frontiers | Research Topics

# Parental questionnaires as a reliable instrument for the assessment of child language development

**Edited by**
Maria-José Ezeizabarrena and Melita Kovacevic

**Published in**
Frontiers in Psychology

ISSN 1664-8714
ISBN 978-2-8325-5426-5
DOI 10.3389/978-2-8325-5426-5

## About Frontiers

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

## Frontiers journal series

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

## Dedication to quality

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

## What are Frontiers Research Topics?

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

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

# Parental questionnaires as a reliable instrument for the assessment of child language development

**Topic editors**
Maria-José Ezeizabarrena — University of the Basque Country, Spain
Melita Kovacevic — University of Zagreb, Croatia

**Citation**
Ezeizabarrena, M.-J., Kovacevic, M., eds. (2024). *Parental questionnaires as a reliable instrument for the assessment of child language development*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-5426-5

# Table of contents

frontiers | Frontiers in Psychology

TYPE Editorial
PUBLISHED 02 September 2024
DOI 10.3389/fpsyg.2024.1471747

Check for updates

OPEN ACCESS

EDITED AND REVIEWED BY
Pamela Bryden,
Wilfrid Laurier University, Canada

*CORRESPONDENCE
Maria-José Ezeizabarrena
✉ mj.ezeizabarrena@ehu.eus

RECEIVED 28 July 2024
ACCEPTED 01 August 2024
PUBLISHED 02 September 2024

CITATION
Ezeizabarrena M-J and Kovacevic M (2024) Editorial: Parental questionnaires as a reliable instrument for the assessment of child language development.
*Front. Psychol.* 15:1471747.
doi: 10.3389/fpsyg.2024.1471747



# Editorial: Parental questionnaires as a reliable instrument for the assessment of child language development

Maria-José Ezeizabarrena[1,2]* and Melita Kovacevic[3]

[1]Departement of Linguistics and Basque Studies, University of the Basque Country (UPV/EHU), Vitoria-Gasteiz, Spain, [2]ELEBILAB, Bilingualism Research Laboratory (UPV/EHU), Vitoria-Gasteiz, Spain, [3]Laboratory for Psycholinguistic Research, Department for Speech and Language Pathology, ERF, University of Zagreb, Zagreb, Croatia

KEYWORDS

**parental questionnaire, child language, development, assessment, vocabulary, grammar, pragmatics**

Editorial on the Research Topic
Parental questionnaires as a reliable instrument for the assessment of child language development

Parents and caregivers have the experience of observing their children many hours per week, including school days, weekends and holidays, and in multiple situations of tiredness and lack of attention, but also of excitement, happiness and creativity. Such a varied experience makes them the ideal informants about children's behavior. Therefore, professionals such as pediatricians, psychologists, speech and language therapists, and teachers take great note of the information provided by the parents when assessing their children's health, general and learning skills and communicative abilities.

One of the tools used widely in the assessment of language development and in psycholinguistic research to compile such valuable information is questionnaires. Except for the time needed to fill them out, parental questionnaires do not require a major effort. Informants do not need to disrupt their daily life by scheduling appointments with the professional either, since they can select the time and place in which they feel more comfortable reporting on their children's linguistic experience and verbal (and non-verbal) communicative skills. Moreover, data obtained using this methodology are less likely to be influenced by factors that may mask a child's "true" abilities in the laboratory or clinic, such as child's non-compliance, shyness or time limitations.

This volume presents studies conducted with different types of parental questionnaires in either their original version or in their adaptations to other languages. It comprises fifteen articles—12 original research, one brief research, one mini-review and one opinion—on young children's language development. Twelve of them are based on *The MacArthur-Bates Communicative Development Inventories* (CDI) (Fenson et al., 2007). The remaining three used, respectively, the *Language Use Inventory* (LIU) by O'Neill (2009), the *Parents of Bilingual Children Questionnaire* (PaBiQ) by Tuller (2015), or the *Parental Linguistic Concern Questionnaire* (PLCQ), based on Restrepo (1998).

Five CDI questionnaires are available in many languages for the assessment of infants' and toddlers' communicative skills at different ages. All these five instruments were originally developped to norm the non-verbal and/or verbal communicative skills in English of (mostly monolingual) infants and toddlers living in the USA. The long

*Words and Gestures* (CDI-1) was designed to test gestures, receptive and expressive vocabulary of 8–15 months (originally), whilst the age range has been extended to 18 month-olds or even older in some of the adaptations. Its shorter version, CDI-1sh, tests vocabulary only. The *Words and Phrases* (CDI-2) questionnaire was designed to test expressive vocabulary and grammar of 16- to 24 month-olds, although its short version (CDI-2sh) only tests vocabulary. Finally, the CDI-3, of which there is only a short version, was originally designed to test vocabulary, grammar and language use of children up to age 4 (Fenson et al., 2007). In their opinion article, Marchman and Dale present a comprehensive overview of the contribution of the samplings conducted in the late 20th century with the original USA-English CDI-2 printed versions, and compare those findings with more recent ones, obtained in the current century from new child populations and using online procedures. Despite slight differences demonstrated, rates of vocabulary size and increase between age 16 and 30 months appear as very consistent across samples, confirming the robustness of the data and the reliability of the instrument.

CDI instruments have been adapted to over 100 languages in the world. The number of 12 (mostly European) countries and language varieties involved in this Research Topic is a clear evidence of the international impact of CDIs and their adaptations. Data from Finno-Ugric languages such as Estonian (Tulviste and Schults) and Finnish (Surakka et al.) are presented in addition to Germanic languages, such as British English (Jago et al.), Norwegian (Holm et al.), and Swedish (Eriksson and Myrberg), to Romance languages, such as Catalan (Feijoo et al.), Chilean Spanish (Varela-Moraga et al.), and Galician (Ogneva and Pérez-Pereira), to Semitic languages, such as Hebrew (Ohana and Armon-Lotem) and Maltese (Gatt et al.) and to Slavic languages, such as Croatian (Šmit Brlekovič and Kuvač-Kraljevič).

Some of the CDI articles deal with typically developing, and almost exclusively monolingual, children (Holm et al.; Jago et al.; Šmit Brlekovič and Kuvač-Kraljevič; Marchman and Dale; Surakka et al.; Tulviste and Schults; Varela-Moraga et al.). Others report on children with or at risk of developmental delay (Eriksson and Myrberg; Jago et al.; Šmit Brlekovič and Kuvač-Kraljevič; Ogneva and Pérez-Pereira; Varela-Moraga et al.). A set of papers report on and explore the ways of (better) assessing the linguistic development of children with bi- or multilingual language exposure and use (Eriksson and Myrberg; Ohana and Armon-Lotem) or compare the acquisition of the same language in normal and exceptional pandemic circumstances (Feijoo et al.). Variation was found across CDI studies in participants' profiles, but also in the specific questionnaire used in their assessment. Some used the CDI-1 (Feijoo et al.; Jago et al.; Surakka et al.; Varela-Moraga et al.), alone or together with the CDI-2 (Feijoo et al.; Gatt et al.; Marchman and Dale; Ogneva and Pérez-Pereira; Ohana and Armon-Lotem; Surakka et al.; Varela-Moraga et al.), whilst others used the CDI-3 (Eriksson and Myrberg; Holm et al.; Šmit Brlekovič and Kuvač-Kraljevič; Tulviste and Schults). The majority of papers converge in testing and demonstrating the internal consistency and validity of the instruments across languages. Some provide additional evidence of their validity to predict outcomes even over 2 years later.

