, and for controls. Pearson's correlation coefficient values were used to investigate the correlations between the continuous language and literacy variables measured at 2 and 7 years. All language and literacy variables were also categorized. The 10th percentile cut-off value was used to evaluate the association between early weak language skills at 2 years of corrected age and weak literacy skills at

7 years of age. For the FinCDI, the cut-off value was based on the normative sample, and for the other measures, the 10th percentile cut-off values were derived from the control group. Comparisons between categorical variables were done using cross-tabulation with the Chi-square test or Fisher's exact test. Multiple variable linear regression analysis was conducted to assess how much 2-year language variables explain the variance in literacy skills at 7 years when the effect of background factors were taken into consideration. The dependent variables were reading precursors (sumscore of letter knowledge and phoneme synthesis), reading, and writing skills at 7 years. The independent variables were lexicon size, M3L, and ELS measured at 2 years. Since the independent variables were strongly correlated with each other, they were analyzed separately. Nine regression models were run: in the first three models, lexicon size was used as an independent variable; in the next three models, M3L; and in the last three models, ELS. In the preliminary analyses, the following background factors were associated with the outcome variables and were therefore included in the regression models: gestational age, mother's self-reported reading difficulties, father's self-reported reading difficulties, and paternal education. Maternal education was not included because in the preliminary analysis paternal education level correlated more strongly with the outcome variables. Due to multicollinearity between maternal and paternal education, only paternal education was included in the regression analyses. Lastly, the predictive value of early language development at 2 years for literacy skills at 7 years was analyzed using the AUC values. The AUC is the measure of the ability of a test to distinguish between classes [28]. The greater the AUC values, the better the prediction model. An area of 1. represents a perfect classifier, whereas a ROC curve no better than chance would have an area under the curve of 0.5. AUC values are interpreted as follows: excellent predictive value 0.90–1, good 0.80–0.90, fair 0.70–0.80, poor 0.60–0.70, and fail <0.60 [28]. All statistical analyses were performed using IBM® SPSS® Statistics for Windows, version 26.0. (IBM Corp., Armonk, NY, USA). Two-tailed p-values < 0.05 were considered statistically significant.

3. Results

3.1. Data Description

The background characteristics of the children are presented in Table 1. No statistically significant difference in background factors was found between the children born VP/VLBW who participated in the study and the VP/VLBW children living in Finnish-speaking families whose language and literacy data were unavailable (n = 46, 25%), except that there were more multiple births among participating children (36% of the study children vs. 16% of the dropouts, p = 0.02).

Table 1. Background characteristics of very preterm/very low birth weight (VP/VLBW) children and full-term controls. Numbers (percentages) are shown. If mean (standard deviation) [minimum, maximum] are presented, they are indicated separately.

Characteristic	**Children Born VP/VLBW**	**Controls**
	(n = 136) n (%)	(n = 137) n (%)
Gestational age (weeks); M (SD), (min., max)	28.9 (2.7) (23.0, 35.9)	40.2 (1.2) (37.1, 42.3)
Birth weight (grams); M (SD) (min., max)	1116 (303) (400, 1820)	3663 (442) (2830, 4980)
Small for gestational age [a],	39 (29)	0
Prenatal corticosteroids	129 (95)	–
Multiple birth	49 (36)	0
Male	83 (61)	67 (49)

Table 1. *Cont.*

Characteristic	Children Born VP/VLBW	Controls
Bronchopulmonary dysplasia	22 (8)	–
Laser-treated retinopathy of prematurity	4/127 [d] (3),	–
Neurodevelopmental impairment	13 (10)	0
Mental Developmental Index <70	3/134 [d] (2)	0
Cerebral palsy	9 (7)	0
Hearing impairment (threshold >40)	4 (3)	0
Visual impairment	0	0
Brain pathology, MRI at term age [b]		–
Normal finding or minor abnormality	94/135 [d] (69)	–
Major abnormality	41/135 [d] (30)	–
Maternal education [c]		
High	64/127 [d] (47)	43 (31)
Intermediate	52/127 [d] (38)	70 (51)
Low	11/127 [d] (8)	24 (18)
Paternal education [c]		
High	36/126 [d] (27)	36 (26)
Intermediatel	80 (59)	72 (53)
Low	10 (7)	29 (21)

[a] Small for gestational age was defined as a birth weight < −2.0 SD, according to the age- and gender-specific Finnish growth charts. [b] See specific magnetic resonance imaging (MRI) protocol and details about the classification elsewhere [29]. [c] High is defined as a Bachelor's degree, Master's degree, or Doctoral degree; Intermediate is defined as high school or vocational school; Low is defined as primary or lower secondary or less. [d] The percentages were calculated from the data available. VP/VLBW = very preterm (<32 gestational weeks)/very low birth weight (≤1500 g).

Descriptive statistics for language and literacy variables were measured and group comparisons are presented in Table 2. A statistically significant difference between the groups was found in every language (p-values from 0.04 to < 0.001) and literacy variable (p-values from 0.002 to 0.003). When children with neurodevelopmental impairment were excluded, the group differences remained statistically significant in ELS and in every literacy variable.

Table 2. Descriptive statistics and group comparisons for the language variables at 2 years of age and literacy variables at 7 years of age for all VP/VLBW children, for VP/VLBW children without neurodevelopmental impairment (NDI), and for full-term controls.

	VP/VLBW Children	**Controls**	**Group Comparison for the Mean**		**VP/VLBW Children**	**Controls**
Measure	Mean (SD) (min, max)	Mean (SD) (min, max)	95% CI	p	Weak skills *n* (%)	Weak skills *n* (%)
Language at 2 years						
Lexicon size	236 (159) (4, 574)	281 (164) (9, 581)	5.87 to 82.76	0.017	21 (15%)	6 (4%)
M3L	5 (3) (1, 14)	6 (4) (1, 21)	0.02 to 1.61	0.036	26 (21%)	18 (14%)
ELS	9 (5) (0, 15)	11 (5) (0, 15)	0.86 to 3.31	<0.001	21 (17%)	13 (10%)
Literacy skills at 7 years						
Reading precursors	30 (8) (3, 38)	33 (6) (11, 38)	1.08 to 4.56	0.002	27 (20%)	13 (10%)
Reading	4 (4) (0, 10)	6 (4) (0, 10)	0.55 to 2.57	0.003	49 (36%)	25 (18%)
Writing	3 (4) (0, 13)	4 (4) (0, 13)	0.70 to 2.65	0.001	64 (48%)	45 (33%)
VP/VLBW children without NDI						
Language at 2 years						
Lexicon size	247 (155) (4, 574)		−5.08 to 73.1	0.087	16 (13%)	
M3L	5 (3) (1, 14)		−0.20 to 1.43	0.138	20 (18%)	

Table 2. *Cont.*

	VP/VLBW Children	Controls	Group Comparison for the Mean		VP/VLBW Children	Controls
ELS	9 (5) (0, 15)		0.57 to 3.05	0.004	17 (15%)	
Literacy skills at 7 years						
Reading precursors	31 (8) (4, 38)		0.64 to 4.06	0.007	23 (19%)	
Reading	4 (4) (0, 10)		0.42 to 2.50	0.006	36 (44%)	
Writing	3 (4) (0, 13)		0.59 to 2.60	0.002	60 (49%)	

VP/VLBW = very preterm (<32 gestational weeks)/very low birth weight (≤1500 g); NDI = Neurodevelopmental Impairment; *(SD)* = Standard Deviation; (min, max) = minimum and maximum; *n* = number; (%) = percentages; 95% CI = Confidence Interval for the mean; *p* = significance level; M3L = mean length of the three longest utterances value; ELS = Expressive Language Score.

3.2. Associations between Language Development at 2 Years of Corrected Age and Literacy Skills at 7 Years

Statistically significant positive correlations were found in both groups between all variables measured at 2 and 7 years (r-values between 0.29 and 0.43, $p < 0.01$) (Table 3). When children with neurodevelopmental impairment were excluded, the correlations remained statistically significant. The r-values were slightly smaller for precursors of reading but remained the same or even slightly increased in reading and writing.

Table 3. Pearson correlation coefficient values (r-values) between language measures at 2 years and literacy measures at 7 years of age for all VP/VLBW children, for VP/VLBW children without NDI, and for the full term controls.

	7 y		
	Reading Precursors	**Reading Skills**	**Writing Skills**
2 y			
Children born VP/VLBW			
Lexicon size	0.37 **	0.40 **	0.33 **
M3L	0.43 **	0.41 **	0.31 **
ELS	0.36 **	0.39 **	0.29 **
Children born VP/VLBW without NDI			
Lexicon size	0.33 **	0.42 **	0.34 **
M3L	0.39 **	0.43 **	0.32 **
ELS	0.30 **	0.39 **	0.31 **
Controls			
Lexicon size	0.32 **	0.39 **	0.32 **
M3L	0.34 **	0.39 **	0.36 **
ELS	0.29 **	0.38 **	0.33 **

** Correlation is significant at the 0.01 level (2-tailed). VP/VLBW = very preterm (<32 gestational weeks)/very low birth weight (≤1500 g); NDI = neurodevelopmental impairment; y = years.

Based on the cross-tabulation, 33% to 74% of VP/VLBW children who had weak early language development (10th percentile) had also weak literacy skills at 7 years (see Table 4). The corresponding proportions for VP/VLBW children with typical language development at 2 years were 11% to 44%. In the controls, the corresponding proportions for children with weak language at 2 years were 15% to 83%. However, the results of the cross-tabulation were statistically significant only between weak lexicon size and weak reading skills and between weak M3L and weak reading and writing skills.

Table 4. Results of the cross-tabulation with Chi-square or Fisher's exact tests of the associations between weak lexicon size (10th percentile, <30 words), weak M3L (10th percentile, <2.06), weak ELS (10th percentile, <1.60) measured at 2 years, and weak reading precursors (10th percentile, <25), weak reading (<10th percentile, 0 words), and weak writing (10th percentile, 0 words) measured at 7 years. Results for all VP/VLBW children, for VP/VLBW children without NDI, and for full-term controls are presented.

	Weak Reading Precursors at 7 years					
	Children born VP/VLBW		Children without NDI		Controls	
Measured at 2 years	n (%)	*p*	n (%)	*p*	n (%)	*p*
Lexicon size		0.001		0.039		0.101
Weak	10 (48%)		6 (38%)		2 (33%)	
Normal	17 (15%)		17 (16%)		11 (8%)	
M3L		<0.001		<0.001		0.058
Weak	13 (50%)		9 (45%)		4 (22%)	
Normal	11 (11%)		11 (12%)		8 (7%)	
ELS		0.037		0.169		0.325
Weak	7 (33%)		5 (29%)		2 (15%)	
Normal	16 (15%)		15 (15%)		10 (8%)	
	Weak reading skills at 7 years					
Lexicon size		0.007		0.068		0.301
Weak	13 (62%)		9 (56%)		2 (33%)	
Normal	36 (31%)		35 (33%)		23 (18%)	
M3L		0.009		0.058		0.013
Weak	14 (54%)		10 (50%)		7 (39%)	
Normal	27 (27%)		26 (28%)		17 (15%)	
ELS		0.08		0.101		0.253
Weak	11 (51%)		9 (53%)		4 (31%)	
Normal	34 (32%)		32 (32%)		20 (16%)	
Lexicon size		0.019		0.027		0.019
Weak	14 (74%)		12 (75%)		5 (83%)	
Normal	50 (44%)		48 (45%)		40 (31%)	
M3L		0.005		0.004		<0.001
Weak	17 (71%)		15 (75%)		14 (78%)	
Normal	39 (39%)		37 (40%)		28 (24%)	
ELS		0.028		0.019		0.268
Weak	14 (70%)		13 (77%)		6 (46%)	
Normal	46 (44%)		44 (44%)		38 (31%)	

M3L = mean length of the three longest utterances value; ELS = Expressive Language Score; VP/VLBW = very preterm (<32 gestational weeks)/very low birth weight (≤1500 g); NDI = neurodevelopmental impairment.

The regression models of the VP/VLBW group, including early lexicon size as a predictor, explained 23% of the variance in reading precursors, 27% of the variance in reading skills, and 17% of the variance in writing skills at 7 years (Table 5). In these models, early lexicon size and paternal education were statistically significant independent predictors. The regression models of the controls are presented in Appendix A.

Table 5. Results of multiple variable linear regression analysis with reading precursors, reading and writing skills at 7 years of age as dependent variables, and with lexicon size at 2 years of corrected age and background factors as independent variables. Results of VP/VLBW children are presented.

	Reading Precursors			Reading			Writing		
	b	95% CI	*p*	*b*	95% CI	*p*	*b*	95% CI	*p*
Gestational age	−0.05	−0.66 to 0.38	0.589	0.04	−0.20 to 0.32	0.642	0.03	−0.20 to 0.27	0.764
Reading difficulties									
Mothers	0.15	−0.40 to 9.50	0.071	0.10	−0.93 to 4.01	0.223	0.07	−1.29 to 3.25	0.462
Fathers	−0.07	−5.63 to 2.39	0.432	−0.02	−2.28 to 1.73	0.791	−0.07	−2.73 to 0.95	0.412
Paternal education	0.25	1.47 to 7.56	0.004	0.31	1.42 to 4.46	<0.001	0.20	0.26 to 3.05	0.024
Lexicon size	0.31	0.01 to 0.03	0.001	0.32	0.004 to 0.01	<0.001	0.30	0.002 to 0.01	0.001
Fit statistics									
F	7.0			9.0			5.0		
P for F	<0.001			<0.001			0.001		
R^2	0.23			0.27			0.17		
ΔR^2	0.20			0.24			0.13		

VP/VLBW = very preterm (<32 gestational weeks)/very low birth weight (≤1500 g); F = value of F-statistic; *p* = significance value; R^2 = coefficient of determination; ΔR^2 = adjusted coefficient of determination.

The models including M3L as a predictor explained 27% of the variance in reading precursors, 28% of the variance in reading skills, and 16% of the variance in writing skills at 7 years (Table 6). M3L and paternal education were statistically significant independent predictors.

Table 6. Results of multiple variable linear regression analysis with reading precursors, reading, and writing skills at 7 years of age as dependent variables, and with M3L at 2 years of corrected age and background factors as independent variables. Results of VP/VLBW children are presented.

	Reading Precursors			Reading			Writing		
	b	95% CI	*p*	*b*	95% CI	*p*	*b*	95% CI	*p*
Gestational age	−0.06	−0.72 to 0.34	0.484	0.03	−0.22 to 0.32	0.711	0.03	−0.21 to 0.29	0.762
Reading difficulties									
Mothers	0.12	−1.47 to 8.67	0.163	0.08	−1.35 to 3.82	0.351	0.07	−1.49 to 3.30	0.472
Fathers	−0.06	−5.51 to 2.63	0.491	−0.02	−2.28 to 1.87	0.842	−0.08	−2.78 to 1.07	0.294
Paternal education	0.25	1.50 to 7.66	0.004	0.32	1.44 to 4.57	<0.001	0.21	0.27 to 3.18	0.021
M3L	0.37	0.57 to 1.54	<0.001	0.35	0.25 to 0.75	<0.001	0.26	0.08 to 0.55	0.008
Fit statistics									
F	8.0			8.5			4.0		
P for F	<0.001			<0.001			0.003		
R^2	0.27			0.28			0.16		
ΔR^2	0.23			0.25			0.12		

M3L = mean length of the three longest utterances value; VP/VLBW = very preterm (<32 gestational weeks)/very low birth weight (≤1500 g); F = value of F-statistic; *p* = significance value; R^2 = coefficient of determination; ΔR^2 = adjusted coefficient of determination.

The models including ELS as a predictor (Table 7) explained 20% of the variance in reading precursors, 25% of the variance in reading skills, and 14% of the variance in writing skills at 7 years. ELS, paternal education, and mother's self-reported reading difficulties were statistically significant independent predictors.

Table 7. Results of multiple variable linear regression analysis with reading precursors, reading, and writing skills at 7 years of age as dependent variables, and with ELS at 2 years of corrected age and background factors as independent variables. Results of VP/VLBW children are presented.

	Reading Precursors			**Reading**			**Writing**		
	b	95% CI	*p*	*b*	95% CI	*p*	*b*	95% CI	*p*
Gestational age	−0.01	−0.58 to 0.50	0.901	0.07	−0.15 to 0.39	0.401	0.06	−0.17 to 0.33	0.531
Reading difficulties									
Mothers	0.18	0.24 to 10.48	0.039	0.13	−0.58 to 4.56	0.132	0.11	−0.98 to 3.72	0.253
Fathers	−0.06	0.5.75 to 2.71	0.501	−0.02	−2.33 to 1.91	0.824	−0.08	−2.80 to 1.08	0.384
Paternal education	0.20	0.46 to 7.10	0.029	0.27	0.88 to 4.19	0.003	0.18	−0.09 to 2.93	0.069
ELS	0.26	0.13 to 0.72	0.005	0.28	0.08 to 0.38	0.002	0.21	0.01 to 0.28	0.029
Fit statistics									
F	6.0			7.0			3.4		
P for F	<0.001			<0.001			0.007		
R^2	0.20			0.25			0.14		
ΔR^2	0.17			0.21			0.10		

ELS = Expressive Language Score; VP/VLBW = very preterm (<32 gestational weeks)/very low birth weight (≤1500 g); F = value of F-statistic; *p* = significance value; R^2 = coefficient of determination; ΔR^2 = adjusted coefficient of determination.

The exclusion of children with neurodevelopmental impairment did not alter the results. In the control group (Appendix A), the same models explained a smaller proportion of the variance in outcome relative to children born VP/VLBW.

In VP/VLBW group, the AUC values of language variables at 2 years regarding literacy skills at 7 years varied between 0.70 and 0.77 ($p < 0.001$) (Table 8). Exclusion of children with neurodevelopmental impairment did not significantly alter the results. In controls, the values varied between 0.62 and 0.73 (*p*-values from 0.18 to < 0.001), respectively.

Table 8. Area under the receiver operating characteristic (ROC) curve (AUC) values for weak reading precursors (sum score < 25, 10th percentile), weak reading (0 words, 10th percentile), and weak writing (0 words, 10th percentile) skills at 7 years with lexicon size/M3L/ELS at 2 years as predictor variables.

	AUC Value of Lexicon Size	**95% CI**	***p***
Children born VP/VLBW			
Reading precursors	0.70	0.58 to 0.83	0.001
Reading	0.72	0.63 to 0.81	<0.001
Writing	0.72	0.64 to 0.80	<0.001
Controls			
Reading precursors	0.65	0.50 to 0.80	0.081
Reading	0.67	0.56 to 0.78	0.009
Writing	0.69	0.60 to 0.78	<0.001
Children born VP/VLBW			
Reading precursors	0.77	0.66 to 0.89	<0.001
Reading	0.74	0.65 to 0.83	<0.001
Writing	0.73	0.64 to 0.82	<0.001
Controls			
Reading precursors	0.73	0.58 to 0.87	0.009
Reading	0.65	0.51 to 0.78	0.029
Writing	0.73	0.64 to 0.82	<0.001
Children born VP/VLBW			
Reading precursors	0.72	0.61 to 0.82	0.001
Reading	0.71	0.62 to 0.80	<0.001
Writing	0.74	0.65 to 0.83	<0.001

Table 8. *Cont.*

	AUC Value of Lexicon Size	95% CI	*p*
Controls			
Reading precursors	0.62	0.44 to 0.80	0.182
Reading	0.64	0.52 to 0.77	0.029
Writing	0.71	0.62 to 0.80	<0.001

AUC values are interpreted as follows: excellent predictive value 0.90–1, good 0.80–0.90, fair 0.70–0.80, poor 0.60–0.70, and fail < 0.60 [28]. AUC = Area Under the ROC Curve; M3L = mean length of the three longest utterances value; ELS = Expressive Language Score; VP/VLBW = very preterm (<32 gestational weeks)/very low birth weight (≤1500 g); NDI = neurodevelopmental impairment.

4. Discussion

To the best of our knowledge, this is the first controlled follow-up study providing longitudinal information on the associations between very early language development at 2 years of corrected age and later literacy skills in VP/VLBW children and their controls. Significant correlations between every language and literacy variable were found both in the VP/VLBW group and in the control group. Most of the children born VP/VLBW with weak language skills at 2 years had also weak literacy skills at 7 years. Lexicon size, M3L, and ELS measured at 2 years were statistically significant predictors in the regression models explaining the variance in literacy skills, especially in the VP/VLBW group. Every language variable at 2 years had a fair predictive value for literacy skills 5 years later in children born VP/VLBW when measured using AUC values.

Previously, the associations between language and literacy ability in children born preterm have been analyzed at 4 years of age at the earliest [17]. In a longitudinal study consisting of 110 children born VP and 113 term controls, Pritchard and colleagues [17] found an association between school readiness domains including language at age 4 years, and literacy and numeracy skills at ages 6 and 9 years. In addition, in the study of Perez-Pereira et al. [30], morphosyntactic production and comprehension of syntactic structures at 5 years were associated with reading outcome at 9 years in preterm children. The knowledge of letters and words at 5 years [31] and phonological awareness and expressive and receptive language at 6 years [16] have been found to be associated with reading and writing outcome at 7 years [31] and at 8 years [16] in VP populations. In two previous cross-sectional studies [14,15], reading performance at 8 years in children born VP was correlated with lexical production and grammar comprehension [14] and with phonological awareness and rapid naming [15]. To date, it has been unclear whether an association between very early language development and later literacy outcome can be detected in VP/VLBW children. Our findings fill in this gap, suggesting that the association between language and literacy at 7 years of age is evident already at age 2 years in children born VP/VLBW.

In the present cohort, most children born VP/VLBW with small lexicon size, short M3L, and weak ELS at 2 years had weak literacy skills at 7 years compared with those with more advanced early language. In previous studies considering the continuum of language in children born VP/VLBW small lexicon size and short utterance length at 2 years have been shown to predict later language development [3,4,10,12]. Our study extends this knowledge to literacy skills. These results together emphasize the need for early screening of weak language development in the vulnerable group of VP/VLBW children

Another novel finding was that language skills at 2 years explained a significant amount of the variance in literacy skills 5 years later, especially in the VP/VLBW group. Thus, the present findings provide evidence for the existence of a continuum between language development at 2 years and literacy ability at 7 years in these children. This study offers an interesting perspective on the association of early language with later literacy skills in preterm children using the Finnish language, which has a highly regular grapheme–phoneme correspondence. Different language versions of the CDI have been shown to be a cost-effective way of identifying small lexicon and short utterance length at 2 years [2–4,12]. In this study, the variables of the FinCDI were even stronger predictors for later literacy skills than ELS, which is a performance-based subtest of BSID-II [24]. Thus, our results support the view that parents can provide valuable information on early language development of their children, when structured, validated measure, such as CDI, is used.

Paternal education was a significant background variable in the regression models, especially in children born VP/VLBW. For the controls, the effect of paternal education was not as clear. The effects of paternal education are less studied than those of maternal education. However, in previous studies regarding the same PIPARI cohort [22,23], paternal education was found to relate also to precursors of reading at 5 years [5], and to verbal comprehension at 11 years [32] in children born VP/VLBW. Our findings suggest that fathers may have a significant role in supporting the language development of preterm children in the home environment during childhood years, at least in societies which emphasize the role of both parents in early childhood care, as in Finland.

In this study, lexicon size, M3L, and ELS measured at 2 years had fair predictive value (AUC values varied between 0.70 and 0.77, $p < 0.001$) for literacy skills 5 years later in the VP/VLBW group. The explaining value of early language at 2 years of age for literacy performance at school age has been investigated previously in full-term populations, e.g., in children with a familial risk of dyslexia [19,20]. Parallel results have been noted for late talkers, i.e., children with small expressive lexicon at 2 years but with an absence of cognitive delay or any other neurological condition explaining the slow language acquisition [33,34]. Late talkers perform consistently lower on language and literacy tasks at school age and even in adolescence than their peers [33–35]. In the present study, the predictive value of early language at 2 years of age for literacy skills at 7 years was established for the first time in the vulnerable population of preterm children born at very early gestational age (<32) and/or with very low birth weight ($\leq$1500 g). Comparable findings detected in different populations support the view that very small lexicon size and/or very short utterance length at 2 years of age are risk factors for later language and literacy deficits after controlling for background factors. Furthermore, the predictive value of early language skills for later literacy outcome was better in children born VP/VLBW than in controls. This finding may be explained by the fact that the VP/VLBW sample included more children with early weak language skills relative to controls [3]. These results may also reflect the persistence of language-related difficulties among children born VP/VLBW with early weak language.

This study has several implications. First, it shows very clear longitudinal associations between very early language skills and later literacy outcome in preterm children. Thus, our findings propose the clinical importance of screening language skills of preterm children born at very early gestational age and/or with very low birth weight at 2 years of corrected age. Our findings highlight the usefulness of the validated parent-reported form, such as CDI [13,25] in the follow-up of the vulnerable group of high-risk prematurely born VP/VLBW children for identifying children at risk for later literacy deficits. Identifying developmental problems as early as possible is important, since it enables targeted early interventions and support. In addition, standardized parental report forms, such as CDI, may promote parents' active involvement in observing and supporting the language development of their preterm-born child. Our results provide information also for the educational professionals working with school-aged children born VP/VLBW showing the higher percentage of weak pre-reading, reading and writing skills in this population when compared with full-term control children.