Vocabulary and grammar are not the only linguistic components assessed through parental questionnaires. The mini-review by Pesco and O'Neill presents the *Language Use Inventory* (LIU), an instrument designed to measure children's pragmatic knowledge, originally in English, and an overview of its adaptation to seven additional languages, namely Arabic, French, Italian, Mandarin, Norwegian, Polish, and Portuguese. Based on the instrument's sensitivity to age and its usefulness across different linguistic and cultural contexts, Pesco and O'Neil conclude that LIU is valuable for clinical and research purposes.

Auza et al.'s paper analyses the strengths and weaknesses of the *Parental Language Concern Questionnaire* (PLCQ) in the identification of monolingual Mexican Spanish-speaking children with delay in language development. They conclude that a reduced questionnaire conformed by four out of the eight items in the list, in combination with one of the four items of the additional list of *Biological and Environmental Conditions Questions*, based on Peñaloza (2018) is a reliable screening method for identifying children with language disorders.

The usefulness of parental questionnaires extends to the assessment of older than pre-school aged children, as demonstrated by Pourquié et al. in their investigation, in which data obtained with the parental questionnaire HEGA (Haur Elebidunen Gurasoentzako Galdetegia 'Questionnaire for parents of bilingual children'), the Basque adaptation of *Parents of Bilingual Children Questionnaire* (PABIQ) were tested against performance data in Basque of 4- to 9-year-old children. They found a correlation between the parental responses to questions on their children's linguistic experience and children's accuracy at several scales of expressive vocabulary and grammar in Basque.

The studies compiled in this volume confirm: (a) the interest of the community of researchers and professionals of language therapy for the development and use of parental questionnaires to assess language development; (b) the consistency of the data, inter-individually, intra-individually and across languages; and (c) the reliability, validity, and usefulness of these tools for identifying atypical development in children's early and later communicative skills.

## Author contributions

M-JE: Writing – original draft, Writing – review & editing. MK: Writing – review & editing.

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the Basque Governement (IT-1627-22).

## Conflict of interest

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

## Publisher's note

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

## References

Fenson, L., Marchman, V., Thal, D., Dale, P., Reznick, S., and Bates, E. (2007). *MacArthur-Bates communicative development inventories – Users guide and technical manual*. Baltimore, MD: Paul H. Brookes Publishing Co.

O'Neill, D. (2009). *Language Use Inventory (LUI): Manual*. Waterloo: Knowledge in Development, Inc.

Peñaloza, C. (2018). *Habilidades lingüístico-discursivas y características familiares y socioambientales de niños preescolares de Ciudad de México* [PhD]. Mexico: Universidad Nacional Autónoma de México.

Restrepo, M. (1998). Identifiers of predominantly Spanish-speaking children with language impairment. *J. Speech Lang. Hear. Res.* 41, 1398–1411. doi: 10.1044/jslhr.4106.1398

Tuller, L. (2015). "Clinical use of parental questionnaires in multilingual contexts," in *Methods for assessing multilingual children: disentangling bilingualism from Language Impairment,* eds. S. Armon-Lotem, J. de Jong, and N. Meir (Bristol: Multilingual Matters).

frontiers | Frontiers in Psychology

TYPE Opinion
PUBLISHED 31 May 2023
DOI 10.3389/fpsyg.2023.1170303



OPEN ACCESS

EDITED BY
Maria-José Ezeizabarrena,
University of the Basque Country UPV/EHU, Spain

REVIEWED BY
Ewa Haman,
University of Warsaw, Poland
Hanne Gram Simonsen,
University of Oslo, Norway

*CORRESPONDENCE
Virginia A. Marchman
✉ marchman@stanford.edu

RECEIVED 20 February 2023
ACCEPTED 11 May 2023
PUBLISHED 31 May 2023

CITATION
Marchman VA and Dale PS (2023) The MacArthur-Bates Communicative Development Inventories: updates from the CDI Advisory Board. *Front. Psychol.* 14:1170303.
doi: 10.3389/fpsyg.2023.1170303



# The MacArthur-Bates Communicative Development Inventories: updates from the CDI Advisory Board

Virginia A. Marchman[1]* and Philip S. Dale[2]

[1]Department of Psychology, Stanford University, Stanford, CA, United States, [2]Department of Speech and Hearing Sciences, University of New Mexico, Albuquerque, NM, United States

KEYWORDS

**MacArthur-Bates CDI, parent report, language development, Wordbank, Web-CDI, CDI-CAT**

## 1. Introduction

Beginning with the classic diary studies (e.g., Stern and Stern, 1907; Leopold, 1949), parents (and other caregivers)[1] have been a source of valuable insights on their child's early language and communicative skills. The logic behind parent report is simple. Parents are generally keenly aware of their child's behaviors and their impressions are based on hours of observation in diverse settings, rather than the brief time available in a standard clinic or office visit. Moreover, their reports are less likely to be influenced by factors that may mask a child's "true" abilities in the laboratory or clinic (e.g., child non-compliance).

The MacArthur-Bates Communicative Development Inventories (CDIs) evolved from early efforts to harness this parental expertise in the 1970 and 1980's by Elizabeth Bates and colleagues (Bates et al., 1975, 1988). These instruments were further developed in the mid-1990's and beyond (Fenson et al., 1993, 1994, 2007), first for families with children learning American English, followed shortly by Mexican Spanish (Jackson-Maldonado et al., 1993, 2003) and Italian (Caselli and Casadio, 1995). For Bates, the main keys to success for parent report were to ask parents to choose from a list of example words or behaviors (e.g., recognition), rather than to recall them from memory, and to focus on emerging, salient behaviors, rather than to require retrospective reports.

Most versions of the CDIs come in two levels/forms. The Words and Gestures (CDI:WG) form, for children 8–18 months, asks parents to indicate on a vocabulary checklist which words or phrases their child "understands" or "understands and says," and to choose among examples of early communicative and symbolic gestures. The Words and Sentences (CDI:WS) form, for children 16–30 months, asks parents to select the words their child produces on their own, and also to indicate their child's use of grammatical forms (e.g., plural -s) and word combinations (e.g., "mommy sock"). While these original "long form" versions provide a comprehensive picture of early language, they typically require 20–30 min or longer for the parent to complete. Consequently, short form versions focusing only on vocabulary have been developed for when a comprehensive assessment is not needed or parental time commitment is limited (Fenson et al., 2000; Jackson-Maldonado et al., 2013). Versions appropriate for slightly older children, the CDI-III, have also been developed (e.g., Dale et al., 2023; Jackson-Maldonado et al., 2023).

1 We use the term "parent" to refer to any individual who takes care of the child on a regular basis, which could include parents, grandparents, step-parents, and others.

The CDIs have been used to explore many questions relevant to researchers and clinicians, for example, to what extent do demographic or environmental factors influence language development? Does a low score on the CDI predict continued or future language delays? Most analyses rely on aggregate scores from the vocabulary checklist, e.g., total words understood or total words produced. But individual item responses can also be analyzed, investigating questions such as the relative difficulty of words, or whether some words are more likely to be learned by boys vs. girls (Braginsky et al., 2019).

In the late 1980's, the developers of the American English and Mexican Spanish CDIs came together to form an Advisory Board. For more than 25 years, the Board has used proceeds from the sales of these instruments, distributed by Brookes Publishing Co (https://brookespublishing.com/product/cdi/), to support a variety of initiatives in the U.S. and internationally. Thanks to the strong interest and considerable effort of researchers around the globe, the Board has authorized versions of CDIs in more than 100 languages, with each instrument adapted (not just translated) to fit the linguistic and sociocultural features of that language. These important contributions are too numerous to mention individually, so we invite readers to peruse the full list here: https://mb-cdi.stanford.edu/adaptations.html. It is gratifying to reflect on the role of the CDIs in the crosslinguistic child language landscape, making contributions to our understanding of the normative course, as well as the individual differences, that characterize early language development (e.g., Bornstein et al., 2004; Bleses et al., 2008; Tardif et al., 2008; Jørgensen et al., 2010; Frank et al., 2021). At the same time, we acknowledge that significant gaps remain in the availability of CDIs across the world's languages (Kidd and Garcia, 2022).

In this Opinion, we seek to remember Liz Bates and the contributions that she made by briefly reviewing four recent significant innovations directed by the MacArthur-Bates CDI Advisory Board. First, we overview Wordbank, an open repository of CDI administrations from dozens of languages (Frank et al., 2017, 2021). Second, we report on an online platform for administration and scoring called Web-CDI (deMayo et al., 2021). Third, we discuss the development of a new, computer-adaptive testing instrument, the CDI-CAT (Kachergis et al., 2022). Finally, we announce the expanded and improved normative data for the American English long forms included in the 3rd Edition of the User's Guide and Technical Manual (Marchman et al., 2023).

# 2. Four major initiatives

## 2.1. Harnessing the power of "big data"

Many samples of CDI data to date have been limited in both size and scope because, with few exceptions, they were collected at a single site or laboratory. When data are combined across laboratories, the resulting datasets are not only larger, but are also likely to be considerably more representative along key dimensions (e.g., socioeconomic status). Building the infrastructure to enable data sharing across laboratories is non-trivial, but an enterprise that has a history in our field, for example, ChiLDES and CLEX (MacWhinney, 2000; Jørgensen et al., 2010). Inspired by this work, Michael Frank and his team developed Wordbank (http://wordbank.stanford.edu, Frank et al., 2017), a structured database of cross-linguistic CDI data currently consisting of more than 80,000 CDI administrations in 38 different languages. Such amazing progress would not have been possible without the many researchers who contributed their data (http://wordbank.stanford.edu/contributors). Wordbank also provides a powerful, browseable web interface that allows interactive exploration at the level of individual children (aggregating across words) and of individual words (aggregating across children). Recent analyses reveal remarkable insights into both the consistency and variability in early language development across languages (Braginsky et al., 2019; Frank et al., 2021). In just a few short years, Wordbank has become an invaluable tool with many research and teaching applications.

## 2.2. Moving beyond paper-and-pencil

Traditionally, CDIs are completed on paper: parents check off responses using pencil/pen and scores are later hand-tabulated. Today many prefer to engage with an electronic or online format on a laptop, tablet, or smart phone. Electronic administration eliminates postage costs, does not require face-to-face contact, and minimizes the chance of lost forms. Moreover, scoring is simplified since responses need not be transferred from the paper into an electronic format. Two options for electronic administration are available through Brookes Publishing Co. First, users can purchase fillable pdfs of the American English and Mexican Spanish forms, which can be emailed to families. Tabulation of scores is straightforward, but requires additional tools, such as Excel. A second option is Web-CDI, which offers a comprehensive online administration, data management, and scoring platform. Similar to platforms in other languages (Kristoffersen et al., 2013; Gendler-Shalev and Dromi, 2022), users share URLs (web links to a researcher's or clinician's own Web-CDI study) via email or social media, facilitating access to families at a distance. Importantly, Web-CDI's infrastructure ensures anonymity and participant privacy. Moreover, pictorial instructions are provided to facilitate uptake of critical information (see http://mb-cdi.stanford.edu/about). End-users can download tabulated summary scores, percentiles, and individual item responses automatically, facilitating analyses at both the child and item levels.

A recent analysis showed that demographic trends are similar for the American English long forms collected with Web-CDI and paper (deMayo et al., 2021). Moreover, Web-CDI has been successfully used to recruit American English-speaking families from diverse backgrounds, offering hope that the platform may help increase representation across ethnic/racial and educational groups. Managed in parallel at Stanford University and the Max Planck Institute, CDIs for American English, Mexican Spanish, Canadian French, Korean, Hebrew, Dutch, and Argentinian Spanish are currently available in Web-CDI, and there is a straightforward procedure for adding more languages.

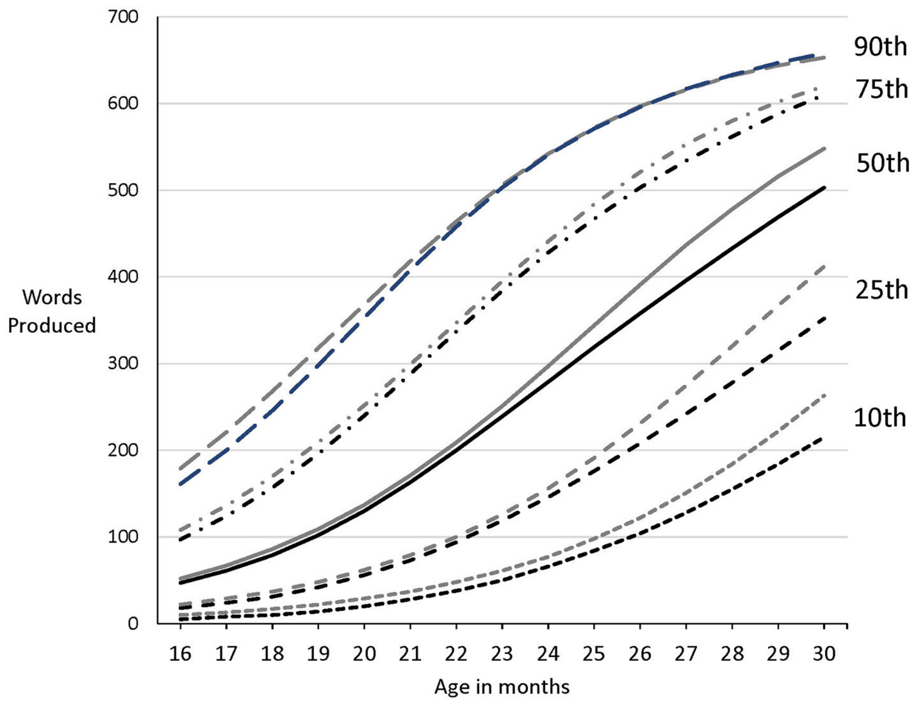


FIGURE 1
Percentile values (10, 25, 50, 75, and 90th) for words produced on the American English MacArthur-Bates Communicative Development Inventory: Words and Sentences form from Fenson et al. (2007; gray lines) and Marchman et al. (2023; dark lines).