Strengths of the study include its longitudinal design with a well-defined cohort of children born VP/VLBW together with a control group born in the same hospital. The longitudinal data from altogether 274 children provided a great possibility to assess the associations between early language development and later literacy performance. Both a validated parent-report form [13,25] and a test-based measure [24] were used to gather information on early lexical and grammatical development. The use of different types of method to assess early language development strengthened our findings. In our study, the phonological awareness task, which included three to seven-letter words said aloud phoneme by phoneme, also relates to working memory and actually measures both domains. However, the participants also had visual aid: at the same time as they heard the phonemes, they saw pictures of the correct word and three other alternatives. The participants had to mark one picture out of four alternatives they thought would best match the word. This might have reduced the burden of the working memory during the task. As a limitation, measures used in the study provided information on expressive language only. Information on receptive language would have provided an even more comprehensive view of early language development. This should be taken into consideration when applying these results to a clinical context.

5. Conclusions

Language development is essential for academic learning of children starting school. It is important to recognize potential risks for learning disorders as early as possible. Our study shows, for the first time, that problems in literacy skills at the beginning of formal schooling at 7 years of age may be identified already at age 2 years in preterm children born at very early gestational age and/or with very low birth weight. Early identification enables early interventions for those preterm children at risk for later literacy deficits. If concern arises regarding the early language ability of preterm children based on the results of a parental report form, such as CDI, a broader assessment of language skills by a speech-language pathologist is recommended. We emphasize the need for further studies (randomized controlled trials) regarding effective early interventions for VP/VLBW children at risk for literacy deficits.

Author Contributions: Conceptualization, E.J., P.M., H.L. and S.K.J.S.; methodology, E.J., P.M., S.S., J.L., M.H., H.L. and S.K.J.S.; investigation, E.J., P.M., S.S., M.H., H.L. and S.K.J.S.; formal analysis, E.J. and J.L.; writing—original draft preparation, E.J.; writing—review and editing, E.J., P.M., S.S., J.L., M.H., H.L. and S.K.J.S.; supervision, P.M. and S.K.J.S.; project administration, H.L. and S.K.J.S. All authors have read and agreed to the published version of the manuscript.

Funding: This study was funded by the Doctoral Programme in Psychology, Learning and Communication, University of Helsinki, Finland.

Institutional Review Board Statement: The PIPARI study was conducted according to the guidelines of the Declaration of Helsinki, and the study protocol was approved by the Ethics Review Committee of the Hospital District of Southwest Finland (code: 2000/12, 19/12/2000, §333, 12/06/2001, §175).

Informed Consent Statement: Informed consent was obtained from all parents who agreed to participate in the study after they had received written and oral information.

Data Availability Statement: This manuscript is based on health data. Access to these data is regulated by Finnish legislation and Findata, the Health and Social Data Permit Authority. The disclosure of data to third parties without explicit permission from Findata is prohibited. Only those fulfilling the requirements established by Finnish legislation and Findata for viewing confidential data are able to access the data. See https://www.findata.fi/en/about-us/data-protection-and-the-processing-of-personal-data/ (accessed on 29 May 2021).

Acknowledgments: We thank for the following persons in the PIPARI Study Group: Satu Ekblad; Eeva Ekholm; Linda Grönroos; Leena Haataja; Pentti Kero; Riikka Korja; Katri Lahti; Liisa Lehtonen; Tuomo Lehtonen; Marika Leppänen; Annika Lind; Mira Mattsson, study coordinator; Jonna Maunu; Laura Määttänen; Pekka Niemi; Anna Nyman; Riitta Parkkola; Liisi Ripatti; Päivi Rautava; Katriina Saarinen, physiotherapist; Tiina Saarinen; Susanna Salomäki; Virva Saunavaara; Matti Sillanpää; Päivi Tuomikoski-Koiranen; Timo Tuovinen; Karoliina Uusitalo; Anniina Väliaho; Milla Ylijoki. Open access funding provided by the University of Helsinki.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in study design; in collection, analyses, or interpretation of data; in writing of the manuscript; or in the decision to publish the results.

Appendix A

Table A1. Results of multiple variable linear regression analysis with reading precursors, reading, and writing skills at 7 years of age as dependent variables, and with lexicon size at 2 years of age and background factors as independent variables. Results of full-term controls are presented.

	Reading Precursors			**Reading**			**Writing**		
	b	95% CI	*p*	*b*	95% CI	*p*	*b*	95% CI	*p*
Gestational age	0.01	−0.87 to 0.94	0.94	0.02	−0.53 to 0.65	0.81	−0.04	−0.78 to 0.52	0.73
Reading difficulties									
Mothers	0.08	−2.67 to 3.90	0.38	0.06	−0.85 to 3.43	0.45	0.13	−0.58 to 4.12	0.22
Fathers	−0.10	−6.62 to 1.95	0.25	−0.02	−3.21 to 2.36	0.80	−0.04	−3.75 to 2.38	0.66
Paternal education	0.06	−1.36 to 3.45	0.47	0.19	0.38 to 3.51	0.029	0.12	−0.50 to 2.94	0.26
Lexicon size	0.38	0.01 to 0.19	<0.001	0.39	0.01 to 0.15	<0.001	0.34	0.005 to 0.01	<0.001
Fit statistics									
F	4.2			5.3			3.3		
P for F	0.001			<0.001			0.008		
R^2	0.15			0.19			0.12		
ΔR^2	0.12			0.15			0.09		

Table A2. Results of multiple variable linear regression analysis with reading precursors, reading, and writing skills at 7 years of age as dependent variables, and with M3L at 2 years of age and background factors as independent variables. Results of full-term controls are presented.

	Reading Precursors			**Reading**			**Writing**		
	b	95% CI	*p*	*b*	95% CI	*p*	*b*	95% CI	*p*
Gestational age	0.005	−0.90 to 0.94	0.95	0.02	−0.54 to 0.67	0.62	−0.04	−0.80 to 0.51	0.66
Reading difficulties									
Mothers	0.03	−2.80 to 3.80	0.77	0.09	−1.02 to 3.33	0.63	0.12	−0.66 to 4.03	0.16
Fathers	−0.10	−6.70 to 1.90	0.27	−0.03	−3.32 to 2.37	0.72	−0.04	−3.80 to 2.32	0.63
Paternal education	0.05	−1.80 to 3.04	0.61	0.17	0.02 to 3.22	0.04	0.09	−0.82 to 2.62	0.30
M3L	0.34	0.30 to 0.91	<0.001	0.38	0.26 to 0.67	<0.001	0.36	0.24 to 0.68	<0.001
Fit statistics									
F	3.4			5.6			4.1		
P for F	0.006			<0.001			0.002		
R^2	0.13			0.19			0.16		
ΔR^2	0.09			0.16			0.12		

M3L = mean length of the three longest utterances value.

Table A3. Results of multiple variable linear regression analysis with reading precursors, reading, and writing skills at 7 years of age as dependent variables, and with ELS at 2 years of age and background factors as independent variables. Results of full-term controls are presented.

	Reading Precursors			Reading			Writing		
	b	95% CI	*p*	*b*	95% CI	*p*	*b*	95% CI	*p*
Gestational age	0.01	−0.86 to 0.99	0.88	0.02	−0.52 to 0.68	0.80	−0.03	−0.79 to 0.53	0.70
Reading difficulties									
Mothers	0.04	−2.61 to 4.15	0.65	0.11	−0.72 to 3.65	0.38	0.14	−0.40 to 4.34	0.10
Fathers	−0.09	−6.64 to 2.13	0.31	−0.02	−3.19 to 2.48	0.81	−0.04	−0.370 to 2.50	0.68
Paternal education	0.06	−1.60 to 3.32	0.49	0.19	0.20 to 3.38	0.03	0.11	−0.70 to 2.80	0.22
ELS	0.29	0.14 to 0.59	0.002	0.38	0.19 to 0.48	<0.001	0.35	0.16 to 0.47	<0.001
Fit statistics									
F	2.5			5.5			3.7		
P for F	0.04			<0.001			0.004		
R^2	0.09			0.20			0.14		
ΔR^2	0.06			0.16			0.10		

ELS = Expressive Language Score.

References

1. Foster-Cohen, S.; Edgin, J.O.; Champion, P.R.; Woodward, E.J. Early delayed language development in very preterm infants: Evidence from the MacArthur-Bates CDI. *J. Child Lang.* **2007**, *34*, 655–675. [CrossRef] [PubMed]
2. Sansavini, A.; Guarini, A.; Justice, L.M.; Savini, S.; Broccoli, S.; Alessandroni, R.; Faldella, G. Does preterm birth increase a child's risk for language impairment? *Early Hum. Dev.* **2010**, *86*, 765–772. [CrossRef] [PubMed]
3. Stolt, S.; Matomäki, J.; Lind, A.; Lapinleimu, H.; Haataja, L.; Lehtonen, L. The prevalence and predictive value of weak language skills in children with very low birth weight—A longitudinal study. *Acta Paediatr* **2014**, *103*, 651–658. [CrossRef] [PubMed]
4. Twilhaar, E.S.; de Kieviet, J.F.; Aarnoudse-Moens, C.S.H.; van Elburg, R.M.; Oosterlaan, J. Academic performance of children born preterm: A meta-analysis and meta-regression. *Arch. Dis. Child. Fetal Neonatal Ed.* **2018**, *103*, F322–F330. [CrossRef] [PubMed]
5. Munck, P.; Niemi, P.; Väliaho, A.; Lapinleimu, H.; Lehtonen, L.; Haataja, L. Prereading skills of very-low-birth-weight prematurely born Finnish children. *Child Neuropsychol.* **2012**, *18*, 92–103. [CrossRef]
6. Guarini, A.; Bonifacci, B.; Tobia, V.; Alessandroni, R.; Faldella, G.; Sansavini, A. The profile of very preterm children on academic achievement. A cross-population comparison with children with specific learning disorders. *Res. Dev. Disabil.* **2019**, *87*, 54–63. [CrossRef]
7. Alanko, O.; Niemi, P.; Munck, P.; Matomäki, J.; Turunen, T.; Nurmi, J.-E.; Lehtonen, L.; Haataja, L.; Rautava, P. Reading and math abilities of Finnish school beginners born very preterm or with very low birth weight. *Learn Individ. Differ.* **2017**, *54*, 173–183. [CrossRef]
8. van Noort-van der Spek, I.; Franken, M.-C.; Weisglas-Kuperus, N. Language Functions in Preterm-Born Children: A Systematic Review and Meta-analysis. *Pediatrics* **2012**, *129*, 745–754. [CrossRef]
9. Kovachy, V.N.; Adams, J.N.; Tamaresis, J.S.; Feldman, H.M. Reading abilities in school-aged preterm children: A review and meta-analysis. *Dev. Med. Child Neurol.* **2015**, *57*, 410–419. [CrossRef]
10. Nguyen, T.-N.-N.; Spencer-Smith, M.; Zannino, D.; Burnett, A.; Scratch, S.E.; Pascoe, L.; Ellis, R.; Cheong, J.; Thompson, D.; Inder, T.; et al. Developmental trajectory of language from 2 to 13 years in children born very preterm. *Pediatrics* **2018**, *141*, e20172831. [CrossRef]
11. Putnick, D.L.; Bornstein, M.H.; Eryigit-Madzwamuse, S.; Wolke, D. Long-Term Stability of Language Performance in Very Preterm, Moderate-Late Preterm, and Term Children. *J. Pediatr.* **2017**, *181*, 74–79. [CrossRef]
12. Charkaluk, M.-L.; Rousseau, J.; Benhammou, V.; Datin-Dorrière, V.; Flamant, C.; Gire, C.; Kern, S.; Pierrat, V.; Kaminski, M.; Marret, S. Association of Language Skills with Other Developmental Domains in Extremely, Very, and Moderately Preterm Children: EPIPAGE 2 Cohort Study. *J. Pediatr.* **2019**, *208*, 114–120. [CrossRef]
13. Fenson, L.; Marchman, V.A.; Thal, D.J.; Dale, P.S.; Reznick, J.S.; Bates, E. MacArthur-Bates Communicative Development Inventories. In *User's Guide and Technical Manual*, 2nd ed.; Paul H. Brookes Publisihing Co.: Baltimore, MD, USA, 2007.
14. Guarini, A.; Sansavini, A.; Fabbri, C.; Savini, S.; Alessandroni, R.; Faldella, G.; Karmiloff-Smith, A. Long-term effects of preterm birth on language and literacy at eight years. *J. Child Lang.* **2010**, *37*, 865–885. [CrossRef]
15. Wocadlo, C.; Rieger, I. Phonology, rapid naming and academic achievement in very preterm children at eight years of age. *Early Hum. Dev.* **2007**, *83*, 367–377. [CrossRef]

16. Brochers, L.R.; Bruckert, L.; Travis, K.E.; Dodson, C.K.; Loe, I.M.; Marchman, V.A.; Feldman, H.M. Predicting text reading skills at age 8 years in children born preterm and at term. *Early Hum. Dev.* **2019**, *130*, 80–86. [CrossRef]
17. Pritchard, V.E.; Samudragupta, B.; Austin, N.C.; Levin, K.J.; Woodward, L.J. Identifying Very Preterm Children at Educational Risk Using a School Readiness Framework. *Pediatrics* **2014**, *134*, e825–e832. [CrossRef]
18. Holopainen, L.; Ahonen, T.; Lyytinen, H. Predicting Delay in Reading Achievement in a Highly Transparent Language. *J. Learn. Disabil.* **2001**, *34*, 401–413. [CrossRef]
19. Niemi, P.; Nurmi, J.-E.; Lyyra, A.-L.; Lerkkanen, M.-K.; Lepola, J.; Poskiparta, E.; Poikkeus, A.-M. Task avoidance, number skills and parental learning difficulties as predictors of poor response to instruction. *J. Learn. Disabil.* **2011**, *44*, 459–471. [CrossRef]
20. Torppa, M.; Lyytinen, P.; Erskine, J.; Eklund, K.; Lyytinen, H. Language Development, Literacy Skills, and Predictive Connections to Reading in Finnish Children with and Without Familial Risk for Dyslexia. *J. Learn. Disabil.* **2010**, *43*, 308–321. [CrossRef]
21. Lyytinen, P.; Lyytinen, H. Growth and predictive relations of vocabulary and inflectional morphology in children with and without familial risk for dyslexia. *Appl. Psycholinguist.* **2004**, *25*, 397–411. [CrossRef]
22. PIPARI. Development and Functioning of very Low Birth Weight Infants from Infancy to School Age. Available online: https://sites.utu.fi/pipari/en/ (accessed on 19 May 2021).
23. Setänen, S.; Lehtonen, L.; Lapinleimu, H.; Haataja, L. Lessons learnt about the long-term neurodevelopment in very preterm born children in the PIPARI Study. *Lääketieteellinen Aikakausk Duodecim* **2018**, *134*, 118–125.
24. Bayley, N. *Bayley Scales of Infant Development*, 2nd ed.; Psychological Corporation: San Antonio, TX, USA, 1993.
25. Lyytinen, P. *Varhaisen Kommunikaation ja Kielen Kehityksen Arviointimenetelmä*, The adapted Finnish long-form version of the MacArthur-Bates Communicative Development Inventories; Niilo Mäki–Instituutti: Jyväskylä, Finland, 1999.
26. Poskiparta, E. *Miten Ehkäisen Oppilaitteni Luku- ja Kirjoitusvaikeuden Syntymisen? Kielellisen Tietoisuuden Kehittäminen 1. Luokalla Ryhmämuotoisena Erityisopetuksena*; Developing linquistic awareness in Grade 1; University of Turku, Centre for Learning Research: Turku, Finland, 1995.
27. Lerkkanen, M.K.; Poikkeus, A.M.; Ketonen, R. *ARMI–Luku- ja Kirjoitustaidon Arviointimateriaali 1. Luokalle*; ARMI–A tool for assessing reading and writing skills in Grade 1; WSOY: Helsinki, Finland, 2006.
28. Tape, T.G. The Area under an ROC Curve. Available online: http://gim.unmc.edu/dxtests/ROC3.htm (accessed on 18 January 2021).
29. Setänen, S.; Haataja, L.; Parkkola, R.; Lind, A.; Lehtonen, L. Predictive value of neonatal brain MRI on the neurodevelopmental outcome of preterm infants by 5 years of age. *Acta Paediatr.* **2013**, *102*, 492–497. [CrossRef]
30. Perez-Pereira, M.; Martínez-López, Z.; Maneiro, L. Longitudinal relationships between reading abilities, phonological awareness, language abilities and executive functions: Comparison of low risk preterm and full-term children. *Front. Psychol.* **2020**, *11*, 468. [CrossRef]
31. Taylor, R.; Pascoe, L.; Scratch, S.; Doyle, L.W.; Anderson, P.; Roberts, G. A simple screen performed at school entry can predict academic under-achievement at age seven in children born very preterm. *J. Paediatr. Child Health* **2016**, *52*, 759–764. [CrossRef]
32. Nyman, A.; Korhonen, T.; Munck, P.; Parkkola, R.; Lehtonen, L.; Haataja, L. Factors affecting the cognitive profile of 11-year-old children born very preterm. *Pediatr. Res.* **2017**, *82*, 324–332. [CrossRef] [PubMed]
33. Rescorla, L. Do late-talking toddlers turn out to have reading difficulties a decade later? *Ann. Dyslexia* **2000**, *50*, 87–102. [CrossRef] [PubMed]
34. Rescorla, L. Age 17 Language and reading outcomes in late-talking toddlers: Support for dimensional perspective on language delay. *J. Speech Lang. Hear* **2009**, *52*, 16–30. [CrossRef]
35. Preston, J.L.; Frost, S.J.; Mencl, W.E.; Fulbright, R.K.; Landi, N.; Grigorenko, E.; Jacobsen, L.; Pugh, K.R. Early and late talkers: School-age language, literacy and neurolinguistics differences. *Brain* **2010**, *133*, 2185–2195. [CrossRef]

children

MDPI

Article

Effects of ASL Rhyme and Rhythm on Deaf Children's Engagement Behavior and Accuracy in Recitation: Evidence from a Single Case Design

Leala Holcomb * and Kimberly Wolbers

Jane and David Bailey Education Complex, University of Tennessee, Knoxville, TN 37996, USA; kwolbers@utk.edu
* Correspondence: lholcom5@vols.utk.edu

Received: 31 October 2020; Accepted: 24 November 2020; Published: 26 November 2020

Abstract: Early language acquisition is critical for lifelong success in language, literacy, and academic studies. There is much to explore about the specific techniques used to foster deaf children's language development. The use of rhyme and rhythm in American Sign Language (ASL) remains understudied. This single-subject study compared the effects of rhyming and non-rhyming ASL stories on the engagement behavior and accuracy in recitation of five deaf children between three and six years old in an ASL/English bilingual early childhood classroom. With the application of alternating treatment design with initial baseline, it is the first experimental research of its kind on ASL rhyme and rhythm. Baseline data revealed the lack of rhyme awareness in children and informed the decision to provide intervention as a condition to examine the effects of explicit handshape rhyme awareness instruction on increasing engagement behavior and accuracy in recitation. There were four phases in this study: baseline, handshape rhyme awareness intervention, alternating treatments, and preference. Visual analysis and total mean and mean difference procedures were employed to analyze results. The findings indicate that recitation skills in young deaf children can be supported through interventions utilizing ASL rhyme and rhythm supplemented with ASL phonological awareness activities. A potential case of sign language impairment was identified in a native signer, creating a new line of inquiry in using ASL rhyme, rhythm, and phonological awareness to detect atypical language patterns.

Keywords: deaf; preschool; rhyme; rhythm; recitation; engagement; language processing; phonological awareness; sign language; language deprivation; language impairment; single case study; alternating treatments

1. Introduction

The difference between hearing and deaf toddlers' early language access and experience is stark. From birth, hearing children access the playful phonological patterns of spoken language on a minute-by-minute basis [1]. Hearing children benefit greatly from language activities that incorporate rhythmic and rhyming spoken language (e.g., nursery rhymes or Dr. Seuss), especially if paired with spoken phonological awareness tasks [2–4]. What has become ubiquitous among households, daycare, and early childhood education programs for hearing children is largely absent or inaccessible to deaf children. Deaf children can neither fully nor easily access sound, limiting their ability to fully benefit from spoken rhyme and rhythm. As we will see, early childhood education for deaf children is only just beginning to incorporate American Sign Language (ASL) rhyme and rhythm in the classroom. There are different reasons for this gap between deaf and hearing children; one of these reasons is the lack of comparable research as to the benefit of ASL rhyme and rhythm for deaf children's phonological awareness and language development. This study adds to a growing body of research aimed at addressing this gap.

While there are historical documents on language play in American Sign Language (ASL)—such as on rhyme, rhythm, and poetry in the deaf community—this practice has been excluded from most early childhood classrooms that serve deaf children [5]. The absence of early language approaches that utilize ASL can be explained by the long history of the stronghold of oralism (speaking-only) in Deaf education [6]. The exclusion of deaf teachers meant that their unique cultural and linguistic approaches were also precluded from deaf children's language experiences. It was not until the 1970s that ASL began to be embraced in a few deaf schools. A contributing factor to this change was the recognition of ASL as a true, sophisticated language by linguists [7]. Since then, there is accumulating evidence that deaf children immersed in a signing environment during their early years can meet language milestones, develop robust literacy skills, and excel in academic studies [8–10]. Yet negative biases towards the use of ASL with deaf toddlers have not wavered, especially during the early childhood period [11]. Although tremendous potential is evident, approaches that maximize deaf children's visual access to language have still not fully infiltrated the educational system.

In small pockets of deaf education, some doors have opened for deaf educators and leaders to enter the field and devise methods of language instruction that are culturally and linguistically responsive to deaf ways of learning [12]. ASL rhyme and rhythm were among the approaches that materialized in the ASL/English bilingual classroom [13]. Deaf educators and researchers have recently begun to come together to create new ASL rhyming and rhythmic resources in response to the demands from teachers and families (e.g., ASL Mother Goose Program and Hands Land). We ask whether parallel forms of language intervention using ASL rhyme and rhythm can support engagement and recitation skills in deaf children. This single-subject study compared the effects of rhyming and non-rhyming ASL stories on the engagement behavior and accuracy in recitation of five deaf children between three and six years old in an ASL/English bilingual early childhood classroom.

1.1. Literature Review

Chukovsky [14] documented hearing children's fascination with language and their tendency to make up rhyming words. His own four-year-old child was found spontaneously screaming, "I'm a big, big rider! You're smaller than a spider!" Hearing children are regularly exposed to rhyme and rhythm packed with patterns and repetitions in sound that propel language development. Rhymes are the same sounds produced across different words such as in this verse: "the fat bat sat on the cat's hat". Rhythms are the patterned stressed and unstressed beats that sounds make. The developing human brain is biologically captivated by these patterns and fires neurons to build robust connections [15]. The consistent and abundant input of rhyme and rhythm on a daily basis through incidental exposure in the environment (e.g., restaurants, grocery stores, car rides, television) and explicit instruction (e.g., parentese, daycare centers, early childhood classrooms, interventions) make a significant impact on hearing children's development.

There are sequential steps in language acquisition that follow a timetable. As infants begin to acquire language, they attempt to produce their first sounds, words, and then songs by imitating their caregivers' language input. Rhyme, rhythm, and parentese (baby talk) are among the critical early language approaches utilized for healthy brain development [16]. With rich language input, toddlers' language grows exponentially between two and five years old, enabling them to turn their attention to language itself through the development of phonological awareness. Phonological awareness is the knowledge and understanding of how sound structures work [17]. Adams [18] presented a hierarchical trajectory of phonological awareness development. Following the sequence of development, hearing children first develop sensitivity to and awareness of sounds by recognizing that they can be put together to form words. With this awareness, they begin to enjoy playing with language and reciting nursery rhymes. Next, they learn that some words share similar sounds while others do not. They recognize words that rhyme such as 'cat,' 'fat,' and 'bat'. Then, they are able to put together individual sounds to produce a word. They can say "bat" after being given the individual sounds /b/a/t/. Inversely, they can break down a word into individual sounds by saying /b/a/t/ when given the word 'bat'. Finally, they can manipulate sounds in words through deleting or substituting. If asked to say 'fear' without the /f/, they would be able to say 'ear.' If asked to replace individual sounds to create new rhymes, they would be able to say 'bat,' and 'sat,' or 'sit' and 'sip.' With these language opportunities, children often enjoy playing with sounds by singing out aloud by themselves or with each other—dee, dee, bee, bee, tee, tee, rub, rub, dub, dub. With a strong foundation in language, hearing children are primed to navigate the challenges and complexity of literacy development.

A meta-analysis by the National Early Reading Panel [19] concluded that access to lexical store, phonological memory, and phonological awareness during early years is a good predictor for literacy success. This finding signifies the importance of using approaches that maximize young children's language input, and that includes rhyme, rhythm, and phonological awareness [2,4,20–23]. Patscheke et al. [22] investigated the impact that music and phonological training had on phonological awareness in four- to six-year-old children of immigrant families. Thirty-nine preschoolers were randomly assigned to three groups and received a twenty-minute intervention thrice a week. One group received music training along with phonological training. One group received phonological training only. One group served as the control and received sports training. Children in the music group and phonological group significantly increased their scores on the phonological awareness test. However, the effect size of the music group was much larger compared to the phonological group. This finding supports the hypothesis that combining music with phonological awareness activities increases language outcomes. Further expanding on evidence, Dunst and Gorman's meta-analysis [24] also found that all children, regardless of differences in disability, age and gender, had positive outcomes in their language and literacy development after exposure to rhyme and rhythm.

The widespread use of music in every culture makes it common for hearing children to recognize rhymes and enjoy reciting songs without specialized interventions. Hearing children with language disabilities, however, may struggle with phonological awareness tasks and need more explicit instruction. When hearing children show difficulty with imitating or reciting common nursery rhymes, professionals raise red flags for dyslexia or language impairment [25]. Multiple interventions that utilize rhyme and rhythm exist for the purpose of helping develop phonological awareness in hearing children [4]. Similarly, activities to increase engagement and recitation skills are typically a part of interventions in early childhood education.

1.2. Engagement and Recitation

The construct of engagement is defined as having cognitive, affective, and social components [26]. First, there is the cognitive task of having heightened alertness and focused attention. Second, there is the affective behavior of having positive attitudes towards the language and the willingness to use it. Third, there is the social aspect of initiating interactions using language. Rhyme and rhythm are known to increase attention, motivation, imitation, and interaction in hearing children, especially for those with disabilities [27]. A meta-analysis found that there were large effects on attention and motivation from music therapy for hearing individuals with developmental or behavioral disabilities [28].

A single-subject mixed method study conducted by Vaiouli et al. [27] looked into the effectiveness of a music intervention in engaging three hearing kindergarteners with autism. Each music therapy session included a welcome song, a child-led part, an adult-led part, and a goodbye song. Actions that counted as engagement behaviors were the child's instances of focusing on the adult's face, shifting gazes, showing awareness and positive affect with smiles or nods, exchanging looks between the object and the adult, and pointing at or showing objects. Significant gains in engagement behaviors were seen in all three children. Similarly, Perry [29] conducted a qualitative study observing 10 hearing students with multiple disabilities during music therapy sessions. Music activities led to increased attention, turn taking, and expressive language. Some children exhibited behaviors of interest and attention during musical activities that they rarely demonstrate in routine activities. Not only do rhyme and rhythm promote engagement in hearing children, but they also support recitation skills.

Recitation is defined as repeating language aloud from memory. Research shows that being exposed to rhyme and rhythm enriches the ability to recite [30–33]. Calvert and Billingsley [30] looked at whether hearing children were able to recite songs without understanding the meaning of words. They showed an incomprehensible song in French and a comprehensible song in English on television to 48 English-speaking preschoolers and asked them to recite the song verbatim. Children were able to successfully recite the incomprehensible French version of the song in the same way as they recited the comprehensible English song. Children's ability to process and recite words did not depend on their knowledge and understanding of these words, indicating the effects of rhyme and rhythm on recitation skills.

Sheingold and Foundas [34] examined the impact of rhyming and non-rhyming versions of stories on 24 five- and six-year-old hearing children's ability to accurately recall the details and provide the correct picture sequence of stories. More specifically, the researchers wanted to know if the presence of rhyme would impact memory. Both versions (rhyming and non-rhyming) of each story had the same information but with rhyming words removed in the non-rhyming version, and they were counterbalanced in their administration. After the story was told, the child was asked five questions about the content of the story and arranged the picture cards in the correct sequential order. The researchers found that hearing children did better with the rhyming version. More children also chose the rhyming version as their favorite over the non-rhyming version, supporting social validity.

Johnson and Hayes [33] examined the effects of rhyme on 64 preschoolers' recitation of stories by comparing their performance in reciting rhyming and non-rhyming versions of a short story. Both versions were similar in content but had different order of the lines in the stanza to remove the rhyming aspect in the non-rhyming version. Their dependent variable measures were the numbers of story words correctly recited and the number of story words recited in the correct presentation order. The results of two-factor analysis of variance yielded information that the rhyming version increased verbatim recitation in correct sequential order. However, children also did adequately well with paraphrasing the non-rhyming version. Preschoolers were able to perceive and process the phonological patterns, leading to heightened ability to recite rhyming stories. Yet non-rhyming stories still served their purpose in facilitating comprehension and paraphrasing. This study provides evidence of different kinds of language processing and their distinct benefits when it comes to recitation and memory, which serve language development.

Read et al. [35] conducted a group experimental study to see if rhyming words in shared storybook reading helped hearing children retain more words. They split 24 children aged two to four years old into two groups and had parents read either a rhyming or non-rhyming version of the same animal story to their child individually. The results showed that children were able to retain more words in the rhyming condition, supporting the hypothesis that exposure to rhyme boosts word retention and vocabulary development. These findings illustrate the ways rhyme, rhythm, phonological awareness, engagement, and recitation benefit hearing children.

1.3. Deaf Learners

Unlike the literature on the role of spoken rhyme, rhythm, and phonological awareness for hearing children, little is known about the benefits of ASL rhyme, rhythm, and phonological awareness for deaf children. While research shows that ASL does have its own ways of generating rhyme and rhythm through phonological play, syllables, and movements, there is still much to learn about their relevance to language acquisition and emergent literacy. A review of the literature provides a glimpse into the potential of ASL rhyme, rhythm, and phonological awareness sharing similar functions for deaf children as spoken rhyme and rhythm does for hearing children.

Valli [36], a deaf person known for his ASL poetry renditions, conducted his dissertation examining the role of eye gaze, body shift, head shift, handshapes, and movement in creating rhyme and rhythm in ASL poetry. He explored strategies used by teachers in teaching ASL poetry to deaf students. The teachers incorporated phonological patterns in instruction that stressed repeated patterns in handshape, location, and movement. Valli saw that this instruction helped deaf children understand, memorize, and create ASL poetry through visual rhymes. Valli's work brought new knowledge to the field regarding the application of ASL linguistics to ASL poetry within the context of deaf education. Investigations into signed language rhyme and rhythm were also conducted overseas in France.

Blondel and Miller studied the uses of nursery rhymes in French Sign Language (LSF). They said, " … nursery rhymes exist in sign languages. They are part of language games, along with tongue-twisters, lullabies, riddles, and so on. As far as we know, they are created by deaf adults for children" [37] (p. 29). Phonological parameters required for creating rhymes were identified: handshape, location, movement, and non-manual markers. In LSF, just like ASL, rhythm can be formed by maintaining the flow and manipulating the transitions of signs to make the initial parameter (i.e., handshape, location, movement) of the sign match the previous or subsequent signs. Furthermore, syllables are found in the movements and holds in signed words. The researchers marked the benefits of developing original nursery rhymes in sign language rather than translating spoken nursery rhymes, which had linguistic and phonological limitations.

In Canada, original ASL rhyme and rhythm were developed for a parent-infant program called the ASL Parent-Child Mother Goose Program [38]. In interviews, hearing parents shared excitement in witnessing their deaf toddlers responding positively to ASL rhyme and rhythm by babbling along with their arms and hands, producing happy facial expressions, and laughing. These deaf toddlers' behaviors did not depart from what was expected from hearing toddlers when exposed to spoken rhyme and rhythm. Because deaf toddlers demonstrate favorable responses while viewing ASL rhyme and rhythm, some early childhood educators are beginning to see the benefits of incorporating these practices into their instruction.

Crume observed and interviewed teachers of the deaf in a school that was converted into an ASL/English bilingual program in the 1990s. It was discovered that rhyming and rhythmic activities were recently introduced to their early childhood program. Crume defined the sign rhythm activity as "ASL stories incorporating signs with repetitive movements" [13] (p. 105). A teacher said that they were experimenting with ASL rhyme and rhythm with preschoolers by repeating signs with the same handshape in rhythmic movements. Crume expanded on the definition of this practice in his observation, "In sign rhythm activities, teachers incorporate the repetitive use of signs together with clapping or patting on knees. The sign rhythm activities allowed deaf students to learn specific handshapes in signs in a pattern that made learning fun. This provided the deaf students a similar benefit that hearing preschool children enjoy when they incorporate movement and gesture in songs" [13] (p. 99). Teachers remarked on the effectiveness of rhyme and rhythm activities in increasing engagement and motivation in students with limited language.

Researchers identified a predictable progression in phonological skills over the course of natural acquisition of sign language in young children [39–42]. Young children are able to produce signed words in correct locations before they learn to use correct handshapes and movements. Di Perri's experimental study [43] on phonological awareness in 29 deaf children between four and eight years old further reinforces these findings. Deaf children were asked to engage in the tasks of identification, categorization, differentiation, blending, segmentation, and substitution for each phoneme—handshape, location, and movement. Location was found to be the easiest phoneme for deaf children to manipulate, followed by handshape, and then movement. Categorization was the easiest phonological awareness task, followed by identification, and then differentiation. More tasks such as blending, segmenting, and substituting parameters in signed words were also examined. It was discovered that segmenting a signed word into parameters was the easiest, followed by blending, and then substituting. No significant difference could be found in performance across tasks and ages. The hierarchy of ASL phonological awareness parallels English phonological awareness, with the task of identifying being the easiest and substituting being the hardest [18], indicating similarities in language development processes.

Andrews and Baker [44] argued that ASL rhyme and rhythm should be used in early childhood programs to support healthy language acquisition. They laid down a framework in which ASL rhyme and rhythm can promote ASL phonological awareness, ASL vocabulary development, and emergent literacy. Among the prominent advantages of ASL rhyme and rhythm are early communication, executive functions, language, early literacy, and metalinguistic awareness. The authors asserted that ASL rhyme and rhythm are a valuable tool in making language experiences positive and fun for young deaf children and their families alike. Although it may be intuitive to say that ASL rhyme, rhythm, and phonological awareness benefit deaf children, findings from cognitive neuroscience further add to the strength of the argument.

Petitto [15,45–47] has carried out some groundbreaking work in neuroscience research on identifying the functions of ASL in the developing brain. Her twenty years of research have extensively expanded our knowledge of literacy development in relation to brain mechanisms and phonological awareness in deaf children. Whether the modality is spoken or signed, the brain responds in the same way in its search for phonological patterning. Petitto and her colleagues said, "the crucial link for early reading success is not between print and sound, but between print and the abstract level of language organization that we call phonology - signed or spoken ... " [15] (p. 367). Similar conclusions were made by McQuarrie and Abbott, who have done substantial work in exploring the role of ASL phonological awareness in literacy development in deaf children. "Having a strong phonological foundation in any language may be more important than the modality itself" [48] (p. 96). In terms of the relationship between phonological awareness in English and literacy skills, a meta-analysis conducted by Mayberry et al. [49] found that English phonological coding and phonological awareness skills accounted only for 11% of the variance in reading proficiency in deaf participants. Further affirming this phenomenon, a recent experimental study tested deaf children's English vocabulary scores after viewing ASL-English bilingual stories that utilized ASL rhymes, ASL-English bilingual stories without any rhymes, and ASL-English stories that had English rhymes. Results showed that deaf children demonstrated the greatest gains in English vocabulary scores with the ASL rhyming story over the other two conditions [50]. These findings suggest that ASL rhyme, rhythm, and phonological awareness may have a bigger role in language and literacy development than previously realized.

While there is no clear explanation yet on exactly how deaf children transfer their ASL knowledge to the acquisition of English, what can be derived from the current literature is that deaf children can become successful readers through myriad avenues either with or without spoken phonological awareness [8] and that ASL phonological awareness may have a positive influence on both ASL and English literacy [48,50,51]. How the use of two languages—ASL and English—are processed and mediated in the brain during the acquisition and development phase in bilingual deaf children remain understudied. Similarly, the impact of ASL rhyme, rhythm and phonological awareness on deaf children's language has not been thoroughly investigated.

1.4. Sign Language Impairments and Language Deprivation

Since there are widely recorded challenges in phonological processing tasks in hearing children with language impairments, it can be assumed that similar challenges would also show up in the deaf population with ASL. However, only a few studies have been conducted on native signing deaf students suspected of having dyslexia or sign language impairment. In these studies, deaf students performed poorly on short-term sequential memory tests such as fingerspelling words, recalling sequences of items, and repeating ASL sentences [52]. Yet there is a paucity of knowledge regarding the role of ASL rhyme, rhythm, and phonological awareness in capturing dyslexia or sign language impairments early on. What complicates the inquiry is the prevalence of deaf children going through their early years without language access, which confounds the line between signed language impairment and language deprivation syndrome [53,54].

A language deprived person is defined as an individual who went through their first years without language access and has had structural changes in the brain as a consequence [55]. Multiple studies have looked into the ability to imitate and recite in deaf adults who experienced language deprivation. These deaf adults had difficulty signing along simultaneously to what was signed to them and struggled to recall ASL sentences verbatim [56–58]. Knowing that the challenges of phonological processing tasks remain with deaf individuals into adulthood, questions are raised about the type of specialized interventions that should be given to young deaf children already experiencing language deprivation. Whether interventions that incorporate ASL rhyme, rhythm, and phonological awareness can help counteract the effects of language deprivation are yet to be investigated.

The literature review of rhyme, rhythm, phonological awareness, engagement, and recitation prompt a wide range of research questions surrounding the role of ASL rhyme and rhythm in young deaf children's development. To date, there is little qualitative or quantitative research on ASL rhyme and rhythm. Thus, any experimental study conducted to explore the relationship between ASL rhyme and rhythm and other developmental areas in children will be significant and bring new knowledge and discussion to the field. The purpose of this research was to examine the effects of ASL rhyme and rhythm on Deaf children's engagement behavior and accuracy in recitation. The research questions were as follows:

(1) What are the effects of rhyming and non-rhyming conditions of ASL stories on Deaf children's engagement behavior?
(2) What are the effects of rhyming and non-rhyming conditions of ASL stories on Deaf children's accuracy in recitation?
(3) What are the effects of handshape rhyme awareness instruction on Deaf children's engagement behavior and accuracy in recitation?

2. Method

The effects of ASL rhyme and rhythm on deaf children's engagement behavior and accuracy in recitation were examined through single subject design. Individual performance was analyzed using a visual analysis that looks at the level, trend, variability, immediacy of effect, and consistency of data patterns within condition and between phases. Group performance was investigated through the mean and mean difference between both conditions and phases. Variables such as vocabulary knowledge and language ability that may impact overall results were examined. Information derived from social validity questionnaires provided insight into the significance of this intervention.

2.1. Alternating Treatments Design

The alternating treatments design capitalizes on the benefits of single subject research by giving two or more treatments to the same individual and then documenting the effects on target behaviors [59,60]. The quick alternation of two different conditions allow for direct comparison between treatments, minimizing potential confounding factors. This design brings a greater understanding of how deaf children respond to two different stimuli—rhyming ASL stories and non-rhyming ASL stories. A functional relation between independent and dependent variables is established when there is consistent evidence of an effect at a minimum of three different points in time [61]. The students needed to demonstrate higher levels of engagement behavior during viewing and higher levels of accuracy in reciting the rhyming or non-rhyming condition of the ASL stories on at least three days' worth of attempts. A higher standard for assessing a functional relation is at least 4 or 5 data points [62].

2.2. Participants and Setting

Upon obtaining approval from the Institutional Review Board at the University of Tennessee (UTK IRB-18-04313-XP), teacher, child, and family participants were recruited from an early childhood program at an ASL/English bilingual deaf school in the western region of the USA. Teachers interested in participating in the study were asked to sign consent forms. Interested families were asked to complete the packet containing a consent form, a family background questionnaire, and a social validity questionnaire. If a family did not give consent for their child to participate in the study, no data was collected from their child. No child was turned away from being able to participate in the study as long as they maintained regular attendance in school.

Ten deaf children between three and six years old with varying backgrounds in language level, race, gender, sex, disability, hearing status, familial hearing status, home language, and socio-economic status participated in this study. By the end of the study, only five students met What Works Clearinghouse (WWC)'s criteria of having four or more data points in each condition for stronger evidence of functional relations. The other five students missed some sessions due to illness, of pulled out for special services or off-campus appointments, which disqualified them from visual analyses. However, there was a special case with a deaf of deaf child, who did not meet the WWC criteria, but exhibited atypical language patterns. This child is included in the report for the purpose of extending our understanding of a possible case of sign language impairment. Table 1 lists characteristics for each student participant. The teacher participants were two preschool teachers and a prekindergarten teacher. Two teachers were native deaf signers and one teacher was hearing but fluent in ASL. The teachers' teaching experience ranged from five to twelve years.

Table 1. Student Participants' Characteristics.

Name	Class	Age	ASL	Vocab.	VCSL	Sex	Race	Disab.	P.H.S.	H.L.
Daya	P.S.	4.6	3 years old	10/23	2.7	F	White	None	H + H	Eng.
Yair	P.S.	4.10	Birth	16/23	2.4	M	Asian	None	D + D	ASL
Giada	P.K.	5.7	4 months old	21/23	4.5	M	White	None	H + H	ASL & Eng.
Jaslene	P.K.	5.10	Birth	21/23	4.5	M	Mixed	None	D + D	ASL
Lexie	P.K.	6.5	4 years old (adop.)	14/23	2.8	F	Asian	None	D + D	ASL & Sign. Lang.
* Lacey	P.S.	3.10	Birth	19/23	3.7	F	Asian	None	D + D	ASL

Notes: Names are pseudonyms to maintain confidentiality. P.S. = Preschool. P.K. = Prekindergarten. ASL = age of initial acquisition in American Sign Language. Adop. = adopted. Vocab. = scores from picture vocabulary assessment (See Appendix A, Figure A1). VCSL = language age from Visual Communication Sign Language assessment. M = male. F = female. Disab. = disability. P.H.S = Parental hearing status. D = deaf and H = hearing. H.L. = home language. Span. = Spanish. Eng. = English. Sign. Lang. = foreign signed languages. * Lacey did not meet WWC standards but demonstrated atypical language patterns.

Preschool and prekindergarten classes had a daily routine following a schedule of activities. When it was ASL time, students were seated in a semi-circle facing a large Smartboard. To be consistent with the classroom routine, the teacher in each classroom, with the first author's assistance, introduced either the rhyming or non-rhyming condition of ASL stories in a video format on the Smartboard to the whole class, including students not participating in the study. After collecting engagement data from the students via a camera latched to the Smartboard, the first author called them individually to a private space next to the classroom where they were recorded reciting the story. Prior to the study, the first author came to their classroom daily for a week and asked them to view and recite videos from their regular curriculum to get the students acquainted with the researcher, the camera, and the process.

3. Materials

The first author created a total of five ASL videos—four versions of two ASL stories and an ASL story for the preferred condition—for the intervention in this study. The first author used to be an ASL teacher for deaf preschoolers, taught ASL courses in universities, and is a co-founder of Hands Land, a company that develops ASL rhyme and rhythm for young children. These experiences contributed to the qualifications required in developing the materials for this study. Both versions were similar in vocabulary and basic semantic content, but some of the words were ordered differently to eliminate rhyme and rhythm in the non-rhyming version. Both versions had the same rate and inflection, were syntactically correct, and made sense semantically. The videos were reviewed for production similarity by the classroom teachers and a deaf colleague prior to implementation. They confirmed that the only difference in the videos was the ordering of words and the absence of rhythms in the non-rhyming story. (See Tables 2 and 3 for ASL gloss of the rhyming and non-rhyming versions of an ASL story.) Each line in the rhyming version had the same handshape for all the signed words while each line in the non-rhyming version had different handshapes. For handshape rhyme awareness lessons, the first author developed a slideshow presentation with images of rhyming signed words and images of individual handshapes.

Table 2. Rhyming Version of Animals Crossing.

Rhyming Version
(1) SPOT—ONE—MOUSE—CROSSING (1-handshape rhyme)
(2) SEE—TWO—RACCOONS—CROSSING (2-handshape rhyme)
(3) JAW DROP—THREE—ROOSTERS—CROSSING (3-handshape rhyme)
(4) HAIR STAND—FOUR—ZEBRAS—CROSSING (4-handshape rhyme)
(5) SHOCK—FIVE—DEER—CROSSING (5-handshape rhyme)
(6) WALK—FINISH! (5-handshape rhyme)

Table 3. Non-Rhyming Version of Animals Crossing.

Non-Rhyming Version
(1) SHOCK—ONE—ROOSTER—CROSSING (no handshape rhyme)
(2) JAW DROP—TWO—ZEBRA—CROSSING (no handshape rhyme)
(3) SPOT—THREE—DEER—CROSSING (no handshape rhyme)
(4) SEE—FOUR—MICE—CROSSING (no handshape rhyme)
(5) HAIR STAND—FIVE—RACCOONS—CROSSING (no handshape rhyme)
(6) WALK—FINISH!

4. Measures

4.1. Baseline Assessments

The researcher-made picture vocabulary assessment consisted of printed images of the selected 22 out of 45 vocabulary words in ASL Story 1 and ASL Story 2 (See Appendix A, Figure A1). The vocabulary words were mouse, raccoon, rooster, zebra, deer, one, two, three, four, five, red, orange, yellow, green, blue, purple, worm, bison, whale, bird, shark, and skunk. While vocabulary knowledge was not the main focus of this study, knowing whether students already had previously developed a lexicon of the target words in the ASL stories could help explain their performance during the intervention. For the purpose of analysis, three subgroups were formed regarding students' knowledge of the target words in the ASL stories: little (0–8 signed words), some (9–18 signed words), and most (19–23 signed words). The Visual Communication Sign Language (VCSL) scores were collected from classroom teachers. The VCSL checklist tracks young children's sign language development and provides information on whether they are meeting age-appropriate milestones [63]. Two subgroups based on language abilities indicated in the VCSL checklist were formed for this study: typical and delayed. Students behind in language by two years or more were considered delayed and placed in the delayed subgroup.

4.2. Social Vailidity Questionnaire

The researcher-made social validity questionnaire was a 21-item Likert scale ranging from strongly disagree to strongly agree. There were six categories in the questionnaire: knowledge, experience and uses, implementation, language development, preference and skills, and recommendations. A few examples of the items were: "I was familiar with ASL rhyme and rhythm prior to this research", "I have access to ASL rhyme and rhythm videos", "Signing along with ASL rhyme and rhythm videos is easy for me", and "ASL rhyme and rhythm videos are good resources for families". The questionnaire in English or Spanish was sent to families through regular school-to-home communication.

4.3. Independent and Dependent Variables

ASL stories with rhyming and non-rhyming versions were the two treatment conditions used to measure students' engagement behavior and accuracy in recitation. The four dependent variables were: nonverbal engagement (viewing), verbal engagement (imitating), words recited correctly, and words recited in the correct order. The viewing behavior in nonverbal engagement was defined as eyes on the screen or eyes on peers (if their peers were signing in imitation of the source material). The imitating behavior in verbal engagement was defined as signing along with the signer in the video or peers using signed words associated with the ASL story. Disengagement was defined as eyes off the screen, eyes off the signer, or signing words not associated with the ASL story. Disengagement by interruption was defined as teacher interruption, student interruption, or other external distractions interfering with the student's ability to attend and/or engage with the independent variable. Words recited correctly were defined as repeating and signing aloud any words from the ASL story from memory, regardless of the sequence of words. Words recited in the correct order were defined as repeating and physically signing the words of the ASL story from memory in the correct sequence.

4.4. Data-Recording Procedures

A permanent product in the form of videotaping was used to collect data on engagement behavior and recitation data. The videos collected for engagement behavior were immediately reviewed after each intervention session and a 5-second partial interval data recording procedure was used to indicate if the student was engaged or disengaged. The final metric was calculated by dividing the total number of 5-second interval engagement behaviors by the total intervals measured during a viewing session (n = 26–32).

Students' recitation, which took place individually in a private room, was video recorded and measured using event recording procedures. The first part of the analysis awarded a point for each word recited correctly from the ASL story, regardless of the sequence of the signed words. The second part of the analysis gave a point for each word recited in the correct order. The number of words signed correctly and the number of words signed in the correct order in the rhyming condition and the non-rhyming condition were analyzed and compared to determine the preferred condition in increasing accuracy in recitation.

4.5. Procedural Integrity

Teachers involved in this study received one hour of consultation on administering the whole class intervention with integrity. They were asked to stick with their routine of calling students to sit down in a U-circle, asking them to be ready to view a story on screen, clicking 'play,' and then staying behind students. The first author was present at all sessions and provided immediate feedback when teachers did not achieve fidelity. In 42 sessions, 155 teacher behaviors out of 168 of the total amount of planned teacher behaviors were successfully executed for a total of 92% procedural integrity.

The second part of the intervention did not include classroom teachers, as students were individually called by the first author to the conference room to re-watch the ASL story on a laptop and recite it to camera. All of the 194 recitation sessions met the procedural integrity with 100% fidelity.

4.6. Inter-Rater Reliability

The reliability of the student data was established through the inter-rater agreement of 90% accuracy or above by the first author and another deaf colleague with a doctoral degree who is also fluent in ASL. If there were disagreements, both observers viewed the video again and discussed their observations until an agreement was reached. Thirty percent of the task engagement data and 35% of the recitation data were double scored, and interrater reliability was 95% and 96% respectively.