## 2.3. Introducing the CDI-CAT

The vocabulary checklists of the CDI forms typically include hundreds of words, yielding a comprehensive view of children's vocabulary across many different lexical categories. However, asking about many words that a child is unlikely to know is inefficient and provides little information about the specific child relative to their peers. Computerized adaptive testing (CAT; van der Linden and Glas, 2009) offers an alternative approach. Each parent responds to an individualized list of words, each one selected based on their responses to the previous items. Scoring is computed using a statistical model based on Item Response Theory (IRT). Kachergis et al. (2022) reports on the development and testing of CDI-CATs for both comprehension and production vocabulary in both American English and Mexican Spanish (see also Kachergis and Dale, 2023). Like CATs in other languages (Mayor and Mani, 2019; Mieszkowska et al., 2022), even very short American English CDI-CATs (20–25 items) recovered participant abilities very well with little bias across ages. Moreover, a validation study with more than 200 children whose parents completed both the American English CDI-CAT and the American English CDI:WS showed a very strong correlation ($r = 0.92$). CDI-CATs for vocabulary production in American English and Mexican Spanish are available the spring of 2023 within Web-CDI. CDI-CATs are currently being developed in other languages (e.g., French) with others forthcoming.

## 2.4. New and improved normative information for the American English forms

Percentile scores for the major CDI measures place individual children in relation to a large norming sample. Unfortunately, the norming data for the American English long forms (Fenson et al., 2007) were not representative of the educational, racial, and ethnic distributions of the U.S. population. To remedy this situation, more than 4,000 additional CDI administrations have been added to the norming sample, yielding a sample of more than 6,500 children. Data were contributed by a consortium of researchers who used Web-CDI for their own independent research enterprises as well as via targeted online efforts, e.g., Facebook, to reach a broad, demographically diverse sample of caregivers. In addition, we statistically adjusted the data with raking, a technique for reweighting survey data (Lumley, 2020) in the R statistical package (R Core Team, 2020) to achieve a sample distribution that more closely resembled the demographic makeup of the target population. We used 2020 U.S. Census data[2] for race, ethnicity, and caregiver (maternal) education to derive the weights. Models were fit using generalized additive models in the Beta distribution family (GAMLSS, Stasinopoulos et al., 2017), a technique that is more sophisticated than that used in earlier versions. These innovations are important because a norming sample that is biased toward more educated and otherwise advantaged families results

2 https://www.census.gov, 2020.

in norms that are too high, and therefore, may over-classify late talkers. As illustrated in Figure 1, the new norms generally show lower scores, especially for children who fall <90th percentile and who are older than 24 months of age. It is hoped that these types of statistical solutions will be informative to others who are interested in increasing the representativeness of their norming samples. These new normative data are available in Marchman et al. (2023).

## 3. What's next?

There is still much to be done. One important area is facilitating the administrative, analytic, and reporting practices that best serve children learning more than one language. Joining ongoing discussions (e.g., Gatt et al., 2015) and following from key research in this area (e.g., O'Toole et al., 2017), a new chapter on this topic is included in the new manual (Weisleder et al., 2023), which makes recommendations for best practices in assessment and reporting. We also look forward to efforts that stretch the limits of the parent report methodology to more effectively accommodate respondents with low-literacy or limited experience with electronic devices (e.g., Alcock et al., 2015). We also commend ongoing efforts to apply the parent report methodology to older children, as well as beyond the home context (e.g., Morford and Carlson, 2011; Eriksson, 2017; Bleses et al., 2018; Tulviste and Schults, 2020; Kas et al., 2022). The MacArthur-Bates CDI Advisory Board is committed to continuing to strengthen our knowledge and efficacy in these and other domains and welcomes the efforts of scholars around the world in further expanding the availability of CDIs worldwide.

## 4. Conclusion

In 2023, it will be 20 years since the untimely passing of Elizabeth Bates. In this Opinion, we have sought to honor Liz's memory by highlighting a few recent developments in parent report methodology spearheaded by the MacArthur-Bates CDI Advisory Board. We hope that readers of this special issue will appreciate hearing about our continuing efforts to build on her legacy by strengthening cross-laboratory and cross-linguistic collaboration, improving data administration and management techniques, and expanding the representativeness of normative data. We know that these initiatives represent only a few of the CDI-related activities that are ongoing in the child language community and acknowledge that there is still much more for all of us to do. We look forward to many more years of collaborations with the international community to improve and expand parent report as a useful tool for the fields of child language and developmental psycholinguistics.

## Author contributions

VM and PD contributed equally to the drafting of this manuscript. All authors contributed to manuscript revision, and have read and approved the submitted version.

## Funding

The preparation of this article was supported by the MacArthur-Bates CDI Advisory Board.

## Acknowledgments

We would like to thank the members of the MacArthur-Bates CDI Advisory Board (Larry Fenson, Michael C. Frank, Margaret Friend, Judy Reilly, Caroline Rowland, Donna J. Thal, Ann M. Weber, and Adriana Weisleder), as well as our colleagues who have contributed in so many ways to the advancements that we report on here, including Mika Braginsky, Benjamin deMayo, Danielle Kellier, George Kachergis, Jessica Mankewitz, Alvin Tan, Daniel Yurovsky, and Henry Mehta.

## Conflict of interest

The authors of this paper are members of the MacArthur-Bates CDI Advisory Board, a non-profit consortium of researchers. Funds received from the sale of the American English and Mexican Spanish CDI instruments are placed in a non-profit account and used to fund research, assessment, and applications of CDI instruments. None of the Board members directly receive compensation from the sale of CDI instruments.

## Editor's note

Maria-José Ezeizabarrena edited the article in collaboration with Melita Kovacevic, University of Zagreb, Zagreb, Croatia.

## Publisher's note



## References

Alcock, K. J., Rimba, K., Holding, P., Kitsao-Wekulo, A., Abubakar, A., and Newton, C. R. J. C. (2015). Developmental inventories using illiterate parents as informants : Communicative Development Inventory (CDI) adaptation for two Kenyan languages. *J. Child Lang.* 42, 763–785. doi: 10.1017/S0305000914000403

Bates, E., Bretherton, I., and Snyder, L. (1988). *From First Words to Grammar: Individual Differences and Dissociable Mechanisms*. Cambridge: Cambridge University Press.

Bates, E., Camaioni, L., and Volterra, V. (1975). The acquisition of performatives prior to speech. *Merrill-Palmer Quart. Behav. Dev*. 21, 205–226.

Bleses, D., Jensen, P., Højen, A., and Dale, P. S. (2018). An educator-administered measure of language development in young children. *Infant Behav. Dev*. 52, 104–113. doi: 10.1016/j.infbeh.2018.06.002

Bleses, D., Vach, W., Slott, M., Wehberg, S., Thomsen, P., Madsen, T. O., et al. (2008). Early vocabulary development in Danish and other languages: A CDI-based comparison. *J. Child Lang*. 35, 619–650. doi: 10.1017/S0305000908008714

Bornstein, M. H., Cote, L. R., Maital, S., Painter, K., Park, S. Y., Pascual, L., et al. (2004). Cross-linguistic analysis of vocabulary in young children: Spanish, Dutch, French, Hebrew, Italian, Korean, and American English. *Child Dev*. 75, 1115–1139. doi: 10.1111/j.1467-8624.2004.00729.x

Braginsky, M., Yurovsky, D., Marchman, V. A., and Frank, M. C. (2019). Consistency and variability in children's word learning across languages. *Open Mind* 3, 52–67. doi: 10.1162/opmi_a_00026

Caselli, M. C., and Casadio, P. (1995). *Il primo vocabolario del bambino: Guida all'uso del questionario MacArthur per la valutazione della comunicazione e del linguaggio nei primi anni di vita*. Milan: Franco Angeli Editore.

Dale, P. S., Kachergis, G., Weber, A. M., and Frank, M. C. (2023). "The CDI-III," in *The MacArthur-Bates Communicative Development Inventories: User's Guide and Technical Manual, 3rd Edn*, eds V. A. Marchman, P. S. Dale, and L. Fenson (Baltimore, MD: Brookes Publishing Co), 10.

deMayo, B., Kellier, D., Braginsky, M., Bergmann, C., Hendriks, C., Rowland, C., et al. (2021). Web-CDI: A system for online administration of the MacArthur-Bates Communicative Development Inventories. *Lang. Dev. Res*. 1, 55–98. doi: 10.31234/osf.io/8mjx9

Eriksson, M. (2017). The Swedish communicative development inventory III. *Int. J. Behav. Dev*. 41, 647–654. doi: 10.1177/0165025416644078

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., et al. (1994). Variability in early communicative development. *Monogr. Soc. Res. Child Dev*. 5, 1–185. doi: 10.2307/1166093

Fenson, L., Dale, P. S., Reznick, J. S., Thal, D. J., Bates, E., and Hartung, J. (1993). *MacArthur Communicative Development Inventories: User's Guide and Technical Manual*. Baltimore, MD: Brookes Publishing Co.

Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, J. S., and Bates, E. (2007). *MacArthur-Bates Communicative Development Inventories: User's Guide and Technical Manual, 2nd Edn*. Baltimore, MD: Brookes Publishing Co. doi: 10.1037/t11538-000

Fenson, L., Pethick, S. J., Renda, C., Cox, J. L., Dale, P. S., and Reznick, J. S. (2000). Short-form versions of the MacArthur Communicative Development Inventories. *Appl. Psycholinguist*. 21, 95–115. doi: 10.1017/S0142716400001053

Frank, M. C., Braginsky, M., Yurovsky, D., and Marchman, V. A. (2017). Wordbank: An open repository for developmental vocabulary data. *J. Child Lang*. 44, 677–694. doi: 10.1017/S0305000916000209

Frank, M. C., Braginsky, M., Yurovsky, D., and Marchman, V. A. (2021). *Variability and Consistency in Early Language Learning: The Wordbank Project*. Cambridge, MA: MIT Press.

Gatt, D., O'Toole, C., and Haman, E. (2015). "Using parental report to assess early lexical production," in *Methods for Assessing Multilingual Children: Disentangling Bilingualism From Language Impairment (Issue JANUARY)*, eds S. Armon-Lotem, J. De Jong, and N. Meir (Bristol: Multilingual Matters). doi: 10.21832/9781783093137-009

Gendler-Shalev, H., and Dromi, E. (2022). The hebrew web communicative development inventory (MB-CDI): Lexical development growth curves. *J. Child Lang*. 49, 486–502. doi: 10.1017/S0305000921000179

Jackson-Maldonado, D., Marchman, V. A., and Fernald, L. C. H. (2013). Short-form versions of the Spanish MacArthur–Bates Communicative Development Inventories. *Appl. Psycholinguist*. 34, 837–868. doi: 10.1017/S0142716412000045

Jackson-Maldonado, D., Marchman, V. A., Rubio-Codina, M., Friend, M., Weisleder, A., and Dale, P. S. (2023). The MacArthur-Bates CDI-III for Spanish-Speaking Children. *PsyArXiv*. doi: 10.31234/osf.io/7cw4v

Jackson-Maldonado, D., Thal, D. J., Marchman, V. A., Bates, E., and Gutierrez-Clellen, V. (1993). Early lexical development in Spanish-speaking infants and toddlers. *J. Child Lang*. 20, 523–549. doi: 10.1017/S0305000900008461

Jackson-Maldonado, D., Thal, D. J., Marchman, V. A., Newton, T., Fenson, L., and Conboy, B. (2003). *MacArthur Inventarios del Desarrollo de Habilidades Communicativas: User's Guide and Technical Manual*. Baltimore, MD: Brookes Publishing Co.

Jørgensen, R. N., Dale, P. S., Bleses, D., and Fenson, L. (2010). CLEX: A cross-linguistic lexical norms database. *J. Child Lang*. 37, 419–428. doi: 10.1017/S0305000909009544

Kachergis, G., and Dale, P. S. (2023). "Online computer adaptive tests of children's vocabulary (CDI-CAT)," in *The MacArthur-Bates Communicative Development Inventories: User's Guide and Technical Manual, 3rd Edn*, eds V. A. Marchman, P. S. Dale, and L. Fenson (Baltimore, MD: Brookes Publishing Co.), 137–140.

Kachergis, G., Marchman, V. A., Dale, P. S., Mankewitz, J., and Frank, M. C. (2022). Online computerized adaptive tests of children's vocabulary development in English and Mexican Spanish. *J. Speech Lang. Hear. Res*. 65, 2288–2308. doi: 10.1044/2022_JSLHR-21-00372

Kas, B., Jakab, Z., and Lorik, J. (2022). Development and norming of the Hungarian CDI-III: A screening tool for language delay. *Int. J. Lang. Commun. Disord*. 57, 252–273. doi: 10.1111/1460-6984.12686

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

Kristoffersen, K. E., Simonsen, H. G., Bleses, D., Wehberg, S., Jørgensen, R. N., Eiesland, E. A., et al. (2013). The use of the Internet in collecting CDI data—An example from Norway. *J. Child Lang*. 40, 567–585. doi: 10.1017/S0305000912000153

Leopold, W. (1949). *Speech Development of a Bilingual Child: A Linguist's Record*. Evanston, IL: Northwestern University Press.

Lumley, T. (2020). *Survey: Analysis of Complex Survey Samples. R Package Version 4.0*.