4.7. Visual Analysis

Insights from meta-analyses along with WWC's criteria for procedures and standards on alternating treatments design informed the decision to use visual analysis and the total mean and mean difference in this study [62,64–67]. Visual analysis of single subject data addresses whether behavior changed and if that change came from the intervention. Six indicators were used to evaluate within-phase and between-phase data patterns to judge the extent of the effects of the intervention: (a) level, (b) trend, (c) variability, (d) immediacy of the effect, (e) overlap, and (f) consistency of data patterns across similar phases [68]. Examination of within- and between data patterns using these six indicators informed the existence of causal relation and the strength of its evidence [69].

4.8. Total Mean and Mean Difference

The total mean and mean difference procedures were conducted for the whole group to compare the effects of rhyming and non-rhyming conditions. Non-parametric statistical tests, the Wilcoxon Signed-Rank Tests, were performed to compare baseline means across rhyming and non-rhyming conditions. There was no basis for prediction at the baseline, making it exploratory. The treatment means across rhyming and non-rhyming conditions were predicted to be significant. Finally, a comparison of the preference phase and treatment means of the same type was predicted to be non-significant.

5. Procedure

Prior to initiating the baseline sessions, family questionnaires were sent to families by putting the documents in the students' backpacks and then collecting them when they were sent back to school—which was the routine home-to-school communication. The first author administered the researcher-developed picture vocabulary assessment to each student individually in a private room at the educational setting. The first author pointed to the image and then signed: "WHAT–THIS?" If the student did not provide a signed word, a pause of five seconds was given before proceeding to the next picture. If the student signed, "DO NOT KNOW," the first author proceeded to the next image. If the student responded with an incorrect signed word, the first author would again point to the image and sign "WHAT–THIS?". This provided the student another opportunity to look at the picture and correct their mistake. If they did not provide the correct signed word on their second attempt, then the first author proceeded to the next image. Information on students' present language levels (VCSL) was collected from the classroom teachers.

There were four phases of interventions in this study: (1) baseline, (2) handshape rhyme awareness intervention, (3) alternating treatments, and (4) preference. After two weeks of baseline alternating treatments sessions in both conditions, there was no observable bifurcation [70] in students' engagement behavior and accuracy in recitation. Thus, baseline data confirmed the concern of students lacking handshape rhyme awareness, and this evidence led to the decision to implement the handshape rhyme awareness intervention. The added condition was two 20-minute lessons on handshape rhyme awareness given by the first author. The lessons paralleled what was found in lessons that teach rhyme recognition in spoken language. Activities in the lesson involved asking students: "What same handshape was used for all of the signed words in the video?" The first author encouraged students to pay attention and notice the handshape patterns in signed words. Following the handshape rhyme awareness intervention, alternating treatments of the second story took place to determine the effects of the intervention on engagement behavior and accuracy in recitation. After the preferred treatment was identified through visual analysis, the least effective condition was discontinued, and the more effective treatment was replicated on subsequent days using a new story (See Figure 1 for the intervention schedule).

Week	Phase	Monday	Tuesday	Wednesday	Thursday	Friday
1 & 2	Baseline	*Rhyming:* Animals Crossing	*Rhyming:* Animals Crossing	*Non-Rhyming:* Colorful Animals	*Non-Rhyming:* Colorful Animals	*Rhyming:* Animals Crossing
3	Handshape Rhyme Awareness Intervention	*Rhyming:* Animals Crossing	Handshape Rhyme Awareness Intervention 1	Handshape Rhyme Awareness Intervention 2	*Rhyming:* Colorful Animals	
4 & 5	Alternating Treatments	*Non-Rhyming:* Animals Crossing	*Rhyming:* Colorful Animals	*Non-Rhyming:* Animals Crossing	*Rhyming:* Colorful Animals	*Non-Rhyming:* Animals Crossing
6	Preference	*Rhyming or Non-Rhyming:* Fun Day	*Rhyming or Non-Rhyming:* Fun Day	*Rhyming or Non-Rhyming:* Fun Day	*Rhyming or Non-Rhyming:* Fun Day	*Rhyming or Non-Rhyming:* Fun Day

Figure 1. Intervention Schedule.

6. Results

6.1. Engagement

The mean of total percentage of engagement occurrence intervals in the baseline was 38% imitating and 54% viewing in the rhyming condition, and 26% imitating and 61% viewing in the non-rhyming condition. The mean of total percentage of engagement occurrence intervals in the alternating treatments phase was 44% imitating and 32% viewing in the rhyming condition, and 22% imitating and 58% viewing in the non-rhyming condition. The mean of total percentage of engagement occurrence intervals in the preference phase was 45% imitating and 34% viewing in the rhyming condition. There were no statistically significant differences on the Wilcoxon Signed-Rank Test for baseline ($Z = -0.81$, $p < 0.42$), treatment ($Z = -0.67$, $p < 0.5$), or preference phase compared to treatment ($Z = -0.94$, $p < 0.35$). The same was true for imitation for baseline ($Z = -1.75$, $p < 0.08$), treatment ($Z = -1.83$, $p < 0.07$), and preference phase compared to treatment ($Z = -0.41$, $p < 0.69$); however, total means seem to show imitation to be happening in the rhyming condition more than the non-rhyming condition, especially during the treatment phase (M = 43.8 in rhyming condition compared to M = 22.4) (See Figure 2).

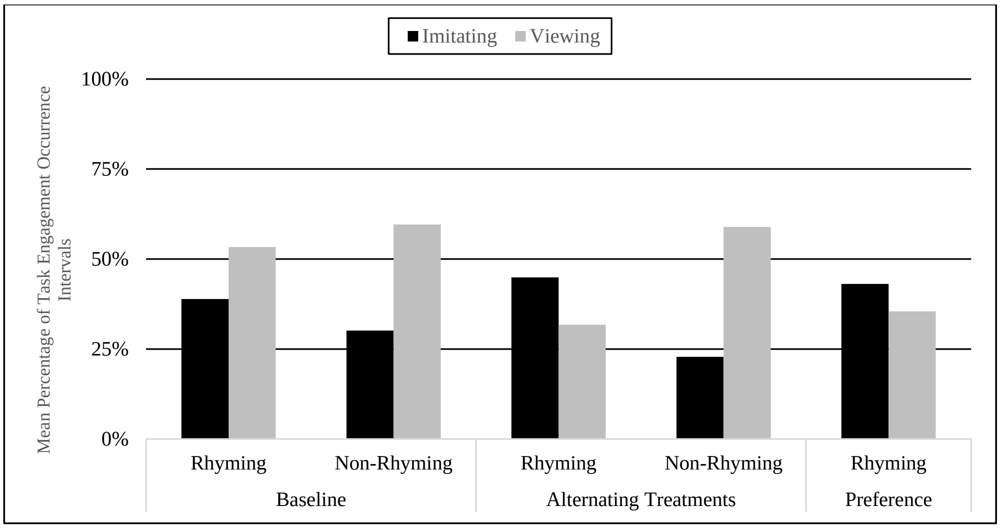


Figure 2. Engagement: Mean Percentage of Task Engagement Occurrence Intervals in Viewing and Imitating Behaviors.

Students' picture vocabulary assessment scores and their language skills were examined as variables that may influence engagement behavior. There were two subgroups of students' knowledge of the target words in the ASL stories: some (9–18 signed words) and most (19–23 signed words). The two subgroups of delayed and typical language skills were determined based on the VCSL checklist. The two students that knew the most vocabulary were the same two students that had age-appropriate language skills. Similarly, the three students that knew some vocabulary had language delays. The most vocabulary knowledge and typical language subgroup demonstrated higher total means in imitation in both conditions (74% in the rhyming condition, 57% in the non-rhyming condition) than the some vocabulary knowledge and language delayed subgroup (21% in the rhyming condition, 3% in the non-rhyming condition) (See Figure 3).

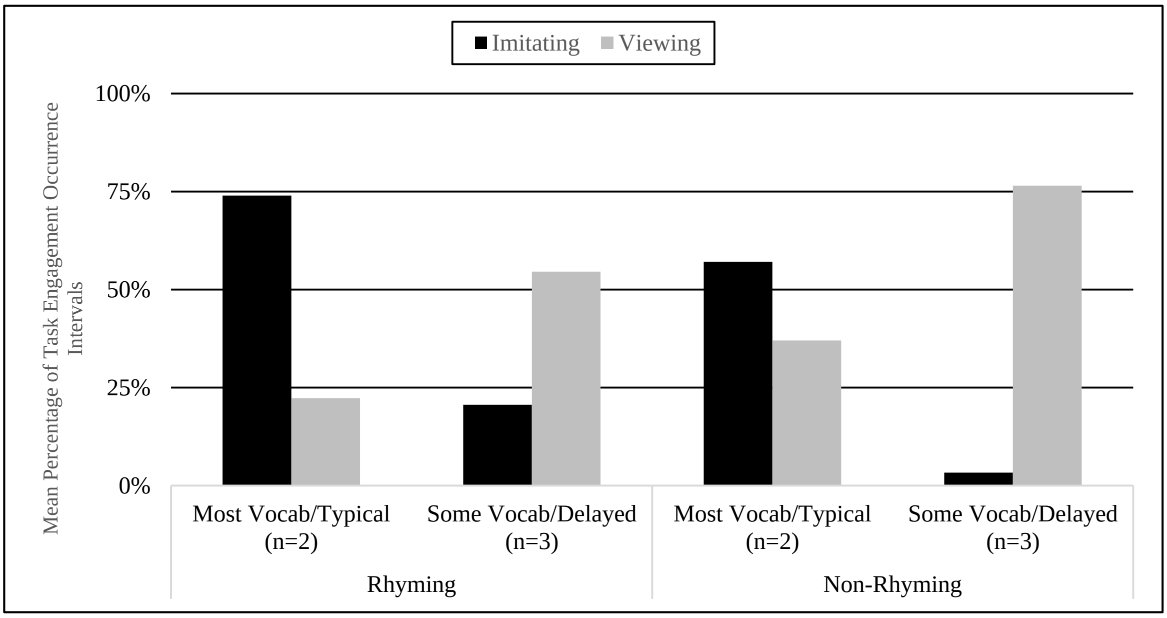


Figure 3. Mean Percentage of Task Engagement Occurrence Intervals in Viewing and Imitating Behaviors Across Phases in Some Vocabulary/Delayed and Most Vocabulary/Typical Groups.

6.2. Recitation

The mean percentages of words signed correctly in rhyming and non-rhyming conditions in each phase and across phases are presented in Figure 4. During the baseline, the mean percentage of words signed correctly in the rhyming condition was 52% and in the non-rhyming condition it was 41%. After handshape rhyme awareness intervention was given, the mean percentage of words signed correctly in the rhyming condition was 69% and in the non-rhyming condition it was 36%. In the preference phase, the mean percentage of words signed correctly in the rhyming condition was 67%. Wilcoxon Signed-Rank Tests indicated that the baseline and treatment means for words signed correctly in the rhyming conditions were statistically significantly higher than the non-rhyming conditions, $Z = -2.03$, $p < 0.04$; $Z = -2.02$, $p < 0.04$, although the total mean in the treatment ($M = 68.8$ in rhyming condition compared to $M = 36.0$) shows more substantial difference than at baseline ($M = 52$ in rhyming condition and $M = 40.8$ in non-rhyming condition). Furthermore, as predicted, the preference condition ranks were not statistically significantly different compared to the treatment condition ranks, $Z = -0.41$, $p < 0.67$.

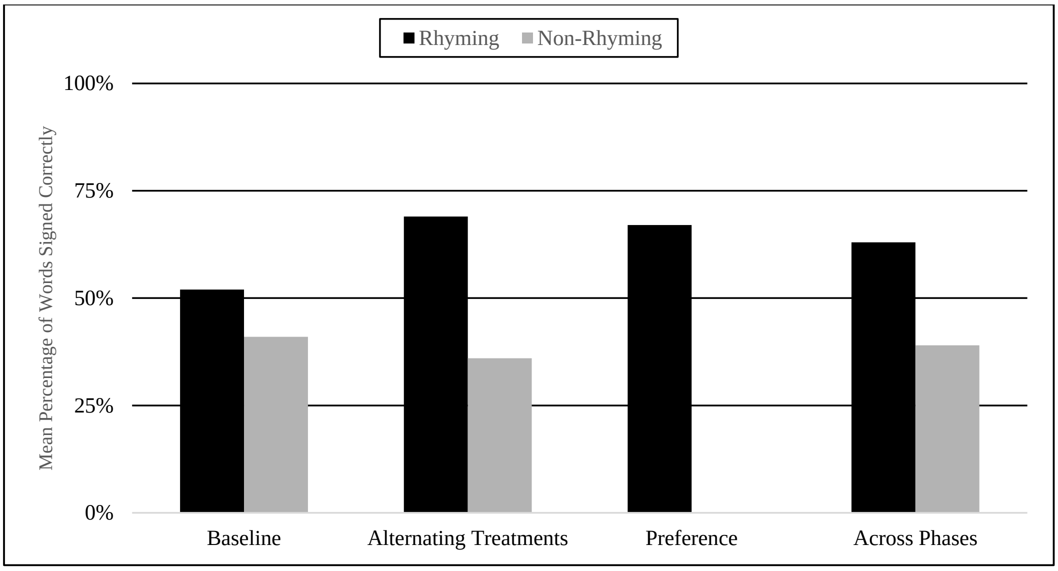


Figure 4. Mean Percentage of Words Signed Correctly in Rhyming and Non-Rhyming Conditions in Each Phase and Across Phases.

The mean percentages of words signed in the correct order in rhyming and non-rhyming conditions in each phase and across phases are demonstrated in Figure 5. In the baseline, the mean percentage of words signed in the correct order in the rhyming condition was 38% and in the non-rhyming condition it was 32%. After the handshape rhyme awareness intervention, the mean percentage of words signed in the correct order in the rhyming condition was 64% and in the non-rhyming condition it was 28% during the alternating treatments phase. In the preference phase, the mean percentage of words signed in the correct order in the rhyming condition was 62%. Wilcoxon Signed-Rank Tests performed for words signed in correct order show the means for rhyming condition in the treatment phase are statistically significantly higher than those in the non-rhyming condition, $Z = -2.02$, $p < 0.04$, while no significant differences were detected at baseline, $Z = -4.05$, $p < 0.69$, or between the preference and treatment conditions, $Z = -0.37$, $p < 0.72$.

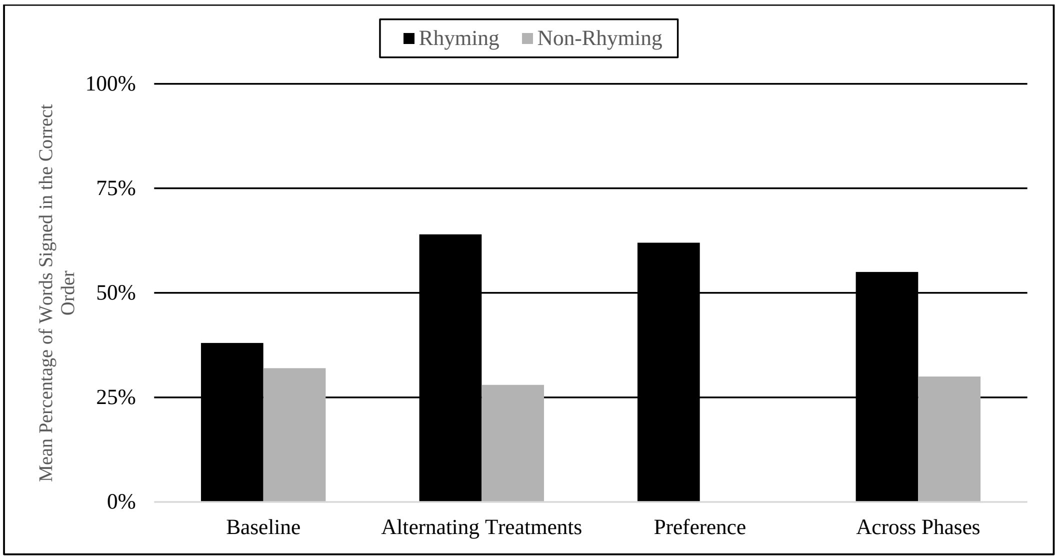


Figure 5. Mean Percentage of Words Signed in the Correct Order in Rhyming and Non-Rhyming Conditions in Each Phase and Across Phases.

In the most vocabulary and typical language subgroup, the mean percentage of words signed in the correct order was 80% in the rhyming condition and 39% in the non-rhyming condition. In the some vocabulary and delayed language subgroup, the mean percentage of words signed in the correct order was 25% in the rhyming condition and 18% in the non-rhyming condition. Total means seem to show more words signed in the correct order in the rhyming condition than the non-rhyming condition in both subgroups (See Figure 6).

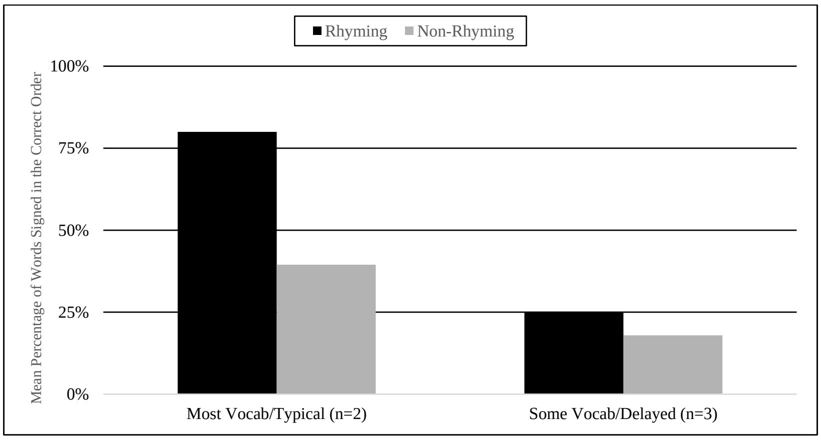


Figure 6. Mean Percentage of Words Signed in the Correct Order Across Phases in Some Vocabulary/Delayed and Most Vocabulary/Typical Language Groups.

6.3. Visual Graphs

Visual graphs showing student performance in the rhyming and non-rhyming conditions with data overlapped for visual comparison are presented for words signed correctly and words signed in the correct order during recitation. The order of students is listed based on age, starting with the youngest, with an exception for Lacey who did not meet the WWC standard of having four or more data points for each condition in each phase. Lacey's visual graph is included at the end for the purpose of exploring a possible case of sign language impairment (See Figures 7–17). Most students performed relatively similarly when it came to accuracy in recitation in both conditions in the baseline, with the rhyming condition being slightly superior. After receiving handshape rhyme awareness intervention, most students' accuracy in recitation increased in the rhyming condition during alternating treatments and preference phases.

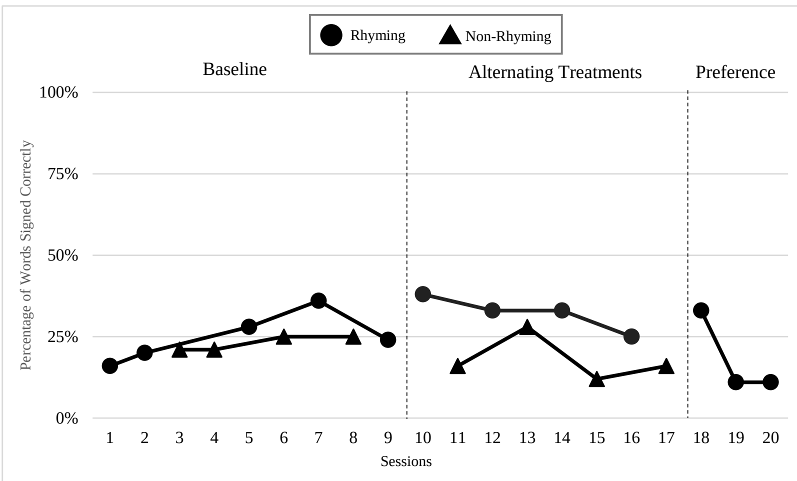


Figure 7. Daya's Percentage of Words Signed Correctly in Recitation.

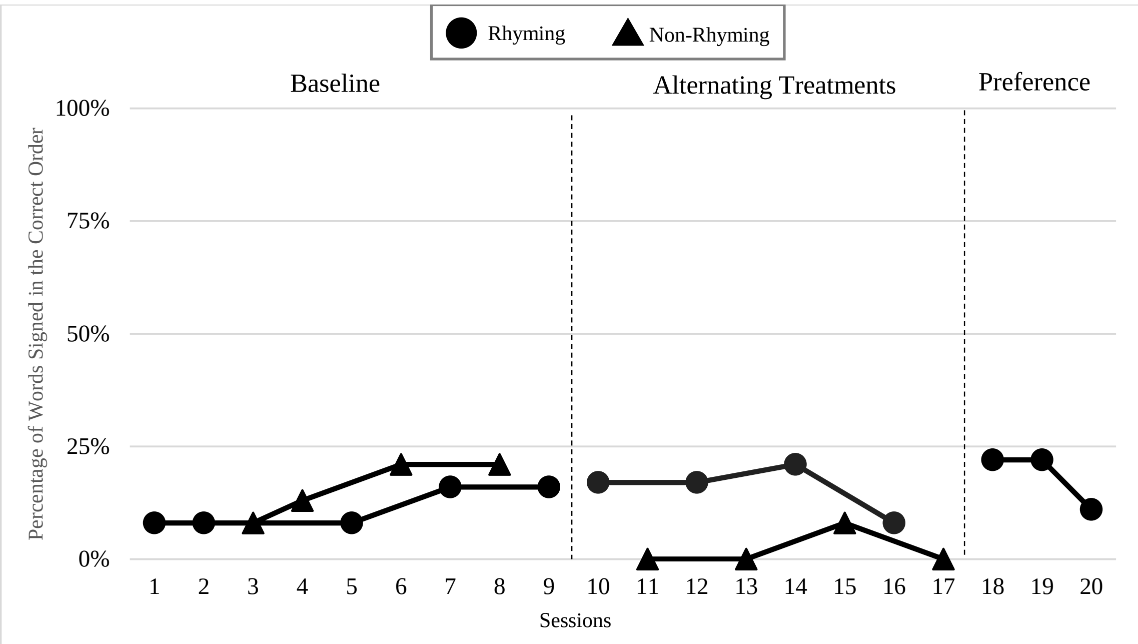


Figure 8. Daya's Percentage of Words Signed in the Correct Order in Recitation.

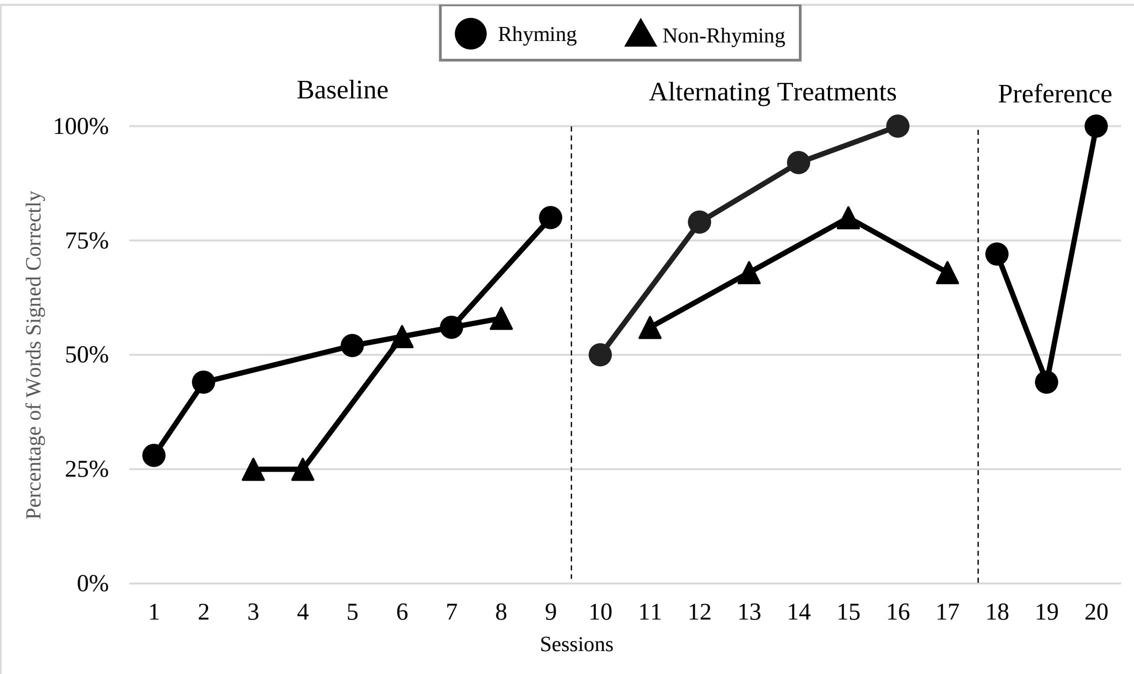


Figure 9. Yair's Percentage of Words Signed Correctly in Recitation.

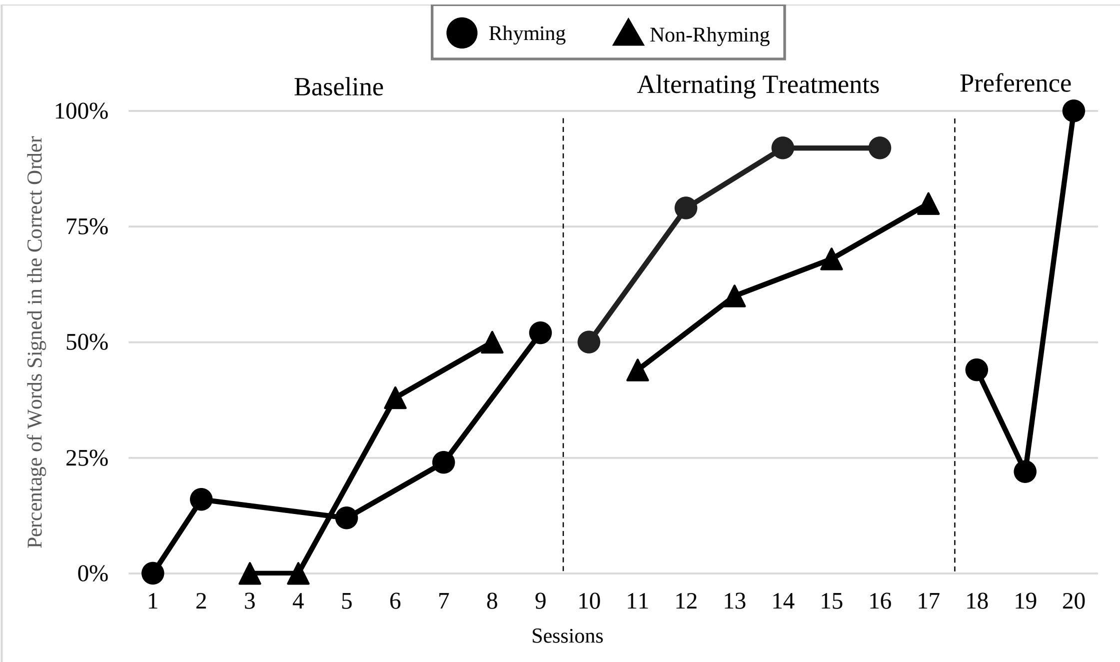


Figure 10. Yair's Percentage of Words Signed in the Correct Order in Recitation.

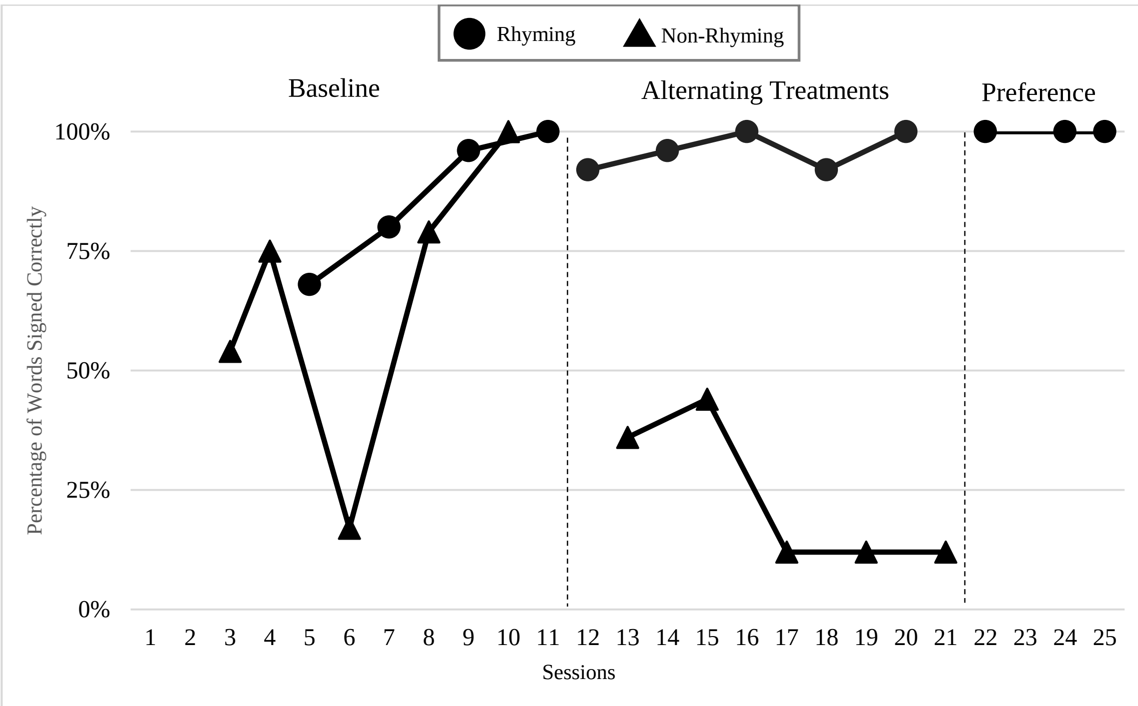


Figure 11. Giada's Percentage of Words Signed Correctly in Recitation.

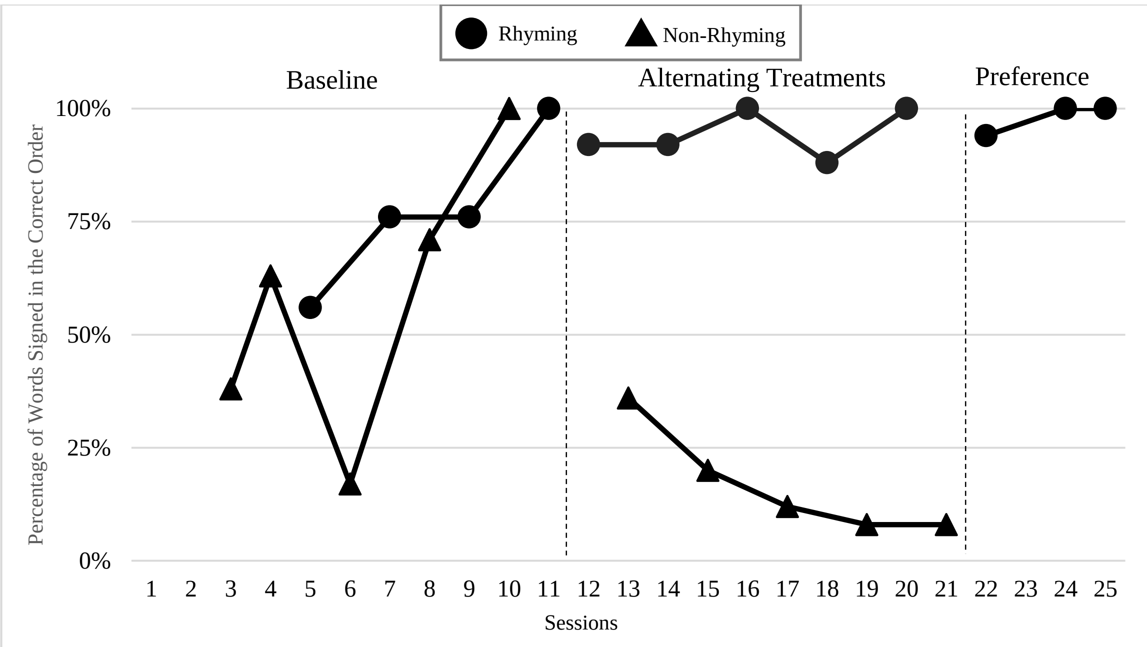


Figure 12. Giada's Percentage of Words Signed in the Correct Order in Recitation.

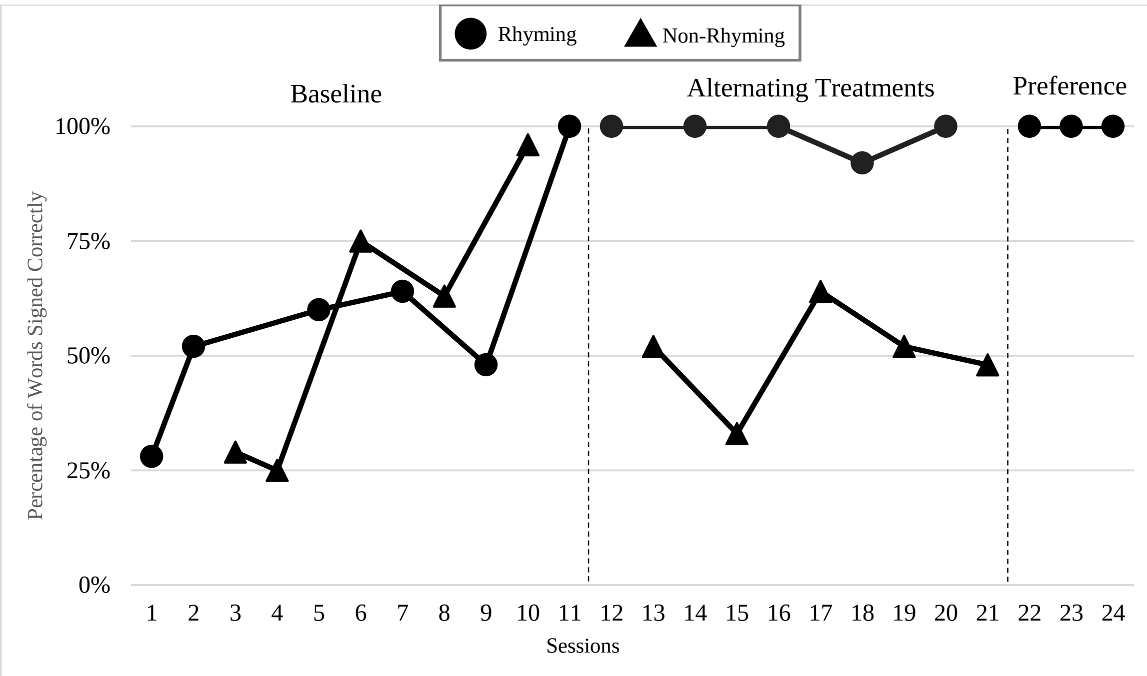


Figure 13. Jaslene's Percentage of Words Signed Correctly in Recitation.

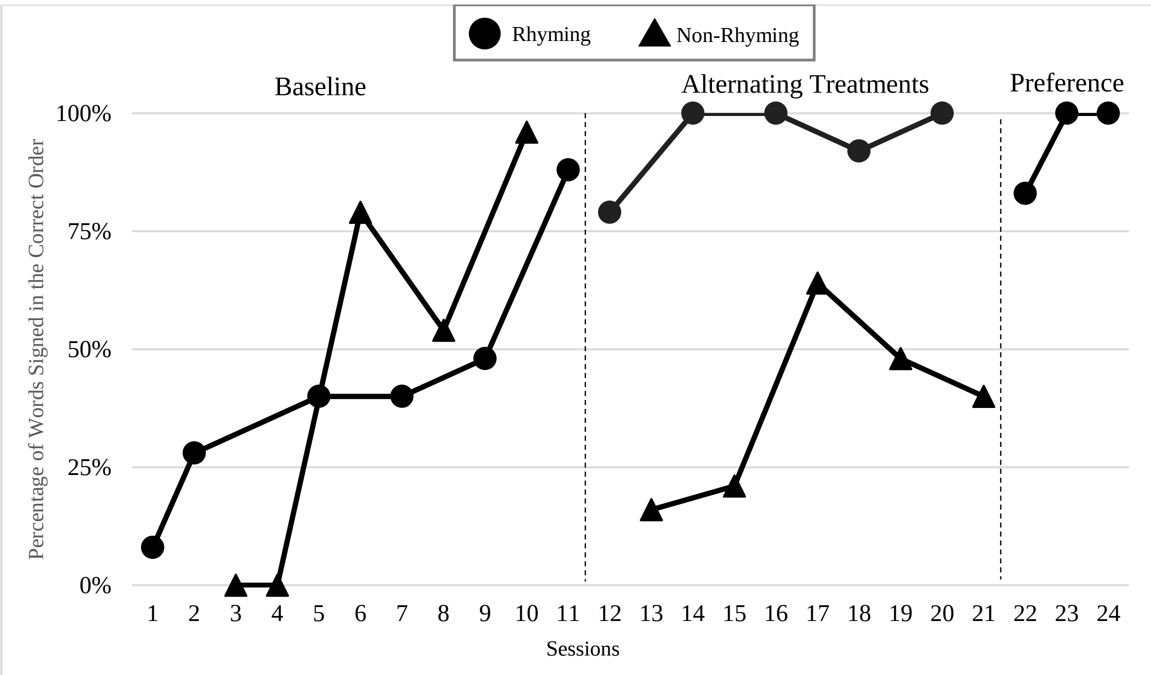


Figure 14. Jaslene's Percentage of Words Signed in the Correct Order in Recitation.

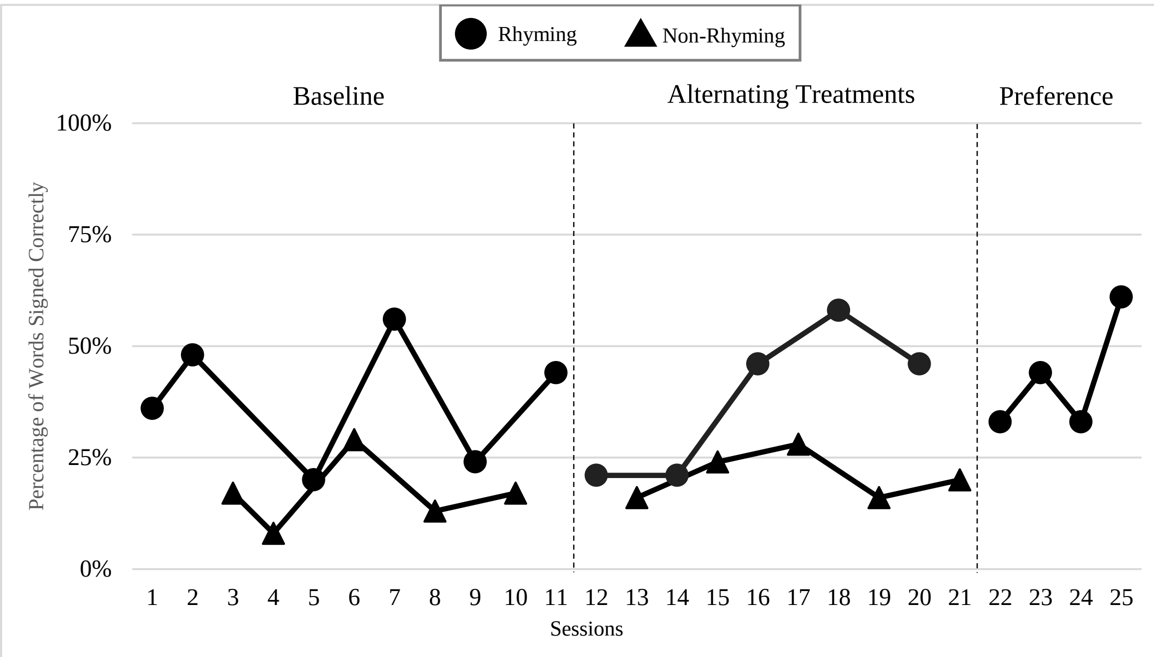


Figure 15. Lexie's Percentage of Words Signed Correctly in Recitation.

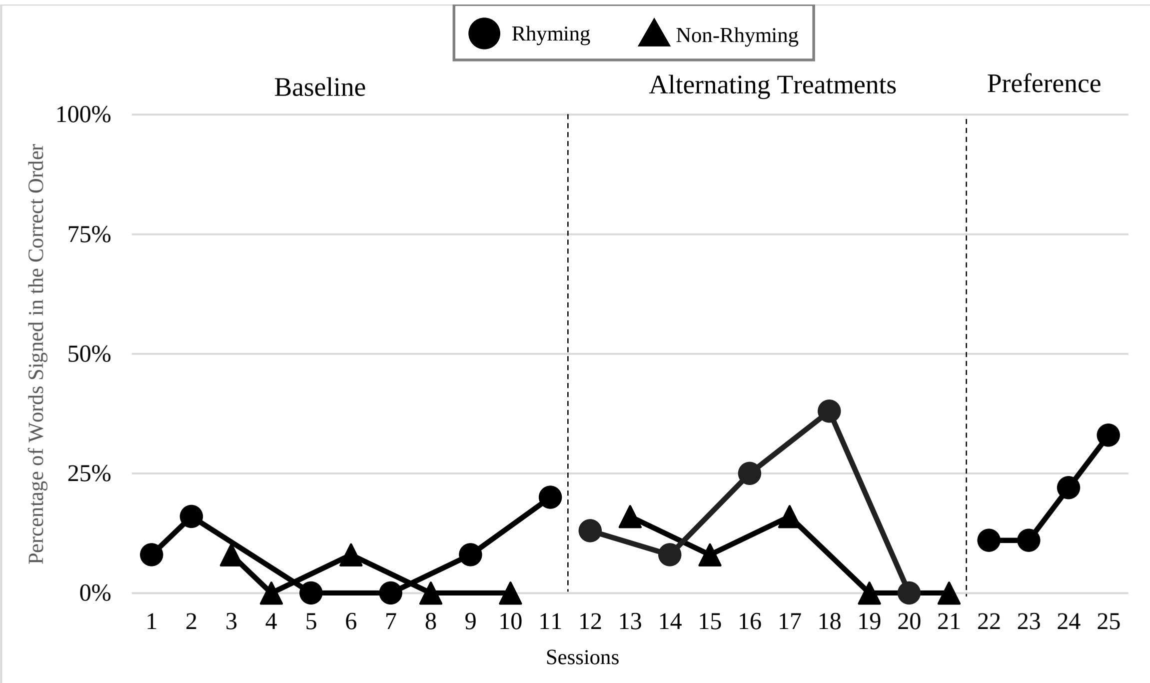


Figure 16. Lexie's Percentage of Words Signed in the Correct Order in Recitation.

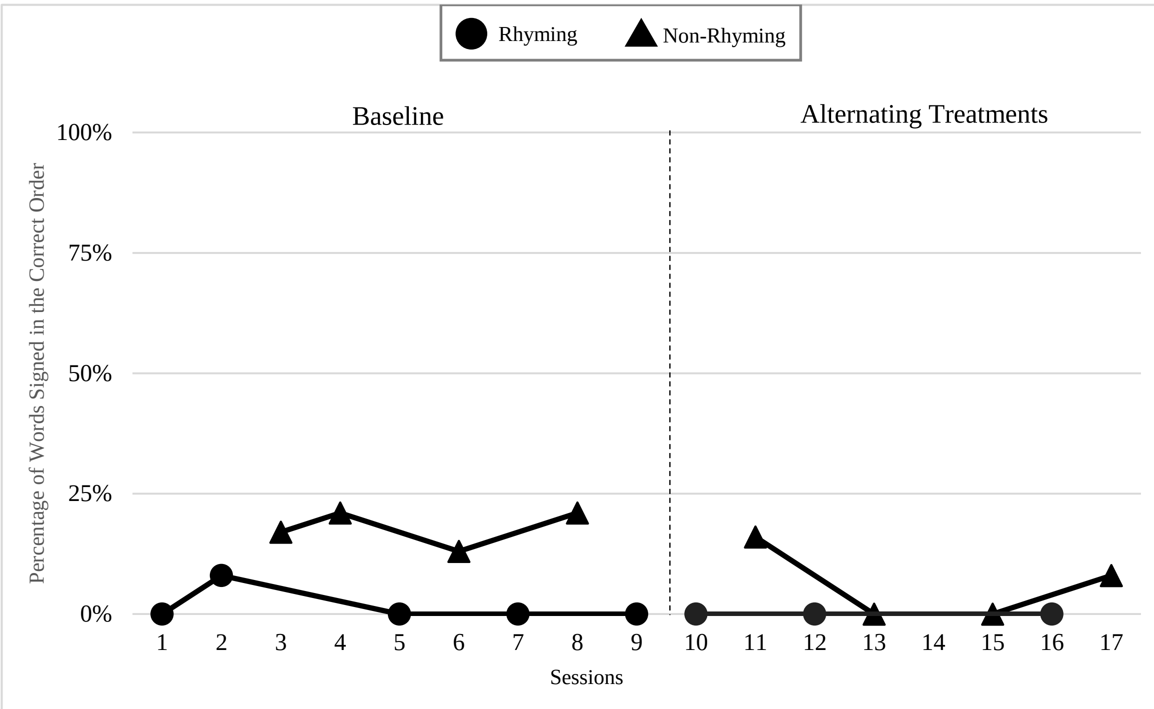


Figure 17. *Lacey's Percentage of Words Signed in the Correct Order in Recitation.

6.3.1. Daya

Daya was four years old and had language delays and some vocabulary knowledge. Daya came from non-signing hearing parents and had been learning ASL for a little over a year. After receiving handshape rhyme awareness intervention, and upon introduction of the alternating treatments phase, Daya's level, trend, and variability between both conditions in the alternating treatments were similar with higher mean percentages in the rhyming condition than the non-rhyming condition (words signed correctly: 14% higher; words signed in the correct order: 14% higher). There was a consistent and small separation of the data paths between both conditions, with the rhyming condition remaining higher over the non-rhyming condition.

6.3.2. Yair

Yair was four years old and came from a deaf family. Yair was delayed in language by two years and had some vocabulary knowledge. Yair's level, trend, and variability between both conditions in the alternating treatments phase were similar with the mean percentage of the rhyming condition being 12% higher than the non-rhyming condition for words signed correctly and 15% higher than the non-rhyming condition for words signed in the correct order. There was a separation in data paths between both conditions of a small to moderate magnitude with the rhyming condition being superior.

6.3.3. Giada and Jaslene

Giada and Jaslene were five-year-old participants who had early access to ASL, age-appropriate language skills, and high vocabulary knowledge. During the baseline phase, both Giada and Jaslene appeared to be oblivious to the existence of handshape rhymes and used the same approaches in their effort to memorize and recite the entire story in rhyming and non-rhyming conditions. For example, they would pause when they could not remember what words came next and demonstrate a "thinking face" as they waited for the words to come to mind. They did not rely on handshape rhymes to clue them as to what come next in the story. They both started the first session of the baseline phase with 56% and 8% words signed in the correct order in the rhyming condition, respectively, and 38% and 0% in the non-rhyming condition, respectively. The accuracy of their recitations in the last session was close to 100% in both conditions. During the handshape rhymes intervention, Giada and Jaslene were shown the rhyming condition of ASL Story 1 and had the handshape rhymes pointed out—both Giada and Jaslene's eyes widened, and their mouths stood agape. Giada put hands to face as if to say: "Why didn't I see it before?!" During the alternating treatments phase, Giada and Jaslene's behavior changed: they immediately pointed out the handshape rhymes in the rhyming condition both during and after viewing and made accurate comments about them. Then, Giada and Jaslene made a big jump in their first session of alternating treatments and recited the rhyming condition of the ASL Story 2 with 92% and 80% of words in the correct order, respectively. Giada and Jaslene's level, trend, and variability in data paths between both conditions in words signed correctly and words signed in the correct order demonstrated a consistent and large magnitude of separation with the rhyming condition having higher mean percentages over the non-rhyming condition (words signed correctly: 73% and 48% higher, respectively; words signed in the correct order: 77% and 56% higher, respectively).

6.3.4. Lexie

Lexie was six years old—the oldest participant in this study—and was adopted from another country two years ago. Lexie came to the United States without any language or lexical vocabulary. After two years of intensive language immersion and support provided by their adoptive deaf parents and teachers of the deaf, Lexie's language skills in ASL grew to the three-year-old level according to the VCSL. Lexie's trend and moderate variability between both conditions in the alternating treatments were comparable with the mean percentage of the rhyming condition, being 17% higher than the non-rhyming condition for words signed correctly and 9% higher than the non-rhyming condition for words signed in the correct order. There was a separation in data paths between both conditions of a small to moderate magnitude with the rhyming condition being higher for words signed correctly.

6.3.5. *Lacey

*Lacey had three data points in the rhyming condition during the alternating treatments phases, which did not meet the WWC criteria for inclusion in this study. However, their recitation data is still included here for the purpose of expanding our understanding of a possible case of sign language impairment. Lacey was almost four years old at the time of the study. Lacey came from deaf parents, had age-appropriate language skills, and high vocabulary knowledge. Lacey did not sign any words in the correct order in four out of five rhyming sessions during the baseline. This trend continued in the alternating treatments phase with 0% of words signed in the correct order for three consecutive sessions. Lacey produced more words in the correct order in the non-rhyming condition with a range of 17–21% of words signed in the correct order in the baseline, and a range of 0–15% words signed in the correct order in the alternating treatments phase (See Figure 17).

6.4. Social Validity Questionnaire

Although only five students met the criteria of having four or more data points in each condition and phase, there were a total of ten students participating in this study. Eight out of ten caregivers, five hearing and three deaf, returned the social validity questionnaire. When asked if they knew how to make rhymes in ASL, four people "agreed," two people "disagreed," one person "strongly disagreed," and one person was "uncertain." When asked if they were familiar with ASL rhyme and rhythm prior to this research, five people "agreed," three people "disagreed" and one person "strongly disagreed." When asked if they had access to ASL rhyme and rhythm videos at home, one person "strongly agreed," four people "agreed," one person "disagreed," and two people were "uncertain." When asked if signing along with ASL rhyme and rhythm videos was easy for them, six people "agreed," and two people were "uncertain." When asked if signing ASL rhyme and rhythm without videos was easy for them, three people "agreed," and five people "disagreed." When asked if they thought ASL rhyme and rhythm were a good way for families to learn sign language, two people "strongly agreed," four people "agreed," and two people were "uncertain." When asked if they thought ASL rhyme and rhythm videos were good resources for families, two people "strongly agreed," and six people "agreed." A caregiver left a comment on the questionnaire, "Anytime spent communicating with your child is particularly important for bonding. Using fun ASL rhymes and rhythms would only enhance this experience." Another caregiver also wrote a note on the questionnaire, "Rhyming in groups I think might be analogous to singing in chorus—a social activity." Caregivers' overall responses were either positive or uncertain as some were familiar with ASL rhyme and rhythm while others did not know ASL nor had any exposure to this practice.