MacWhinney, B. (2000). *The Childes Project*. New York, NY: Lawrence Elbaum Associates.

Marchman, V. A., Dale, P. S., and Fenson, L. (2023). *The MacArthur-Bates Communicative Development Inventories: User's Guide and Technical Manual (V. A. Marchman, P. S. Dale, and L. Fenson (Eds.); 3rd Edn*. Baltimore, MD: Brookes Publishing Co.

Mayor, J., and Mani, N. (2019). A short version of the MacArthur–Bates Communicative Development Inventories with high validity. *Behav. Res. Methods* 51, 2248–2255. doi: 10.3758/s13428-018-1146-0

Mieszkowska, K., Krajewski, G., Sobota, K., Dynak, A., Kolak, J., Krysztofiak, M., et al. (2022). Parental report via a mobile app in the context of early language trajectories: StarWords study protocol. *Int. J. Environ. Res. Public Health* 19, 3067. doi: 10.3390/ijerph19053067

Morford, J. P., and Carlson, M. L. (2011). Sign perception and recognition in non-native signers of ASL. *Lang. Learn. Dev*. 7, 149–168. doi: 10.1080/15475441.2011.543393

O'Toole, C., Gatt, D., Hickey, T. M., Miekisz, A., Haman, E., Armon-Lotem, S., et al. (2017). Parent report of early lexical production in bilingual children: A cross-linguistic CDI comparison. *Int. J. Bilingual Educ. Bilingual*. 20, 124–145. doi: 10.1080/13670050.2016.1179258

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Stasinopoulos, M. D., Rigby, R. A., Heller, G. A., Voudouris, V., and De Bastiani, F. (2017). *Flexible Smoothing Using GAMLSS in R, 1st Edn*. Boca Raton, FL: CRC Press.

Stern, C., and Stern, W. (1907). *Der Kindersprache*. Basel: Barth.

Tardif, T., Fletcher, P., Liang, W., Zhang, Z., Kaciroti, N., and Marchman, V. A. (2008). Baby's first 10 words. *Dev. Psychol*. 44, 929. doi: 10.1037/0012-1649.44.4.929

Tulviste, T., and Schults, A. (2020). Parental reports of communicative development at the age of 36 months: The Estonian CDI-III. *First Lang*. 40, 64–83. doi: 10.1177/0142723719887313

van der Linden, W., and Glas, C. A. W. (2009). "Elements of adaptive testing," in *Elements of Adaptive Testing*, eds W. van der Linder and C. A. W. Glas (Berlin: Springer), 8. doi: 10.1007/978-0-387-85461-8

Weisleder, A., Friend, M., and Marchman, V. A. (2023). "Use of the CDI with bilingual children," in *The MacArthur-Bates Communicative Development Inventories: User's Guide and Technical Manual, 3rd Edn*, eds V. A. Marchman, P. S. Dale, and L. Fenson (Baltimore, MD: Brookes Publishing Co), 101–109.

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 16 June 2023
DOI 10.3389/fpsyg.2023.1167810



OPEN ACCESS

EDITED BY
Maria-José Ezeizabarrena,
University of the Basque Country UPV/EHU, Spain

REVIEWED BY
Peter B. Marschik,
University Medical Center Göttingen, Germany
Draško Kašćelan,
University of Essex, United Kingdom

*CORRESPONDENCE
Lana S. Jago
✉ l.s.jago@ljmu.ac.uk
Caroline F. Rowland
✉ Caroline.Rowland@mpi.nl

†PRESENT ADDRESS
Lana S. Jago,
School of Psychology, Liverpool John Moores University, Liverpool, United Kingdom

RECEIVED 16 February 2023
ACCEPTED 19 May 2023
PUBLISHED 16 June 2023

CITATION
Jago LS, Alcock K, Meints K, Pine JM and Rowland CF (2023) Language outcomes from the UK-CDI Project: can risk factors, vocabulary skills and gesture scores in infancy predict later language disorders or concern for language development?
*Front. Psychol.* 14:1167810.
doi: 10.3389/fpsyg.2023.1167810



# Language outcomes from the UK-CDI Project: can risk factors, vocabulary skills and gesture scores in infancy predict later language disorders or concern for language development?

Lana S. Jago[1*†], Katie Alcock[2], Kerstin Meints[3], Julian M. Pine[1] and Caroline F. Rowland[4,5*]

[1]Department of Psychology, University of Liverpool, Liverpool, United Kingdom, [2]Department of Psychology, Lancaster University, Lancaster, United Kingdom, [3]School of Psychology, University of Lincoln, Lincoln, United Kingdom, [4]Language Development Department, Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands, [5]Donders Institute for Brain, Cognition and Behaviour, Radboud University, Radboud, Netherlands

At the group level, children exposed to certain health and demographic risk factors, and who have delayed language in early childhood are, more likely to have language problems later in childhood. However, it is unclear whether we can use these risk factors to predict whether an individual child is likely to develop problems with language (e.g., be diagnosed with a developmental language disorder). We tested this in a sample of 146 children who took part in the UK-CDI norming project. When the children were 15–18 months old, 1,210 British parents completed: (a) the UK-CDI (a detailed assessment of vocabulary and gesture use) and (b) the Family Questionnaire (questions about health and demographic risk factors). When the children were between 4 and 6 years, 146 of the same parents completed a short questionnaire that assessed (a) whether children had been diagnosed with a disability that was likely to affect language proficiency (e.g., developmental disability, language disorder, hearing impairment), but (b) also yielded a broader measure: whether the child's language had raised any concern, either by a parent or professional. Discriminant function analyses were used to assess whether we could use different combinations of 10 risk factors, together with early vocabulary and gesture scores, to identify children (a) who had developed a language-related disability by the age of 4–6 years (20 children, 13.70% of the sample) or (b) for whom concern about language had been expressed (49 children; 33.56%). The overall accuracy of the models, and the specificity scores were high, indicating that the measures correctly identified those children without a language-related disability and whose language was not of concern. However, sensitivity scores were low, indicating that the models could not identify those children who were diagnosed with a language-related disability or whose language was of concern. Several exploratory analyses were carried out to analyse these results further. Overall, the results suggest that it is difficult to use parent reports of early risk factors and language in the first 2 years of life to predict which children are likely to be diagnosed with a language-related disability. Possible reasons for this are discussed.

KEYWORDS
vocabulary, CDI, health risk factors, demographic risk factors, language development, language impairment, language disorder

## 1. Introduction

Young children with early delays in language acquisition are at an increased risk for developing persisting language impairments (Rice et al., 2008). As a result, policy-makers and lobby groups often recommend that the language development of at risk children is monitored through their first years, and that they receive targeted support aimed at improving their communication environments (All Party Parliamentary Group on Speech and Language Difficulties, 2013; Save the Children, 2015).