7. Discussion

Research on young hearing children shows that engagement, imitation, rhyme awareness, and recitation play an integral role in language development. This study explored whether ASL rhyme and rhythm had similar results in deaf children by comparing the effects of rhyming and non-rhyming conditions of ASL stories after giving explicit instruction in handshape rhyme awareness. The results of the group means did not indicate a significant impact on engagement. However, significant gains could be found in group means of accuracy in recitation.

7.1. Engagement and Imitation

There was a difference in the number of intervals deaf children spent imitating compared to viewing ASL stories with and without rhyme and rhythm, but this was not significant. What is remarkable about this outcome is that deaf children engaged in imitating behavior in both conditions on their own without any instruction or modeling from adults. This spontaneous and naturalistic behavior while attending to stories, including those with rhyme and rhythm, shows how deaf children are not any different from hearing children in their imitating behaviors [71]. Furthermore, deaf children with language delays in this study imitated substantially less than deaf children with typical language skills. This is in align with research that found hearing children with language delays demonstrated fewer imitating behaviors while listening to songs [25]. Although deaf children with language delays imitated less than deaf children with age-appropriate language skills in this study, it is important to note that they still imitated more while viewing ASL rhyme and rhythm than with ASL stories without rhyme and rhythm. It may be that deaf children's brains are naturally seeking phonological patterns, which affirms the hypothesis of Petitto et al. [15] on the significance of sign phonology in language development. The alluring and fun nature of rhyme and rhythm seem to elicit more imitating behaviors in hearing and deaf children alike.

7.2. Recitation and Phonological Awareness

This study departed from past studies by shedding light on whether deaf children are able to independently recognize rhymes and recite them. Although their performances were slightly better in the rhyming condition at the baseline, deaf children did not appear to recognize the existence of rhymes when exposed to ASL rhyme and rhythm and struggled with reciting words in the correct sequential order. Their lack of handshape awareness was incongruent with the literature on rhyme awareness development in the population of young hearing children [18]. This finding was somewhat unsurprising considering that deaf children have extremely limited exposure to and experience with ASL rhyme and rhythm. The first author's experience providing professional development on ASL rhyme and rhythm reinforces this phenomenon. During the training sessions, many of the professionals at the schools serving the deaf children lacked rhyme awareness themselves. Additionally, classroom teachers in this study confirmed that their students had extremely limited exposure to ASL rhyme, rhythm, and phonological awareness. The lack of handshape rhyme awareness may help explain why they did not initially respond to rhymes in the ASL story.

Deaf children may need to first master the prerequisite skills of handshape identification, handshape categorization, and rhyme knowledge in order to have heightened awareness and appreciation of the features found in ASL rhyme and rhythm. Without these foundational skills, children in this study did not think of using handshape rhymes as a tool to support the sequential memory that is required for the task of recitation. Multiple studies on hearing children have stressed supplementing song recitations with phonological awareness activities for better outcomes in engagement and recitation [20,22,23,72]. Overarching theoretical postulations exist on the importance of having the skills to recognize rhymes for enhanced ability in remembering vocabulary [73], word pairs [74], sequences [34] and stories [33]. What can be learned from this study is that when deaf children have minimal, if any, exposure to ASL rhyme and rhythm, they are not given opportunities to develop rhyme awareness. Without rhyme awareness, they are oblivious to the existence of rhymes in ASL songs or stories. Two 20-minute interventions to teach deaf children to recognize handshape rhymes produced mixed results in this study. Deaf children with age-appropriate language picked up on the new skill quickly, while there seem to be variations in language delayed deaf children's performance. Their overall performance was not unlike the literature on variations in hearing children with disabilities' acquisition of rhyme awareness with interventions that span months, or even years [27,28].

7.3. Student Performance

Giada and Jaslene's commensurable age, language abilities, vocabulary knowledge, and the similarities in their performance during the baseline, alternating treatments, and preference phases make a useful case study that parallels the literature on typically developing hearing children. Their ability to successfully recite ASL stories (whether they had rhymes or not) is an age-appropriate skill [75]. While viewing, they demonstrated more imitating behavior than they did in the baseline. This marked gain in imitation is important, as the function of imitation in young children may be related to processing linguistic input from the environment, which helps with memory and furthers their understanding of language. It is argued that children imitate only the phonological information that they can perceive and understand well enough to repeat [76]. While reciting, Giada and Jaslene clearly relied on handshape rhymes as clues to guide them in remembering what signed words should come next. When Giada and Jaslene gave the wrong signed word, they quickly caught their mistake because the signed word did not rhyme with the previous signed word. This behavior shows Giada and Jaslene specifically thinking about the linguistic feature of handshape rhyme and reflecting upon their own language production, judging the handshape in their signed word as an error to be corrected.

This type of self-correction demonstrates metalinguistic awareness. The literature states that hearing children as young as three years old can possess metalinguistic skills that include self-correcting behaviors in language output, and that children with language impairments often struggle in this area [77]. Since interrelationships exist between metalinguistic awareness, phonological awareness, and language abilities [78], Giada and Jaslene's successes in recitation in the rhyming condition could be attributed to these factors.

Daya, on the other hand, rarely imitated any signed words while viewing ASL stories in both conditions across phases. Accuracy in recitation was also low. This performance could be explained by overlapping factors related to language deprivation impacting phonological processing and language development. Multiple studies have looked into deaf adults who experienced language deprivation during their early years and their ability to imitate and recall. These deaf adults had difficulty signing along simultaneously to what was signed to them and struggled to recall sentences verbatim [56–58]. It appears that the language processing gap possibly stemming from language deprivation already begins to widen in four-year-old students of non-signing parents like Daya. Similar issues could be seen in Lexie who also experienced language deprivation during the early years.

As the oldest participant in this study, Lexie's case of language deprivation was the most extreme, not having had any language access until four years old. The fact that Lexie had complete access to language at home in addition to being placed in an ASL-rich environment for two years made a difference in their ability to engage with language and recite. Lexie having progressed from a child with no language to a child who could perform as well as some four-year-old deaf children within a two-year period is encouraging. This informs us that having complete access to language through ASL at school and at home can build skills to assist with language development such as imitation and recitation.

The subpar performance of Lacey, a native signing deaf student raised some questions regarding the understudied phenomenon of dyslexia, signed language impairment, or ADHD in the population of deaf children from signing families. By documenting deaf children's imitating behaviors and their ability to recite ASL rhyme and rhythm in this study, it was possible to recognize more clearly the language processing gaps in certain students. Because Lacey comes from a deaf family, has had access to ASL since birth, and had high vocabulary knowledge and age-appropriate language skills, the results seemed to be an anomaly for someone of this particular background. When asked to recite the ASL rhyme and rhythm, Lacey was not able to sign most of the words accurately nor in the correct order. In a follow-up interview with teachers, they said they noticed something was amiss with Lacey's language and academic performance. In class lessons where children were expected to recall an ASL story, Lacey had a tendency of not following the correct sequence. Lacey also struggled with remembering the sequence of numbers. The teachers attributed the weakness to the child's "free spirit" personality and relatively slow social-emotional development.

In spite of the fact of having been signing since birth, Lacey struggled the most of all students with recitation, demonstrating challenges with language processing tasks pertaining to working and sequential memory. However, classroom teachers did not raise this as a red flag for sign language impairment or ADHD. They were not concerned because Lacey was able to produce ASL sentences independently and engage in meaningful turn-taking conversations, unlike peers who were much more delayed due to language deprivation. In other words, issues with language processing in this child were overshadowed by classmates' even weaker skills. While this is an understandable reality, not raising a red flag for dyslexia or language impairment meant that specialized attention is not being given to Lacey's unique language needs.

Interventions similar to the ones used in this study could provide opportunities for professionals to attend to whether language processing skills are lacking, especially in deaf children who are native users of ASL. These kinds of intervention are also used to alert to the possibilities of dyslexia, language impairment, and ADHD in hearing children. There are well documented difficulties in phonological processing tasks such as recognizing words that sound the same and reciting rhyme and rhythm [79].

Whether this is paralleled to sign language impairments in deaf children from signing parents requires further investigation.

7.4. Social Validity

The social validity of exposing deaf children to rhyming ASL stories was explored through a questionnaire. While deaf parents were enthusiastic and asked for more resources, most hearing parents had limited knowledge of and were uncertain about this practice. A hearing parent's note on the questionnaire illustrates their uncertainty, "Rhyming in groups I think might be analogous to singing in chorus—a social activity." A deaf parent wrote that ASL rhyme and rhythm are fun to create at home. Deaf children are highly motivated to imitate, and the experience often ends with everyone bursting into laughter. This comment paints a picture of the high social and cultural importance of this practice for deaf families. All parents, hearing and deaf, agreed that ASL rhyme and rhythm videos are good resources for families. Most parents said it would be hard to expose their deaf children to ASL rhyme and rhythm at home if there were no videos available.

Teachers, like the deaf parents, spoke highly of the role of ASL rhyme and rhythm in fostering language development and lamented over the lack of resources. They did not feel knowledgeable and confident enough to sign ASL rhyme and rhythm on their own in their instruction. A teacher wrote in the questionnaire that they could incorporate ASL rhyme and rhythm, but limited resources create a stumbling block. Six months after the study took place, the principal requested the first author to return and give an all-day professional development to the department on implementing ASL rhyme and rhythm in their classrooms. During the professional development, teachers commented on witnessing the ways ASL rhyme and rhythm have promoted repetitions and patterns, memorization, creativity and play, metalinguistic awareness, prediction, humor, family-child bond, and turn-taking skills in their classes. Issues of resources and training being scarce present major barriers to exposing deaf children to ASL rhyme and rhythm at home and in school.

7.5. Limitations and Future Directions

There are limitations with the ASL assessments currently available to the public [54]. In this study, the Visual Communication Sign Language (VCSL) checklist was selected to approximate students' language abilities according to developmental milestones. While this assessment was helpful in determining language skills (i.e., typical and delayed) which was important to understanding how language impacted students' engagement behavior and accuracy in recitation, the assessment does have limitations. The VCSL checklist was normed on a small sample of children, which calls into question the ability to accurately quantify language delay by years. Notwithstanding, it is one of the few tools that track young children's language milestones in ASL. The information it provided for this study was valuable. Future studies would benefit from the inclusion of assessment tools that can more precisely capture deaf children's ASL phonological awareness and language skills.

Information extrapolated from this study indicates that not all deaf children have abundant exposure and experience with language—let alone ASL rhyme, rhythm, and phonological awareness—and this impacts their language processing abilities. There are still many unanswered questions. What types of specialized interventions in ASL are effective in closing language gaps? Is training in phonological processing tasks such as imitation and recitation suitable for deaf children as young as three and four years old? What is the role of ASL rhyme and rhythm in these interventions? More specifically, do they need to learn how to successfully imitate, recognize rhymes, and recite ASL rhyme and rhythm as part of the building blocks towards stronger language foundation and emergent literacy skills? Then, there is the question of the number of interventions needed to successfully build these skills. A comprehensive evaluation of ASL phonological awareness activities over a period of time across deaf learners is needed to thoroughly investigate the effects on language and emergent literacy development. Whether ASL rhyme, rhythm, and phonological awareness can also be used to identify potential cases of sign language impairments need to be examined.

8. Conclusions

The review of the literature provides clarity into the large gaps in our empirical knowledge of the role of ASL rhyme, rhythm, and phonological awareness in facilitating language and emergent literacy in young deaf children. A body of research has been built to affirm the significance of exposing hearing children to rhyme and rhythm supplemented with training in phonological awareness for successful literacy development. Yet interventions that incorporate deaf cultural and linguistic approaches using ASL are novel to most classrooms that serve deaf children. The results of this study have implications for potential positive change at the individual, cultural, educational, and societal levels. At the individual level, the results of this study inform the field that certain interventions such as imitation training, handshape rhyme awareness and recitation of rhyming ASL stories may have a favorable impact on deaf children's language processing abilities, which are directly linked to critical emergent literacy skills in the population of young hearing children [80]. At the cultural level, deaf community members have long offered culturally-rich linguistic models through ASL storytelling, poetry, rhyme and rhythm, and language games. When the deaf community sees their linguistic and cultural capital [81] become an important part of deaf students' experience in schools, this may lead to a greater understanding, appreciation, and validation of ASL literature—including the genre of ASL rhyme and rhythm. Should this occur, there may be a shift in deaf children's relationship with language and music, making their experiences more deaf-centric and empowering. At the educational level, this study casts light on how teachers are often untrained in ASL rhyme, rhythm and phonological awareness. Without proper systemic support, deaf children are being deprived from accessing essential language exposure and experience that hearing children have. The lack of proper interventions may have an impact on deaf children's language and literacy, stalling their ability to maximize their linguistic potential. Educators can use these data to advocate for more training in culturally and linguistically responsive approaches. The findings from this study also provide a foundation for future research to explore interventions that are not only "new and better," but also specifically geared to bilingual learners such as deaf children who are primed for the benefits of metalinguistic awareness and linguistic transfers. Considering that this study has sought to address the gaps in pedagogy due to long-standing systemic barriers towards the acceptance of deaf cultural and linguistic practices, outcomes might also have implications at the societal level. This new knowledge about the role of ASL rhyme and rhythm in early childhood development may propel society to take steps towards generating a paradigm shift in uplifting deaf pedagogy.

Author Contributions: L.H. was in charge of overall direction and planning including but not limited to: conceiving the research idea, designing of the research methodology, carrying out experiments, collecting data, creating visual graphs, conducting visual analyses, providing interpretations of data, writing the manuscript, making revisions, and approving the final version. K.W. provided critical feedback during each phase of the study, assisted with running the statistical analysis, revised the manuscript, and approved the final version. All authors have read and agree to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: Holcomb co-directs a non-profit organization called Hands Land, which was established for the purposing of sharing information about ASL rhyme and rhythm to parents, teachers and the Deaf community.

Appendix A

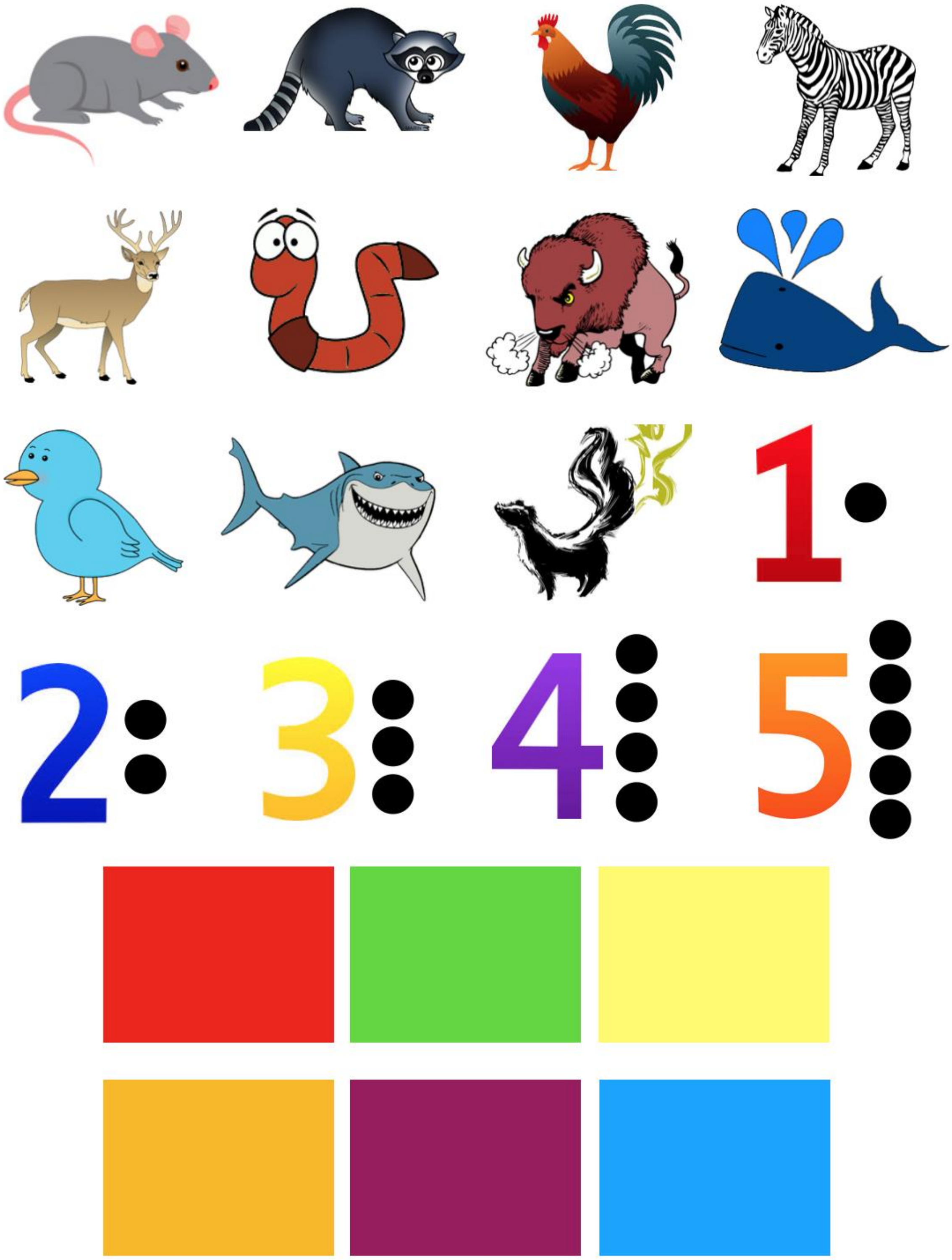


Figure A1. Researcher-Made Picture Vocabulary Assessment.

References

1. Meltzoff, A.N.; Kuhl, P.K. Exploring the infant social brain: What's going on in there. *Zero Three* **2016**, *36*, 2–9.
2. Bryant, P.E.; MacLean, M.; Bradley, L.L.; Crossland, J. Rhyme and alliteration, phoneme detection, and learning to read. *Dev. Psychol.* **1990**, *26*, 429–438. [CrossRef]
3. Brown, C.S. Language and literacy development in the early years: Foundational skills that support emergent readers. *Lang. Lit. Spectr.* **2014**, *24*, 35–49.
4. Moritz, C.; Yampolsky, S.; Papadelis, G.; Thomson, J.; Wolf, M. Links between early rhythm skills, musical training, and phonological awareness. *Read. Writ.* **2013**, *26*, 739–769. [CrossRef]
5. Moses, A.M.; Golos, D.B.; Roemen, B.; Cregan, G.E. The current state of early literacy for deaf and hearing children: A survey of early childhood educators. *J. Early Child. Lit.* **2018**, *18*, 373–395. [CrossRef]
6. Gannon, J. *Deaf Heritage: A Narrative History of Deaf America*; Gallaudet University Press: Washington, DC, USA, 2011; ISBN 978-1-56368-515-6.
7. Stokoe, W.C.; Casterline, D.C.; Croneberg, C.G. *A Dictionary of American Sign Language on Linguistic Principles*; Gallaudet University Press: Washington, DC, USA, 1965.
8. Goldin-Meadow, S.; Mayberry, R.I. How do profoundly deaf children learn to read? *Learn. Disabil. Res. Pract.* **2001**, *16*, 222–229. [CrossRef]
9. Hoffmeister, R.J.; Caldwell-Harris, C.L. Acquiring English as a second language via print: The task for deaf children. *Cognition* **2014**, *132*, 229–242. [CrossRef]
10. Miller, P.; Clark, M.D. Phonemic Awareness is Not Necessary to Become a Skilled Deaf Reader. *J. Dev. Phys. Disabil.* **2011**, *23*, 459. [CrossRef]
11. Hall, W.C. What you don't know can hurt you: The risk of language deprivation by impairing sign language development in deaf children. *Matern. Child Health J.* **2017**, *21*, 961–965. [CrossRef]
12. Strong, M.; Stuckless, R. A review of bilingual/bicultural programs for deaf children in North America. *Am. Ann. Deaf.* **1995**, *140*, 83–94. [CrossRef]
13. Crume, P.K. Teachers' perceptions of promoting sign language phonological awareness in an ASL/English bilingual program. *J. Deaf Stud. Deaf Educ.* **2013**, *18*, 464–488. [CrossRef] [PubMed]
14. Chukovsky, K. *From Two to Five*; University of California Press: Berkeley, CA, USA, 1963; ISBN 978-0-520-00238-8.
15. Petitto, L.A.; Langdon, C.; Stone, A.; Andriola, D.; Kartheiser, G.; Cochran, C. Visual sign phonology: Insights into human reading and language from a natural soundless phonology. *Wiley Interdiscip. Rev. Cogn. Sci.* **2016**, *7*, 366–381. [CrossRef] [PubMed]
16. Ferjan Ramírez, N.; Lytle, S.R.; Kuhl, P.K. Parent coaching increases conversational turns and advances infant language development. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 3484–3491. [CrossRef] [PubMed]
17. Anthony, J.; Francis, D. Development of Phonological Awareness. *Curr. Dir. Psychol. Sci.* **2005**, *14*, 255–259. [CrossRef]
18. Adams, M.J. *Beginning to Read: Thinking and Learning about Print*; The MIT Press: Cambridge, MA, USA, 1990.
19. Shanahan, T. (Ed.) *Developing Early Literacy: Report of the National Early Literacy Panel*; National Institute for Literacy: Washington, DC, USA, 2008.
20. Flett, A.; Conderman, G. Promote phonemic awareness. *Interv. Sch. Clin.* **2002**, *37*, 242–245. [CrossRef]
21. Goswami, U. Early phonological development and the acquisition of literacy. In *Handbook of Early Literacy Research*; Guilford Press: New York, NY, USA, 2003; Volume 1, ISBN 978-1-57230-895-4.
22. Patscheke, H.; Degé, F.; Schwarzer, G. The effects of training in music and phonological skills on phonological awareness in 4- to 6-year-old children of immigrant families. *Front. Psychol.* **2016**, *7*. [CrossRef]
23. Williams, M.; Rask, H. Literacy through play: How families with able children support their literacy development. *Early Child Dev. Care* **2003**, *173*, 527–533. [CrossRef]
24. Dunst, C.J.; Gorman, E. Nursery rhymes and the early communication, language and literacy development of young children with disabilities. *Cent. Early Lit. Learn.* **2011**, *4*, 1–11.
25. Georgiadou, I.; Knight, R.A.; Dipper, L. Exploring duration and isochrony in nursery rhyme reciting for children with language impairments and typically developing children. In Proceedings of the ICPhS, Glasgow, UK, 10–14 August 2015.
26. Svalberg, A.M.-L. Engagement with Language: Interrogating a Construct. *Lang. Aware.* **2009**, *18*, 242–258. [CrossRef]

27. Vaiouli, P.; Grimmet, K.; Ruich, L.J. "Bill is now singing": Joint engagement and the emergence of social communication of three young children with autism. *Autism Int. J. Res. Pract.* **2015**, *19*, 73–83. [CrossRef]
28. Gold, C.; Voracek, M.; Wigram, T. Effects of music therapy for children and adolescents with psychopathology: A meta-analysis. *J. Child Psychol. Psychiatry* **2004**, *45*, 1054–1063. [CrossRef] [PubMed]
29. Perry, M.M.R. Relating improvisational music therapy with severely and multiply disabled children to communication development. *J. Music Ther.* **2003**, *40*, 227–246. [CrossRef] [PubMed]
30. Calvert, S.L.; Billingsley, R.L. Young children's recitation and comprehension of information presented by songs. *J. Appl. Dev. Psychol.* **1998**, *19*, 97–108. [CrossRef]
31. Calvert, S.L.; Tart, M. Song versus verbal forms for very-long-term, long-term, and short-term verbatim recall. *J. Appl. Dev. Psychol.* **1993**, *14*, 245–260. [CrossRef]
32. Fernandez-Fein, S.; Baker, L. Rhyme and alliteration sensitivity and relevant experiences among preschoolers from diverse backgrounds. *J. Lit. Res.* **1997**, *29*, 433–459. [CrossRef]
33. Johnson, J.L.; Hayes, D.S. Preschool children's retention of rhyming and nonrhyming text: Paraphrase and rote recitation measures. *J. Appl. Dev. Psychol.* **1987**, *8*, 317–327. [CrossRef]
34. Sheingold, K.; Foundas, A. Rhymes for some reasons: Effect of rhyme on children's memory for detail and sequence in simple narratives. *Psychol. Rep.* **1978**, *43*, 1231–1234. [CrossRef]
35. Read, K.; Macauley, M.; Furay, E. The seuss boost: Rhyme helps children retain words from shared storybook reading. *First Lang.* **2014**, *34*, 354–371. [CrossRef]
36. Valli, C. Poetics of American Sign Language. Ph.D. Thesis, The Union Institute, Cincinnati, OH, USA, 1993.
37. Blondel, M.; Miller, C. Movement and rhythm in nursery rhymes in LSF. *Sign Lang. Stud.* **2001**, *2*, 24–61. [CrossRef]
38. Snoddon, K. Action Research with a Family ASL Literacy Program. *Writ. Pedagog.* **2011**, *3*, 265–288. [CrossRef]
39. Bonvillian, J.D.; Siedlecki, T. Young Children's Acquisition of the Location Aspect of American Sign Language Signs: Parental Report Findings. *J. Commun. Disord.* **1996**, *29*, 13–35. [CrossRef]
40. Siedlecki, T.; Bonvillian, J.D. Location, handshape & movement: Young children's acquisition of the formational aspects of American Sign Language. *Sign Lang. Stud.* **1993**, *78*, 31–52. [CrossRef]
41. Braem, P.B. Acquisition of the handshape in American Sign Language: A preliminary analysis. In *From Gesture to Language in Hearing and Deaf Children*; Springer Series in Language and Communication; Volterra, V., Erting, C.J., Eds.; Springer: Berlin/Heidelberg, Germany, 1990; pp. 107–127. ISBN 978-3-642-74859-2.
42. Conlin, K.E.; Mirus, G.R.; Mauk, C.; Meier, R.P. The acquisition of first signs: Place, handshape, and movement. In *Language Acquisition by Eye*; Lawrence Erlbaum Associates Publishers: Mahwah, NJ, USA, 2000; pp. 51–69. ISBN 978-0-8058-2937-2.
43. Di Perri, K. ASL Phonemic Awareness in Deaf Children: Implications for Instruction. Ph.D. Thesis, Boston University, Boston, MA, USA, 2004.
44. Andrews, J.F.; Baker, S. ASL Nursery Rhymes: Exploring a Support for Early Language and Emergent Literacy Skills for Signing Deaf Children. *Sign Lang. Stud.* **2019**, *20*, 5–40. [CrossRef]
45. Petitto, L.A.; Zatorre, R.J.; Gauna, K.; Nikelski, E.J.; Dostie, D.; Evans, A.C. Speech-Like Cerebral Activity in Profoundly Deaf People Processing Signed Languages: Implications for the Neural Basis of Human Language. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 13961–13966. [CrossRef]
46. Petitto, L.A.; Katerelos, M.; Levy, B.G.; Gauna, K.; Tétreault, K.; Ferraro, V. Bilingual signed and spoken language acquisition from birth: Implications for the mechanisms underlying early bilingual language acquisition. *J. Child Lang.* **2001**, *28*, 453–496. [CrossRef]
47. Petitto, L.-A. New Discoveries from the Bilingual Brain and Mind across the Life Span: Implications for Education. *Mind Brain Educ.* **2009**, *3*, 185–197. [CrossRef]
48. McQuarrie, L.; Abbott, M. Bilingual Deaf Students' Phonological Awareness in ASL and Reading Skills in English. *Sign Lang. Stud.* **2013**, *14*, 80–100. [CrossRef]
49. Mayberry, R.I.; del Giudice, A.A.; Lieberman, A.M. Reading achievement in relation to phonological coding and awareness in deaf readers: A meta-analysis. *J. Deaf Stud. Deaf Educ.* **2011**, *16*, 164–188. [CrossRef]
50. Gietz, M.R.; Andrews, J.F.; Clark, M.D. ASL Stories with Handshape Rhyme: An Exploratory Intervention to Support English Vocabulary with Signing Deaf Readers. *Arch. Psychol.* **2020**, *4*, 1–24. [CrossRef]
51. Corina, D.P.; Hafer, S.; Welch, K. Phonological awareness for American Sign Language. *J. Deaf Stud. Deaf Educ.* **2014**, *19*, 530–545. [CrossRef]

52. Quinto-Pozos, D.; Singleton, J.L.; Hauser, P.C. A case of specific language impairment in a deaf signer of American Sign Language. *J. Deaf Stud. Deaf Educ.* **2017**, *22*, 204–218. [CrossRef] [PubMed]
53. Hall, W.C.; Levin, L.L.; Anderson, M.L. Language deprivation syndrome: A possible neurodevelopmental disorder with sociocultural origins. *Soc. Psychiatry Psychiatr. Epidemiol.* **2017**, *52*, 761–776. [CrossRef] [PubMed]
54. Henner, J.; Novogrodsky, R.; Reis, J.; Hoffmeister, R. Recent issues in the use of signed language assessments for diagnosis of language disorders in signing deaf and hard of hearing children. *J. Deaf Stud. Deaf Educ.* **2018**, *23*, 307–316. [CrossRef] [PubMed]
55. Cheng, Q.; Roth, A.; Halgren, E.; Mayberry, R.I. Effects of Early Language Deprivation on Brain Connectivity: Language Pathways in Deaf Native and Late First-Language Learners of American Sign Language. *Front. Hum. Neurosci.* **2019**, *13*. [CrossRef] [PubMed]
56. Mayberry, R.I. First-language acquisition after childhood differs from second-language acquisition: The case of American Sign Language. *J. Speech Hear. Res.* **1993**, *36*, 1258–1270. [CrossRef]
57. Mayberry, R.I.; Eichen, E.B. The long-lasting advantage of learning sign language in childhood: Another look at the critical period for language acquisition. *J. Mem. Lang.* **1991**, *30*, 486–512. [CrossRef]
58. Mayberry, R.I.; Fischer, S.D. Looking through phonological shape to lexical meaning: The bottleneck of non-native sign language processing. *Mem. Cognit.* **1989**, *17*, 740–754. [CrossRef]
59. Barlow, D.H.; Hayes, S.C. Alternating treatments design: One strategy for comparing the effects of two treatments in a single subject. *J. Appl. Behav. Anal.* **1979**, *12*, 199–210. [CrossRef]
60. Hains, A.H.; Baer, D.M. Interaction effects in multielement designs: Inevitable, desirable, and ignorable. *J. Appl. Behav. Anal.* **1989**, *22*, 57–69. [CrossRef]
61. Horner, R.H.; Carr, E.G.; Halle, J.; McGee, G.; Odom, S.; Wolery, M. The use of single-subject research to identify evidence-based practice in special education. *Except. Child.* **2005**, *71*, 165–179. [CrossRef]
62. What Works Clearinghouse Reviewer Guidance for Use with the Procedures and Standards Handbook. Available online: https://ies.ed.gov/ncee/wwc/Studyreviewguide (accessed on 15 March 2019).
63. Simms, L.; Baker, S.; Clark, D.M. The Standardized Visual Communication and Sign Language Checklist for Signing Children. *Sign Lang. Stud.* **2013**, *14*, 101–124. [CrossRef]
64. Hammond, D.; Gast, D.L. Descriptive analysis of single subject research designs: 1983–2007. *Educ. Train. Autism Dev. Disabil.* **2010**, *45*, 187–202.
65. Manolov, R.; Onghena, P. Analyzing data from single-case alternating treatments designs. *Psychol. Methods.* **2018**, *23*, 480–504. [CrossRef] [PubMed]
66. Shadish, W.R.; Sullivan, K.J. Characteristics of single-case designs used to assess intervention effects in 2008. *Behav. Res. Methods* **2011**, *43*, 971–980. [CrossRef] [PubMed]
67. Smith, J.D. Single-case experimental designs: A systematic review of published research and current standards. *Psychol. Methods* **2012**, *17*. [CrossRef] [PubMed]
68. Kratochwill, T.R.; Hitchcock, J.H.; Horner, R.H.; Levin, J.R.; Odom, S.L.; Rindskopf, D.M.; Shadish, W.R. Single-case intervention research design standards. *Remedial Spec. Educ.* **2013**, *34*, 26–38. [CrossRef]
69. Lane, J.D.; Gast, D.L. Visual analysis in single case experimental design studies: Brief review and guidelines. *Neuropsychol. Rehabil.* **2014**, *24*, 445–463. [CrossRef]
70. Gast, D.L. *Single Subqject Research Methodology in Behavioral Sciences*; Routledge: New York, NY, USA, 2010; ISBN 978-1-135-59319-3.
71. Lindahl, M.; Pramling-Samuelsson, I. Imitation and Variation: Reflections on toddlers' strategies for learning. *Scand. J. Educ. Res.* **2002**, *46*, 25–45. [CrossRef]
72. Bolduc, J.E.; Lefebvre, P.A.E. Using nursery rhymes to foster phonological and musical processing skills in kindergarteners. *Creat. Educ.* **2012**, *3*, 495–502. [CrossRef]
73. Tillmann, B.; Dowling, W.J. Memory decreases for prose, but not for poetry. *Mem. Cognit.* **2007**, *35*, 628–639. [CrossRef]
74. Bower, G.H.; Bolton, L.S. Why are rhymes easy to learn? *J. Exp. Psychol.* **1969**, *82*, 453–461. [CrossRef]
75. Engel, S. Storytelling in the First Three Years. *Zero Three* **1997**, *2*, 3–6.
76. Bloom, L.; Hood, L.; Lightbown, P. Imitation in language development: If, when, and why. *Cogn. Psychol.* **1974**, *6*, 380–420. [CrossRef]
77. Long, L.E. Productions of Metalinguistic Awareness by Young Children with SLI and Typical Language. Master's Thesis, East Tennessee State University, Johnson City, TN, USA, 2015, unpublished dissertation.

78. Chaney, C. Language development, metalinguistic skills, and print awareness in 3-year-old children. *Appl. Psycholinguist.* **1992**, *13*, 485–514. [CrossRef]
79. Fazio, B.B. Memory for Rote Linguistic Routines and Sensitivity to Rhyme: A Comparison of Low-Income Children with and without Specific Language Impairment. *Appl. Psycholinguist.* **1997**, *18*, 345–372. [CrossRef]
80. Ozernov-Palchik, O.; Norton, E.S.; Sideridis, G.; Beach, S.D.; Wolf, M.; Gabrieli, J.D.E.; Gaab, N. Longitudinal stability of pre-reading skill profiles of kindergarten children: Implications for early screening and theories of reading. *Dev. Sci.* **2017**, *20*, 12471. [CrossRef]
81. Bourdieu, P. The forms of capital. In *Handbook of Theory and Research for the Sociology of Education*; Greenwood Press: New York, NY, USA, 1986; pp. 241–258.

children

MDPI

Case Report

Characterization of Speech and Language Phenotype in GLUT1DS

Martina Paola Zanaboni [1,†], Ludovica Pasca [1,2,*,†], Barbara Valeria Villa [1], Antonella Faggio [1], Serena Grumi [1], Livio Provenzi [1], Costanza Varesio [1,2,†] and Valentina De Giorgis [1,†]

1 Department of Child Neurology and Psychiatry, IRCCS Mondino Foundation, 27100 Pavia, Italy; martinapaola.zanaboni@mondino.it (M.P.Z.); barbaravaleria.villa02@universitadipavia.it (B.V.V.); antonella.faggio@mondino.it (A.F.); serena.grumi@mondino.it (S.G.); livio.provenzi@mondino.it (L.P.); costanza.varesio@mondino.it (C.V.); valentina.degiorgis@mondino.it (V.D.G.)
2 Department of Brain and Behaviour Neuroscience, University of Pavia, 27100 Pavia, Italy
* Correspondence: ludovica.pasca01@universitadipavia.it; Tel.: +39-0382-380289
† Member of ERN-EpiCARE.

Abstract: Background: To analyze the oral motor, speech and language phenotype in a sample of pediatric patients with GLUT 1 transporter deficiency syndrome (GLUT1DS). Methods: eight Italian-speaking children with GLUT1DS (aged 4.6–15.4 years) in stable treatment with ketogenic diet from a variable time underwent a specific and standardized speech and language assessment battery. Results: All patients showed deficits with different degrees of impairment in multiple speech and language areas. In particular, orofacial praxis, parallel and total movements were the most impaired in the oromotor domain; in the speech domain patients obtained a poor performance in the diadochokinesis rate and in the repetition of words that resulted as severely deficient in seven out of eight patients; in the language domain the most affected abilities were semantic/phonological fluency and receptive grammar. Conclusions: GLUT1DS is associated to different levels of speech and language impairment, which should guide diagnostic and therapeutic intervention. Larger population data are needed to identify more precisely a speech and language profile in GLUT1DS patients.

Keywords: GLUT 1 transporter deficiency syndrome (GLUT1DS); language; speech; oral motor; dysarthria

Citation: Zanaboni, M.P.; Pasca, L.; Villa, B.V.; Faggio, A.; Grumi, S.; Provenzi, L.; Varesio, C.; De Giorgis, V. Characterization of Speech and Language Phenotype in GLUT1DS. *Children* **2021**, *8*, 344. https://doi.org/10.3390/children8050344

Academic Editor: Eva Aguilar Mediavilla

Received: 25 March 2021
Accepted: 23 April 2021
Published: 27 April 2021

1. Introduction

GLUT 1 transporter deficiency syndrome (GLUT1DS) is a rare, treatable, metabolic encephalopathy due to mutations in SLC2A1 gene [1], which causes a non-functional glucose uptake by GLUT1 transporter, primarily expressed in endothelial cells forming the blood-brain barrier and in astrocytes [2]. Ketogenic dietary therapies (KDTs) are recognized as the gold standard treatment for GLUT1DS since they provide alternative fuel, namely ketone bodies, for brain energy metabolism [2]. Symptoms develop in age-specific pattern [2] and the classical disease phenotype includes a wide range of movement disorders, drug-resistant epilepsy, neurodevelopmental impairment, and acquired microcephaly. Moreover, ataxia, dystonia, dysarthria, persistent tremor, spasticity are typical findings at neurological examination. Milder and atypical phenotypic variants are continuously reported and GLUT1DS phenotypic spectrum is progressively expanding. KDTs introduction, especially when occurring early in life, could lead to improvement of certain symptoms, as often occurs with epileptic manifestations, intellectual and social adaptive skills [3], whereas its beneficial effects on movement disorders are less evident [2].

Intellectual disability (ID) is a usual finding in GLUT1DS patients, ranging from severe to mild; only a minority of affected patients shows normal intelligence quotient (IQ) [4]. Genotype-phenotype correlation has not been clearly defined so far. Some reports described a milder phenotype or later disease onset in patients with missense mutations or higher cerebrospinal fluid (CSF)/blood glucose ratio [5]. The existence of a proportional relationship between ID and disease severity is debated: there have been reports both supporting and disproving this relation [6–8]. De Giorgis et al. defined a typical cognitive phenotype of GLUT1DS [4], evaluating the neuropsychological profile of 25 patients before and after KDTs introduction. The typical phenotype observed was characterized by a performance IQ more affected than verbal IQ (VIQ), together with greater difficulties in visuospatial and visuomotor skills. In the same study, a significant direct correlation between IQ (total IQ (TIQ) and VIQ) and CSF/blood glucose ratio values was observed.

Speech and language functions in patients with GLUT1DS are the least documented in literature and, to the best of our knowledge, they have only been assessed in the context of broad neuropsychological batteries [5,6,9,10] and have not been deeply characterized yet. To date, the presence of different degrees of speech and language impairments, with varying degrees of motor incoordination, have been described as common features in children and adults with GLUT1DS [2]. In particular, Hully and colleagues [6] reported language delay with dysarthric speech in almost 80% of GLUT1DS patients. In the study by Ramm-Pettersen et al. [10], an improvement of those aspects was recorded after KDTs introduction: the caregivers of six patients reported progress concerning general alertness, expressive language, articulation, and physical endurance in the wake of the dietary treatment. The sooner the KDTs is introduced, the greater is its potential of changing the disease course [2].

The phenotypic variability and the different response to KDTs therapy according to introduction timing should be read considering that glucose utilization in the brain increases threefold from infancy to 3 years of age [11] and thus, for patients with GLUT1DS, that period represents a critical frame if left untreated. Moreover, the neuroconstructivism approach underlies the bidirectional interactions between human biology and social environment, the development itself might be seen as playing a crucial role in shaping phenotypical outcomes [12].

We believe that recognition of a typical speech and language profile might be crucial for disease assessment and targeted rehabilitation.

For this purpose, the present study aimed to characterize in detail language development and to analyze the oral-motor, speech and language phenotype in a pediatric population of patients with GLUT1DS, through the use of standardized tests.

2. Materials and Methods

2.1. Participants

This was a mono-center study. Among 25 patients with established clinical and genetic diagnosis of GLUT1DS of all ages treated with KDTs and regularly followed at our clinic [4], we included eight monolingual Italian-speaking patients (four females and four males), aged 4.6–15.4 years (median age 10; standard deviation (SD) 3.79; range 4.6–15.4 years).

For each patient, information such as sociodemographic and clinical variables were collected (child's pre-peri and post-natal clinically events, developmental history, presence of otitis' history or auditory impairment, previous diagnosis other than GLUT1DS, time of GLUT1DS diagnosis, CSF/blood glucose ratio, SLC2A1 mutation, family history of GLUT1DS, family history of epilepsy, epilepsy history, presence of movement disorder, KDTs and rehabilitation initiation timing) (see Table 1).

Table 1. Patients' clinical history

	Pt1	Pt2	Pt3	Pt4	Pt5	Pt6	Pt7	Pt8	All (n = 8)
Age at evaluation (years)	6.1	13.9	15.4	4.6	10.3	11.2	11.4	7.11	M 10 (SD 3.79) range 4.6–15.4
Age at GLUT1DS Diagnosis (months)	59	60	75	24	122	79	95	69	M 72.87 (SD 26.70) range 24–122
CSF/serum glucose ratio	0.37	NA	0.34	0.27	NA	0.35	0.43	0.37	-
Genotype	DeletionExones1–10	Missense R400C	Deletion 1p34.2	Truncatin Y366X	Missense G314S	Missense N34S	Missense G17R	Frameshift L486fs	-
GLUT1DS Family history	no	no	no	no	yes	yes	no	no	yes 25%
Perinatal Period	un	un	un	un	un	un	un	Prematurity jaundice	yes 12.5%
Recurrent otitis	no	yes	no	no	no	no	no	no	yes 12.5%
Auditory deficits	no	no	no	no	no	no	no	no	yes 0%
Babbling onset (months)	18	12	24	10	9	10	10	na	M 13.29 (SD 5.61) range 9–24
Age at first word (months)	30	24	24	20	12	12	12	12	M 18.25 (SD 7.2) range 12–30
Age at combinatory speech (months)	60	24	60	24	24	24	24	36	M 34.5 (SD 16.27) range 24–60
Preschool speech intelligibility	no	yes	no	no	yes	yes	yes	yes	no 37.5%
Language disorder	dysarthria	speech	dysarthria	speech	speech	none	none	none	speech 37.5% dysarthria 25%
Psychomoto development	delayed	normal	delayed	normal	normal	normal	normal	normal	delayed 25%
Intellectual disability	borderline	borderline	moderate	normal	borderline	borderline	mild	borderline	-
Epilepsy Family history	yes	no	yes	no	yes	yes	no	no	yes 50%
Epilepsy	yes	yes	yes	no	yes	yes	yes	yes	yes 87.5%
Epilepsy onset (months)	24	36	26	na	84	60	12	12	M 36.28 (SD24.71) range 12–84
Movement disorder	yes	yes	yes	yes	yes	yes	yes	yes	yes 100%

Table 1. *Cont.*

	Pt1	Pt2	Pt3	Pt4	Pt5	Pt6	Pt7	Pt8	All (n = 8)
Movement disorder severity	moderate	mild	mild	mild	mild	mild	mild	mild	mild 87.5% moderate 12.5%
Age at KDTs Initiation (months)	60	60	98	24	121	88	94	69	M 76.75 (SD 29.85) range 24–121
Rehabilitation therapy	speech	psychomotor & speech	speech	psychomotor & speech	speech & cognitive	psychomotor & speech	none	psychomotor	speech 75% psychomotor 50% cognitive 12.5%
Rehabilitation onset (months)	24	60	24	24	108	60	na	48	M 49.71 (SD 30.53) range 24–108
Rehabilitation's duration (years)	6	6	8	3	1.5	7	na	2	M 4.78 (SD 2.57) range 1.5–8

Abbreviations: M, mean; SD, standard deviation; n, number; na, not available; KDTs, Ketogenic dietary therapies; Un, unremarkable.

In detail, five patients have a missense mutation, two patients have a deletion of SLC2A1 gene and one patient a truncating mutation. Pregnancy was uncomplicated in all but one patients. No patient suffered from recurrent otitis media during the first years of life, nor auditory deficits. Based on parental reports, the following information was obtained: babbling onset was mainly delayed (age 13.29 months; SD 5.61; range 9–24) and so was the age of first-word onset (age 18.25 months; SD 7.2; range 12–30) and age of combinatory speech (age 34.5 months; SD 16.27; range 24–60). Three patients showed unintelligible speech during the preschool years. Psychomotor development was delayed in two patients. Before the genetic and clinical diagnosis of GLUT1DS, three patients were diagnosed with a speech disorder and two patients with 'dysarthria and ataxia' of unknown origin at first disease manifestations. Age at GLUT1DS diagnosis ranged from 24 to 122 months (mean 72.87 months; SD 26.70). Two patients had a family member affected by GLUT1DS and four patients had a family history of epilepsy. Epilepsy onset in the described sample occurred from 12 to 84 months (mean 36.28 months; SD 24.71). The median age at KDTs initiation was 76.75 months (SD 29.85; range 24–121 months). All patients had movement disorders: seven with mild severity and one with moderate severity. Rehabilitation was provided in all cases but one: six patients underwent speech and language therapy, four patients psychomotor therapy, and one patient underwent cognitive rehabilitation. Rehabilitation started at a median age of 49.71 months (SD 30.53, range 24–108) and the therapy duration reported lasted from 1.5 years to eight years (median 4.78 years, SD 2.57).

All patients were on stable KDTs therapy from more than six months at the evaluation time.

2.2. Materials and Procedures

All patients underwent a cognitive and speech and language evaluation. Speech and language assessment investigated three domains: oromotor, speech, and language abilities (see Table 2 for a detailed list of standardized tests performed).

Table 2. Description of the tests included in the administered battery.

Task	Description
Parental report on clinical history	Child's pre-peri and post-natal clinically events and speech and language milestones acquisition.
Oromotor skills	Oromotor skills were examined with Orofacial Praxis [13]. Oromotor skills were the ability to plan and execute movements or sequences of voluntary movements, meaningful or not, using the muscles of the pharyngo-buccofacial system or the orofacial region. The Orofacial Praxis Test, consisting of 36 gestures, 24 single and 12 complex, elicited through verbal and imitative request.
Phonetic inventory	Phonetic inventory was investigated with the Articulation Test of Fanzago [14].This instrument was based on spontaneous/repetition elicited denomination of 114 figures which named allow to verify whether the target phoneme (place in different positions within the word) has been produced correctly or replaced/omitted/distorted.
Phonological Planning	Phonological Planning was tested by the Repetition of 31 words pronounced by the examiner [15]. This subtest is designed to assess phonological encoding and decoding through the repetition of words and it allows to detect the presence of phonological processes. For each word it is possible to calculate the number and the type of phonological processes produced. It is possible to identify two phonological processes, simplification and atypical. Simplification processes represent the persistence of normal primitive processes in successive stages of phonological development. Atypical idiosyncratic processes included types of simplifications rarely found in normal language development, or those that that are never found in normal developmental processes.
Diadochokinesis	Diadochokinesis was assessed with Maximum performance rate [16]. This task is used to test the ability to repeat a syllable sequence (/pataka/) as quickly as possible for 20 s in order to look at motor speech skills separate from the effects related to word familiarity.
Receptive vocabulary	Receptive vocabulary was evaluated by PPVT-III [17]. The PPVT is a receptive vocabulary test in which the child points to one of four pictures on a page that is named by the examiner.

Table 2. *Cont.*

Task	Description
Expressive vocabulary	Expressive vocabulary was tested by the Name BVN 5-11 [18]. For this task, the subject is asked to name 20 (for children aged from 5 to 11) or 88 (for children aged from 12 to 18) figures in order to measure patient's vocabulary ability.
Receptive grammar	Receptive grammar was examined with Comprehension of Instructions NEPSY [19]. This task assesses receptive language and it involves understanding verbal instructions and processing them into actions.
Expressive grammar	Expressive grammar was evaluated by Sentence repetition NEPSY II [19]. This task was used to investigate the production of grammar structures.
Verbal Fluency	Verbal Fluency was examined with Word generation NEPSY II [19]. This subtest is designed to assess verbal productivity through the ability to generate words and it consists of two tasks: semantic or phonemic fluency. The participants are given 1 min to generate as many words as possible within a semantic category or they are asked to say words that start with a given letter.
Cognitive assessment	Age-appropriate versions of the Wechsler scales were administered to assess intellectual ability: —from 6 to 16 years the Wechsler Intelligence Scale for Children—Fourth Edition [20] or 2–6 year olds, the Wechsler Preschool and Primary Scale of Intelligence—Third Edition, Italian version (WPPSI-III) [21] and –Full-scale IQ (FSIQ) scores were derived and classified according to test manual normative data.