However, though it is widely accepted that children's early vocabulary and gesture use correlates with their later language ability at the group level, at least in children within the typical range (Westerlund et al., 2006; Henrichs et al., 2011; Rescorla, 2011), the role of early language in *predicting* later language-related disability in individuals with sufficient accuracy is less clear cut. Many children who experience an early delay in language acquisition are later diagnosed with a developmental language disorder, but many also catch up with their peers before they enter school (Grossheinrich et al., 2019). Therefore, there is not a straightforward, linear, relationship between early and later language. This makes it very difficult to identify predictors of, for example, later language disorder (Dale et al., 2003; Westerlund et al., 2006). In particular, Westerlund et al. (2006) reported that a measure of vocabulary size at 18 months was not sensitive enough to identify children with language impairment at 3 years. The authors conclude that, although early language skills were associated with later language skills, alone, they were not sensitive enough to identify children with language impairment.

Exposure to a number of adverse health and demographic risk factors is also associated with delays in language development (Paul, 2000; Campbell et al., 2003; Reilly et al., 2010). However, once again, while a number of risk factors predict later language outcomes on a group level (Reilly et al., 2007, 2010), it is not clear that they accurately predict later language proficiency on an individual level. This means that we are currently unable to identify which children are likely to develop problems in their language acquisition using only exposure to risk factors as a predictor (Roos and Ellis Weismer, 2008; Bishop et al., 2012; Duff et al., 2015).

It is, however, possible that the accuracy of our predictive models could be improved by combining detailed measures of early language abilities (vocabulary and gesture use) with health and demographic risk factors (e.g., combining productive vocabulary with measures of receptive vocabulary, gestures, risk factors and early concern for language impairment). Thus, the goal of the present paper was to determine if parental assessments of vocabulary and gesture use in infancy, together with health and demographic risk factors, could be used to predict whether a child was later (by the age of 4–6 years) diagnosed with a disability that was likely to affect spoken English language proficiency (e.g., developmental disability, language disorder, hearing impairment). Because we recognized that some 4-6-year-olds may not yet have received a diagnosis, we also tested the effect of our risk factors on later language using a broader, more inclusive, measure; whether, by the age of 4–6 years, the child's language had raised any concern, either by a parent or professional. By including both measures we can gain a fuller picture of the link between early risk factors and later susceptibility to a language-related disability. In the remainder of this introduction, we review the relevant research on risk factors, before outlining the current study.

### 1.1. Health risk factors

A number of health factors have been shown to be robustly associated with language in young children; in particular child sex/gender[1], prematurity, low birth weight, ear infections, family history of speech or language impairments, and associated (non-language) developmental disabilities (Huttenlocher et al., 1991; Marschik et al., 2007; Barre et al., 2011; Kenyhercz and Nagy, 2022). For example, Reilly et al. (2010) found that multiple health-related risk factors (child sex/gender, low birth weight and family history of speech and language difficulties) predicted variance in language skills at age 4 years. Similarly, Jansson-Verkasalo et al. (2004) reported that being born premature was associated with poorer performance on measures of comprehension at 2 years.

However, such research often yields weak effects (Harrison and McLeod, 2010; Dale and Hayiou-Thomas, 2013; Fisher, 2017; Jin et al., 2020). This means that though the risk factor is correlated with language, it cannot be used to classify children as having a language delay or disorder. For example, Harrison and McLeod (2010) found that while health risk factors correctly classified children who were not attending speech-language pathology services, they only correctly classified 2.6% of children who were attending these services, showing that health risk factors were not sensitive enough to identify children with language impairment.

### 1.2. Demographic risk factors

Two demographic factors that have been shown to robustly predict language development over time are maternal education and household income (Reilly et al., 2010; Fisher, 2017). It is likely that these factors impact language development together because they interact; rate of household income can be dependent on parental education and together they determine socioeconomic status (SES). Previous research

1 We use the term sex/gender to indicate that a child's development can be affected both by their biological sex and socially-constructed gender. Although children at the age studied in this paper are unlikely to have already chosen their preferred gender, their development is very likely to have been influenced by the gender that their parents/society has assigned to them, as well as their biological sex. Thus we use the term sex/gender to encompass both these effects (see Bölte et al., 2023).

has suggested that the role of maternal education on language development may act *via* the mother's own language skills; parents with higher educational attainment have more advanced linguistic skills and are therefore more likely to produce rich linguistic input from which children can learn (Rowe et al., 2009; Rowe, 2018; Vernon-Feagans et al., 2020) although there is also a genetic component making it hard to disentangle cause and effect (Kovas et al., 2005). In addition, there are other factors at play; parents suffering from living in chaotic and crowded conditions will have less opportunity to engage in long periods of linguistic interaction with their children (Evans et al., 1999; Rowe, 2018; Fan et al., 2021). However, again, an association is not enough for accurate classification. For example, Harrison and McLeod (2010) found family income did not predict whether or not children were attending speech-language pathology services, and also reported that the effect of maternal education on predicting later language is weak (Harrison and McLeod, 2010).

## 1.3. Early concern about language development

Parental concern for language development is an often neglected, but promising factor, since it has been shown to support the identification of developmental delays and language impairment (Glascoe, 1991; Sim et al., 2019; Wallisch et al., 2020). A delay in vocabulary development is one of the reasons that parents first seek support for their children's development (Rescorla, 2011; Solgi et al., 2022). When combined with clinical observations, parental evaluation of their children's language development increases the accuracy with which paediatricians can detect developmental complications (Glascoe et al., 1989; Glascoe and Dworkin, 1995). Therefore, it may be possible to use parental concern, in combination with other risk factors, to support the detection of delays in language development, before language impairment is identified by a clinician or therapist.

The research outlined above shows that although multiple studies have identified associations between early risk factors and language, they have not concluded that it is possible to use these factors to identify, in individual children, the likelihood of later language-related disabilities such as developmental language disorder [DLD, previously called specivic language impairment (SLI)]. A plausible explanation is that the size of the effect of these risk factors in language acquisition are quite weak overall. In support of this idea is the finding that where effects are found, effect sizes are often small (Stolt et al., 2009; Carroll and Breadmore, 2017), and that studies with larger sample sizes are often more successful at detecting relationships between risk factors and language acquisition (Kennedy et al., 2006; Winskel, 2006; Barre et al., 2011; Van Noort-Van Der Spek et al., 2012).

If the effects are weak (i.e., effect sizes are small), this might explain why some measures discriminate well between faster learners, or learners in general (i.e., correlate with language), but have very little predictive power when identifying which individual children will develop a language-related disability. For example, boys tend to be somewhat slower word learners than girls on average (Eriksson et al., 2012) but the differences are so small, and the overlap in standard deviation so large, that sex/gender has very little discriminatory power on the level of the individual. In particular, Reilly et al. (2010) found that nine risk factors were successful in predicting individual differences in language scores at 4 years, but only three of them (family history of speech or language problems, low maternal education and SES) allowed the authors to discriminate between children with and without expressive DLD.

If the effects are weak, then, although risk factors when examined in isolation may not be strong enough to predict language impairment, they might gain more substantial predictive power when combined. Using multiple risk factors, combined with measures of earlier language skills, might then increase the chances of successfully predicting later language impairment. However, very little research has investigated how earlier language skills and multiple risk factors might interact to increase the risk of developing persistent language impairments. If we can show that a combination of risk factors and language skills identified early can predict later language impairment, we can use these factors as a starting point for informing early intervention based on risk.