Test administration was carried out individually by a professional neuropsychologist and a speech and language therapist. Both speech and language and cognitive assessments were performed by administering a comprehensive battery of tests depending on patient's age. Language development patterns were reviewed by an expert neuropsychologist and a speech and language therapist through a standardized questionnaire and by collecting detailed medical and developmental milestones history.

This study was approved by our Ethical Committee (P-20190033749), IRCCS Mondino Foundation, Pavia. Written informed consent was obtained from caregivers.

2.3. Statistical Analysis

We firstly performed a set of descriptive analyses. Then, a set of Spearman's rank correlation (rho) coefficients were calculated between the age onset of babbling, age at first word, age at combinatory speech, and the neuropsychological tasks' results.

3. Results

3.1. Tests Evaluation

3.1.1. Oromotor Skills

Oromotor skills were evaluated with nonverbal tasks, and demonstrated impaired functioning on at least one subtest. The worst performances were obtained in orofacial praxis verbal requests, parallel movements, and total score. In particular, in orofacial praxis verbal request (mean z-score −2.72, SD 3.1; range −9.66 0.75) 4/8 subjects showed a severely impaired performance and 3/8 had a poor performance; in 'parallel movements' 5/8 subjects scored a severely impaired performance (mean z-score −3.08, SD 3.42; range −6.09–1.10) and in 'total score verbal request condition' (mean z-score −2.63, SD 2.29; range −6.83 0.43) 4/8 patients had a severely impaired result and 2/8 patients a mildly impaired score. See Table 3.

Table 3. Results of neuropsychological assessment.

	Pt1	Pt2	Pt3	Pt4	Pt5	Pt6	Pt7	Pt8	All (*n* = 8)
Voiced praxis verbal request (zs)	−2	1.37	−2	−1.70	−3.68	−1.16	1.37	−0.3	M −1.01 (SD 1.74) range −3.68 1.37
Voiced praxis imitation (zs)	−3.22	0.54	0.54	−1.17	−5.11	−1.33	0.54	−3.22	M −1.55 (SD 2.12) range −5.11 0.54
Orofacial praxis verbal request (zs)	−1.33	−3.41	−1.33	−2.11	−9.66	−3.41	0.75	−1.33	M −2.72 (SD 3.1) range −9.66 0.75
Orofacial praxis imitation (zs)	−1.90	0.53	−1.73	0.71	−1.90	−1.90	0.53	0.53	M −0.64 (SD1.3) range −1.9 0.71
Sequence movements verbal request (zs)	−2.91	−0.5	−2.84	0.38	−0.5	−0.5	0.69	−2.91	M −1.13 (SD 1.51) range −2.91 0.69
Sequence movements imitation (zs)	−4.74	0.43	−3	−0.28	0.43	0.43	0.43	−1.29	M −0.94 (SD 1.95) range −4.74 0.43
Parallel movements verbal request (zs)	−6.09	−2.80	0.69	1.10	−6.09	−6.09	−6.09	0.69	M −3.08 (SD 3.42) range −6.09 1.10
Parallel movements imitation (zs)	−10.31	0.21	0.21	0.75	0.21	−5.05	−5.05	0.21	M −2.35 (SD 4.03) range −10.31 0.75
Total score verbal request (zs)	−4.41	−0.77	−3.2	−1.15	−6.83	−3.2	0.43	−1.98	M −2.63 (SD 2.29) range −6.83 0.43
Total score imitation (zs)	−6.26	0.75	−1.87	0.74	−2.75	−1.87	−0.12	−1.52	M −1.61 (SD 2.27) range −6.26 0.75
Phonetic inventory: absent phonemes (des)	d g ʧ ʤ s z ts ʣ r ʎ	r ʎ	z s r ʎ st ts	r ʎ	none	r	ʎ r	ts ʣ ʎ r	-
Present phonemes (consonants and vowels) (*n*)	16	24	22	24	26	25	25	22	M 23 (SD 3.16) range 16–26
Phonological Planning (zs)	−31.5	−10.96	−22.38	−9.73	−21.61	−9.23	−1.29	na	M −15.24 (SD 9.52) range −31.5 −1.29
Phonological Planning (word accuracy) (*n*)	0/31	24/31	20/31	8/31	18/31	26/31	29/31	na	-
Phonological Planning (syllabic structure accuracy) (*n*)	26/31	28/31	28/31	31/31	30/31	31/31	31/31	na	-
Phonological Planning (word length accuracy) (*n*)	8/31	24/31	26/31	15/31	25/31	30/31	31/31	na	-
Diadochokinesis (zs)	−7.86	−0.53	−1.2	−3.2	−0.86	−3.2	0.8	na	M −0.27(SD 3.48) range −7.86 0.8
Receptive vocabulary (st)	75	113	83	106	91	87	85	77	M 89.63 (SD 13.42) range 75–113
Expressive vocabulary (zs)	−1.9	0.03	−5.45	−0.1	−1	0	0.35	0.77	M −0.91 (SD 2.01) range 5.45–0.77
Receptive Grammar (ss)	3	6	1	8	8	1	5	6	M 4.75 (SD2.81) range 1–8
Expressive Grammar (ss)	7	9	1	10	6	5	9	12	M 7.37 (SD 3.42) range 1–12
Word generation: semantic fluency (ss)	3	7	3	9	4	6	4	9	M 5.62 (SD 2.5) range 3–9
Word generation: phonemic fluency (ss)	na	3	3	na	6	6	4	na	M 4.40 (SD 1.51) range 3–6
FSIQ (st)	76	81	41	96	71	73	69	80	M 73.38 (SD 15.55) range 41–96
VCI (st)	88	82	66	98	90	84	74	98	M 85 (SD 11.11) range 66–98
PRI (st)	71	102	48	93	80	76	65	73	M 76 (SD 16.56) range 48–102
WMI (st)	na	76	52	na	64	76	91	na	M 71.80 (SD 14.63) range 52–91
PSI (st)	58	79	56	na	74	82	82	67	M 71.14 (SD 10.99) range 56–82

Abbreviations: zs = zeta score; M = mean; SD = standard deviation; des = descriptive; n = number; na = not available; ss = scaled score; FSIQ = Full scale intelligence quotient; st = standard score; VCI = Verbal Comprehension Index; PRI = Perceptual Reasoning Index; WMI = Working memory index; PSI = Processing speed index.

3.1.2. Speech

All but one individual demonstrated impaired articulation, characterized by imprecise production of consonants (7/8) and vowels (1/8). Only one child had acquired all speech sounds expected for his age. Most absent consonants were vibrant [r] (7/8), followed by voiced palatal lateral approximant [ʎ] (6/8, fricatives [z], [s] (2/8) and affricates [ʤ], [ʦ] (2/8), both voiced and voiceless. Phonological planning was strongly impaired in all children but one, who showed a mild impairment (mean z-score −15.24, SD 9.52; range −31.5 −1.29). All patients showed a high percentage of phonetically inaccurate production of words.

We identified simplification processes (backing 3/8; fronting 4/8; stopping 3/8; epenthesis 6/8; metathesis 2/8; cluster reduction 5/8; gliding 1/8; voicing2/8; de-voicing 3/8; affrication 2/8; affrication 3/8; assimilation 4/8; diphthong reduction 2/8) and atypical processes (stops deletion/reduction in clusters 3/8; conflicting processes 3/8; idiosyncratic processes 4/8). Speech was typified by the imprecise articulation of consonants and vowels, abnormal nasal resonance, low pitch, and prosodic errors (e.g., excessive stress on unstressed parts of speech, slow rate, short phrases). Evaluation of Diadochokinesis rate was significantly slower in 3/7 patients, mildly impaired in 1/7 (mean z-score 0.7, SD 3.48; range −7.86 0.8). Conversation speech intelligibility was adequate in seven out of eight patients. See Table 3.

3.1.3. Language

Lexical abilities were preserved in 6/8 patients, both in receptive (mean standard score 89.63, SD 13.42; range 75–113) and expressive vocabulary (mean z-score −0.91, SD 2.01; range −5.45 0.77). Receptive vocabulary resulted in the normal borderline range in 2/8 patients. Expressive vocabulary was severely impaired in one patient, moderately impaired in two patients, and normal in five patients. Receptive grammar was the most impaired domain (mean scaled score 4.75, SD 2.81; range 1–8): a severe impairment was seen in 4/8 patients, two patients showed a borderline normal score, and 2/8 patients obtained a normal score. Expressive grammar (mean scale score 7.37, SD 3.42; range 1–12) was found impaired just in 1/8 patients, 3/8 performed borderline normal and 4/10 normal. See Table 3.

3.1.4. Intelligence Quotient

Full-scale IQ (FSIQ) scores varied from moderately impaired to normal (range 41–96; median 73.38; SD 15.55). Verbal Comprehension Index (VCI) was more conserved compared to the remaining sub-IQ, with a median score of 85 (SD 11.11; range 66–98). See Table 3.

3.2. Correlations

The results of the Spearman correlation indicated that there was a significant negative association between age at the babbling onset and phonemic fluency tasks' score and phonemic inventory. Age at the babbling onset was inversely related to phonemic fluency task and the number of consonants acquired: the greater the delay in babbling's onset, the smaller the number of words produced at the phonemic fluency task [$\rho(5) = -0.89$, $p = 0.042$] and the number of consonants acquired [$\rho(7) = -0.90$, $p = 0.005$]. Age at the first word onset significantly correlated with phonemic fluency tasks' score [$\rho(5) = -0.91$, $p = 0.03$] and phonemic inventory [$\rho(8) = -0.71$, $p = 0.049$]. Children who produced their first word earlier had a better performance in phonemic fluency tasks [$\rho(5) = -0.91$, $p = 0.03$] and they had a greater number of stable consonants [$\rho(8) = -0.71$, $p = 0.049$]. See Table 3.

Moreover, we attempted a correlation between the severity of language evaluations (lower scores in language and oromotor assessments) and genotype. Due to the small sample, we could not find a significance; in particular, as reported in literature about clinical phenotype severity [7], patients with deletions ($n = 2$) or truncating mutation ($n = 1$) do not seem to have a more impaired profile compared to patients with missense mutations ($n = 5$).

The same correlation was attempted between language evaluation and CSF/serum glucose ratio without significative results due to the small numbers of patients (n = 6) and the small range of values (mean 0.35, range 0.27 −0.043).

It has not been feasible to search for different functioning trajectories according to age of KDTs initiation. Nevertheless, in our sample, patients who started KDTs after six years of age achieved lower scores in language assessment (n = 4, mean 100.25 months, range 88–121 months). Regarding oromotor skills, we did not observe the same trend.

4. Discussion

Speech and language impairment have already been recognized in patients with GLUT1DS, but have not been fully characterized compared to other disease symptoms. In available studies, language functioning in GLUT1DS is depicted as extremely variable, ranging from no apparent deficit to the absence of expressive speech, with most affected individuals having reduced language skills [22,23]. In the present study, we deeply investigated speech and language profile in eight Italian-speaking children with GLUT1DS.

Based on parental reports, we documented a delay of early vocal behavior and early language milestones with a late onset of first word and combinatory speech in the majority of patients. We also found a significant negative association between babbling onset and the number of words produced in the phonemic fluency task and phonetic inventory. The delay in the mean age of babbling onset represents a crucial finding, since several studies support the predictive value of babbling onset timing and characteristics to determine subsequent speech and language abilities and communication disorders [24,25]. Babbling represents a linguistic and articulatory exercise and the experience of frequent self-producing consonants and vowels syllables makes infants more aware of similar patterns in their environmental language, acting as potential building blocks for word representations [25]. Moreover, in our sample the age at the first word onset significantly correlated with phonemic fluency tasks' score and phonemic inventory, meaning that children who produced their first word earlier had a better performance in phonemic fluency tasks and a greater number of stable consonants. Importantly, it is often hypothesized that the first speech-like articulation and the babbling phase, which occur at approximately ten months of age, allow infants to develop a link between articulatory settings and the resulting auditory consequences, thus contributing to the development of the phonetic inventory and adaptation to the ambient language [26]. In this connection, the early signs of speech and language deviance and slow acquisition of expressive words in the second and third years of life may set off a cascade, negatively affecting a variety of following additional linguistic capabilities [24]. This scenario, which is frequently reported in cognitive and language disorders, has never been described as associated with GLUT1DS previously.

Oral-motor skills were impaired in most subjects in our sample. Development of orofacial praxis is impaired in a series of developmental disorders such as Developmental Coordination Disorder, Developmental Apraxia of Speech and Speech disorder [27]. These conditions have in common the combined presence of motor and language deficits, as observed in patients with GLUT1DS.

Speech was often characterized by phonetically inaccurate production of words, imprecise articulation of consonants and vowels, abnormal nasal resonance, low pitch, and prosodic errors. The most represented impairment was found in the phonological planning. This task resulted as severely deficient in seven out of eight patients, confirming the presence of a speech and language disorder, still active in some patients, and partially compensated in others. Receptive and expressive language abilities revealed different degrees of impairment in our patients; some of them showed severe receptive and expressive linguistic deficits, others had a mild impairment and only one had a normal profile. In all patients, a more conserved expressive and receptive lexical competence was observed, while linguistic grammar ability was impaired with a greater compromise of the receptive abilities. We may assume that a severe impairment at the morpho-syntactic level of language organization could be interpreted as the less likely domain to recover in patients with a previous speech and language disorder, as observed in GLUT1DS patients.

Several reports describe a mild-to-severe intellectual disability of GLUT1DS patients, in most cases proportional to the disease's severity [6–8]. In our sample, FSIQ scores varied from moderately impaired to normal, one child showed a normal intelligence, five patients had a borderline intellectual functioning, two patients received a diagnosis of intellectual disability on mild and moderate ranges. VCI showed up as more conserved compared to the remaining sub-IQ: these data confirm the results of our previous work, where PRI was more affected than VCI [4]. A less impaired verbal quotient could lead at first to a misidentification of language deficits but, as shown by our results, an impairment of several linguistic domains can be documented with focused tests.

Due to the small number of patients included, it has not been feasible to obtain a phenotype-genotype and/or a phenotype/glycorrhachia correlation, as well as to search for different functioning trajectories according to age of KDTs initiation or total IQ level.

Nevertheless, in our sample, patients who started KDTs later in life (mean 8.5 years) achieved lower scores in language assessment and the patient with lowest IQ achieved one of the worst performances. Definitely larger samples are needed to assess whether KDTs initiation timing and mutation type might influence chances of recovery of speech and language. Unfortunately, in our sample KDTs introduction was late for all included patients.

Children with GLUT1DS are at a disadvantage in the development of cognitive functions since the disease itself causes a lower supply of energy for the correct functioning of the brain, resulting in a multilevel dysfunction affecting cognitive, speech and language abilities, as evidenced by the neuropsychological and language assessment carried in our sample [4]. Our data confirm the presence of a potentially heterogeneous cognitive and linguistic profile with different degrees of impairment in multiple speech and language areas. The variability of the linguistic profiles observed could be explained based on the general theoretical framework of neuroconstructivism [12].

This model is suitable to understand the interaction between biological and socio-environmental factors determining the linguistic development of patients with GLUT1DS.

The neuroconstructivism approach highlights how tiny variations in the initial state could give rise to domain-specific differences in end states [12]. If brain energy requirements are not satisfied in the first years of life, an impairment of input processing and starting points such as language circuitry will occur. Variability of genetic mutations, adaptive strategies, successful behavior as well as intact domains leads to inter-individual outcome differences, that could explain the relative heterogeneity of language profile in our small sample. We did not find factors determining language outcome; nevertheless, we believe that focus must be placed in at-risk populations in early infancy, even before onset of language, and that this time window should represents the optimal timing to start therapy, namely KDTs.

Limitations of this study are represented by the small number of subjects included—also due to the low prevalence of the disease—the age heterogeneity and the absence of a language and speech assessment before KDTs introduction.

5. Conclusions

In conclusion, GLUT1DS can be considered a multilevel condition affecting cognitive, motor, speech, and language competences. Our results confirm the importance of a complete speech and language evaluation to obtain a detailed profile, that is crucial to plan early and specific rehabilitative intervention.

GLUT1DS patients are often diagnosed with aspecific language disorder or delay in the first years of life, before other symptoms manifest. In this scenario, recognizing typical and atypical language fragilities and searching for a common linguistic phenotype in these patients could help to guide early diagnosis. An early diagnosis of GLUT1DS would allow a prompt start of target dietary treatment and of rehabilitative intervention inclusive of speech and language training. Further studies are needed to evaluate the effects of KDTs on language function.

Author Contributions: Conceptualization, M.P.Z. and L.P. (Ludovica Pasca) and B.V.V. and A.F.; methodology, M.P.Z., L.P. (Ludovica Pasca), B.V.V., V.D.G.; formal analysis, S.G. and L.P. (Livio Provenzi); writing—original draft preparation, M.P.Z. and L.P. (Ludovica Pasca); writing—review and editing, L.P. (Ludovica Pasca) C.V., V.D.G. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Italian Ministry of Health RC 2020–2021.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by our Ethical Committee (P-20190033749), IRCCS Mondino Foundation, Pavia.

Informed Consent Statement: Informed consent was obtained from the parent of each subject and all children agreed to participate in the study.

Acknowledgments: Our thanks to the Italian association GLUT1DS ONLUS and the families. This work was supported by the Italian Ministry of Health RC 2017–2019.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Daci, A.; Bozalija, A.; Jashari, F.; Krasniqi, S. Individualizing Treatment Approaches for Epileptic Patients with Glucose Transporter Type1 (GLUT-1) Deficiency. *Int. J. Mol. Sci.* **2018**, *19*, 122. [CrossRef]
2. Klepper, J.; Akman, C.; Armeno, M.; Auvin, S.; Cervenka, M.; Cross, H.J.; De Giorgis, V.; Della Marina, A.; Engelstad, K.; Heussinger, N.; et al. Glut1 Deficiency Syndrome (Glut1DS): State of the art in 2020 and recommendations of the international Glut1DS study group. *Epilepsia Open* **2020**. [CrossRef] [PubMed]
3. Alter, A.S.; Engelstad, K.; Hinton, V.J.; Montes, J.; Pearson, T.S.; Akman, C.I.; De Vivo, D.C. Long-term clinical course of Glut1 deficiency syndrome. *J. Child. Neurol.* **2015**, *30*, 160–169. [CrossRef] [PubMed]
4. De Giorgis, V.; Masnada, S.; Varesio, C.; Chiappedi, M.A.; Zanaboni, M.; Pasca, L.; Filippini, M.; Macasaet, J.A.; Valente, M.; Ferraris, C.; et al. Overall cognitive profiles in patients with GLUT1 Deficiency Syndrome. *Brain Behav.* **2019**, *9*, 01224. [CrossRef] [PubMed]
5. Ito, Y.; Takahashi, S.; Kagitani-Shimono, K.; Natsume, J.; Yanagihara, K.; Fujii, T.; Oguni, H. Nationwide survey of glucose transporter-1 deficiency syndrome (GLUT-1DS) in Japan. *Brain Dev.* **2015**, *37*, 780–789. [CrossRef] [PubMed]
6. Hully, M.; Vuillaumier-Barrot, S.; Le Bizec, C.; Boddaert, N.; Kaminska, A.; Lascelles, K.; de Lonlay, P.; Cances, C.; Portes, V.D.; Roubertie, A.; et al. From splitting GLUT1 deficiency syndromes to overlapping phenotypes. *Eur. J. Med. Genet.* **2015**, *58*, 443–454. [CrossRef] [PubMed]
7. Larsen, J.; Johannesen, K.M.; Ek, J.; Tang, S.; Marini, C.; Blichfeldt, S.; Kibaek, M.; Von Spiczak, S.; Weckhuysen, S.; Frangu, M.; et al. The role of SLC2A1 mutations in myoclonic astatic epilepsy and absence epilepsy, and the estimated frequency of GLUT1 deficiency syndrome. *Epilepsia* **2015**, *56*, 203–208. [CrossRef] [PubMed]
8. Leen, W.G.; Klepper, J.; Verbeek, M.M.; Leferink, M.; Hofste, T.; Willemsen, M.A. Glucose transporter-1 deficiency syndrome: The expanding clinical and genetic spectrum of a treatable disorder. *Brain* **2010**, *133*, 655–670. [CrossRef] [PubMed]
9. Pearson, T.S.; Akman, C.; Hinton, V.J.; Engelstad, K.; De Vivo, D.C. Phenotypic spectrum of glucose transporter type 1 deficiency syndrome (Glut1 DS). *Curr. Neurol. Neurosci. Rep.* **2013**, *13*, 342. [CrossRef] [PubMed]
10. Ramm-Pettersen, A.; Stabell, K.E.; Nakken, K.O.; Selmer, K.K. Does ketogenic diet improve cognitive function in patients with GLUT1-DS? A 6- to 17-month follow-up study. *Epilepsy Behav.* **2014**, *39*, 111–115. [CrossRef] [PubMed]
11. Chugani, H.T.; Phelps, M.E. Maturational changes in cerebral function in infants determined by 18FDG positron emission tomography. *Science* **1986**, *231*, 840–843. [CrossRef] [PubMed]
12. Karmiloff-Smith, A. Development itself is the key to understanding developmental disorders. *Trends Cogn. Sci.* **1998**, *2*, 389–398. [CrossRef]
13. Bearzotti, F.; Fabbro, F. Test per la valutazione delle Prassie orofacciali nel bambino, Protocollo B e C (in Italian). *G. di Neuropsichiatria dell'Età Evol.* **2003**, *23*, 406–417.
14. Fanzago, F. *Test di Valutazione Dell'articolazione (in Italian)*; Centro Stampa Palazzo Maldura: Padova, Italy, 1986.
15. Cossu, G. *TNL—Test Neuropsicologico Lessicale Per l'età Evolutiva (in Italian)*; Hogrefe: Firenze, Italy, 2013.
16. Chilosi, A.M.; Lorenzini, I.; Fiori, S.; Graziosi, V.; Rossi, G.; Pasquariello, R.; Cipriani, P.; Cion, G. Behavioral and neurobiological correlates of childhood apraxia of speech in Italian children. *Brain Lang.* **2015**, *150*, 177–185. [CrossRef] [PubMed]
17. Dunn, L.M.; Dunn, L.M. *Peabody Picture Vocabulary Test—Revised. Minneapolis: NCS Pearson; 2007 Standardised for Italian Children*; Stella, G., Pizzoli, P., Tressoldi, P.E., Eds.; Omega: Torino, Italy, 2000.

18. Bisiacchi, P.S.; Cendron, M.; Gugliotta, M.; Tressoldi, P.E.; Vio, C. *BVN 5-11 Batteria di Valutazione Neuropsicologica per l'età Evolutiva (in Italian)*; Erickson: Trento, Italy, 2005.
19. Korkman, M.; Kirk, U.; Kemp, S. *NEPSY-II: Clinical and interpretative manual. San Antonio TX: Harcourt Assessment; 2007. Standardised for Italian Children*; Urgesi, C., Campanella, F., Fabbro, F., Eds.; Giunti OS: Firenze, Italy, 2011.
20. Wechsler, D. *WISC-IV: Wechsler Intelligence Scale for Children—Fourth Edition*; Psychological Corporation: San Antonio, TX, USA, 2003.
21. Wechsler, D. *WPSSI III: Wechsler Preschool and Primary Scale of Intelligence*; Psychological Corporation: New York, NY, USA, 1967.
22. Klepper, J.; Scheffer, H.; Leiendecker, B.; Gertsen, E.; Binder, S.; Leferink, M.; Hertzberg, C.; Näke, A.; Voit, T.; Willemsen, M.A. Seizure control and acceptance of the ketogenic diet in GLUT1 deficiency syndrome: A 2- to 5-year follow-up of 15 children enrolled prospectively. *Neuropediatrics* **2005**, *36*, 302–308. [CrossRef] [PubMed]
23. Wang, D.; Pascual, J.M.; Yang, H.; Engelstad, K.; Jhung, S.; Sun, R.P.; De Vivo, D.C. Glut-1 deficiency syndrome: Clinical, genetic, and therapeutic aspects. *Ann. Neurol.* **2005**, *57*, 111–118. [CrossRef] [PubMed]
24. Oller, D.K.; Eilers, R.E.; Neal, A.R.; Schwartz, H.K. Precursors to speech in infancy: The prediction of speech and language disorders. *J. Commun. Disord.* **1999**, *32*, 223–245. [CrossRef]
25. Vihman, M.M. Applied language studies. In *Phonological Development: The Origins of Language in the Child*; Blackwell Publishing: Oxford, UK, 1996.
26. Westermann, G.; Reck Miranda, E. A new model of sensorimotor coupling in the development of speech. *Brain Lang.* **2004**, *89*, 393–400. [CrossRef]
27. Shriberg, L.D.; Aram, D.M.; Kwiatkowski, J. Developmental apraxia of speech: III. A subtype marked by inappropriate stress. *J. Speech Lang. Hear. Res.* **1997**, *40*, 313–337. [CrossRef] [PubMed]

MDPI
St. Alban-Anlage 66
4052 Basel
Switzerland
Tel. +41 61 683 77 34
Fax +41 61 302 89 18
www.mdpi.com

Children Editorial Office
E-mail: children@mdpi.com
www.mdpi.com/journal/children

MDPI
St. Alban-Anlage 66
4052 Basel
Switzerland

Tel: +41 61 683 77 34

www.mdpi.com

ISBN 978-3-0365-6840-9