## 1.4. The current study

The current study had three aims. First, we investigated whether, using a combination of risk factors and early vocabulary and gesture scores at time 1 (15–18 months), we could predict which children would go onto develop a language-related disorder at time 2 (4–6 years). We identified children with language-related disorder at age 4–6 years in two different ways. In our first category (Identified Disability group), we included only those children who had been diagnosed with a disability that was likely to affect spoken English proficiency by the age of 4–6 years (e.g., developmental disability, language disorder or hearing impairment). However, since many children do not have a diagnosis by age 4–6 years, we also created a broader category that included both children with an identified disability and children for whom a parent or a professional had expressed 'concern' about their language development (Overall Concern group). This second category is based on the premise that some children with a disability are not yet diagnosed at 4–6 years in the UK, and that the concern of a parent/professional may be a good indicator of an, as yet, undiagnosed language-related disability (Glascoe et al., 1989; Glascoe and Dworkin, 1995).

Our second aim was to determine whether there was continuity in concern over time; whether overall concern for language development at 4–6 years was predicted by parental concern for their child's language development between 15 and 18 months. We anticipated that parents' concerns about their children's language development in late infancy would predict later overall concern for children's later language development.

Our third and final aim was to investigate if risk factors and early language skills together could predict whether or not children's language development caught up *after* concern for their language development had been registered. We anticipated that children with bigger vocabularies and higher gesture scores at time 1, as well as those who experienced fewer risk factors, would be more likely to catch up compared to children experiencing more risk factors and with lower vocabulary and gesture scores.

# 2. Methods

## 2.1. Participants

The cohort were a subset of the 1,210 parents who had taken part in the UK-CDI Project when their children were between 15 and

18 months. The UK-CDI Project collected parental report data from across the United Kingdom to establish norms for productive and receptive vocabulary, and gestures for children from 8 to 18 months. Parents completed a vocabulary and gesture questionnaire (a Communicative Development Inventory, or CDI; Alcock et al., 2020), and a family questionnaire containing questions about child health, familial risk of language and literacy disorders, and demographic characteristics. The original cohort were representative of the UK population as a whole in terms of a range of demographic factors (e.g., socio-economic status, sex/gender, region, nation, marital status etc).

We sent out follow-up questionnaires (see design section below) to all of the families from the original cohort who had agreed to be contacted for further studies, had provided contact details, and whose children were between 15 and 18 months when they completed the UK-CDI (n = 370; 78 of whom had productive vocabulary scores in the bottom 25th percentile). We received 147 (40%) responses. One family was excluded because their response was incomplete, so the final sample size was 146.

The mean age of the 146 children included in the follow-up project was 16 months 25 days (age range: 15 months, 3 days −18 months, 28 days) during the UK-CDI Project (time 1), and 5 years and 3 months (age range: 4 years, 3 months–6 years, 4 months) at follow-up (time 2). All children were monolingual English learners. Table 1 shows the number of children in the different risk factor categories at time 1 and Table 2 shows the number of children in each quartile for vocabulary and gesture use at time 1 (Note that there were differences in the maternal education and household income of the parents who did/did not take part in the follow-up study, but not in any of the other risk factors; see Supplementary materials at https://osf.io/gvz3x/).

## 2.2. Design

We used a parent-report questionnaire to follow up children who previously took part in the UK-CDI Project, and had agreed to be contacted for future studies. This study was granted ethical approval by The University of Liverpool's Research Ethics Subcommittee for Non-Invasive Procedures for the study Language Development in Late Talkers (Institute Review Board protocol number: RETH000764). Predictor variables were measures that were derived from the answers given at time 1. These were: receptive vocabulary quartile, productive vocabulary quartile and gesture quartile, child sex/gender, prematurity, low birth weight, ear infections at time 1, familial risk for speech or language impairment, ear infection lasting more than 6 months, developmental disability, maternal education, household income, hearing or communication concerns at time 1. Outcome measures were derived from answers at time 2: Identified Disability, Overall Concern, Catch up. (See below for information about how these measures were derived).

## 2.3. Sampling and data collection procedures

The first data collection phase (time 1) took place as part of the UK CDI Project from 2013 to 2015. Details of how time 1 data were collected can be found in Alcock et al., 2020. The data for the follow-up phase (time 2) was collected between 2017 and 2018. We used the database from the UK-CDI Project to follow-up families who had consented to be re-contacted for future research. Parents were given a £5 shopping voucher for completing the follow-up questionnaire.

## 2.4. Measures and procedure

### 2.4.1. The UK-CDI

Communicative Development Inventories (CDI) are parent report checklists of the words, gestures and sentences that young children understand and say. Parents complete these questionnaires by indicating if their child uses or understands the words, sentences and gestures listed. The UK-CDI Words and Gestures is standardized for the UK population for vocabulary and gesture scores in children aged 8–18 months and has good validity and reliability (see Alcock et al., 2020; total possible scores are 395 for vocabulary and 63 for gesture).

### 2.4.2. The family questionnaire

The family questionnaire asks a range of questions about a child's health and family background. This questionnaire was designed for the UK-CDI Project (for details of construction, see Alcock et al., 2020) and was used to collect information about demographic and health risk factors, including prematurity, birth weight, family history of language delay or dyslexia, and SES.

### 2.4.3. Follow-up questionnaire

The follow-up questionnaire was used to investigate the later language outcomes of the children who took part in the UK-CDI project. The key questions for this study were those that asked about parental concern for language development, details of those concern (if any), whether the Healthy Child Programme's 2 Year Review identified a delay in language development, whether the children had been diagnosed with a developmental disability or language disorder, and whether the children had a visual or hearing impairment. The Healthy Child Programme 2 Year Review (Department of Health, 2009) is part of the *Healthy Child Programme: Pregnancy and the first five years of life*, which is run in England and Wales. This review is designed to optimize child development by reviewing all children in England and Wales between 2 and 2 years, 6 months.

At time 2, parents who had provided an email address at time 1 were sent an email containing a link to complete the questionnaire online. For parents who only provided a home address, a paper copy of the questionnaire was sent out with a prepaid return envelope included. See the Supplementary materials at https://osf.io/gvz3x/ for a copy of this questionnaire.

## 2.5. Data coding

### 2.5.1. Risk factor scores at time 1

Ten questions probed the child's susceptibility to 10 risk factors. All information was collected at time 1 when children were 15–18 months old. Answers to risk factor questions were scored as 1 if the child had the risk factor, and 0 if they did not

1. Concern: answering yes to "have you or anyone else had any concerns about your child's hearing or communication?" = score of 1

TABLE 1 Number of children with and without each risk factor at time 1.

| Risk factor | Female | | | Male | | |
|---|---|---|---|---|---|---|
| | *n* with the risk factor | *n* without the risk factor | Missing data | *n* with the risk factor | *n* without the risk factor | Missing data |
| Health problems at time 1 (total) | | | | | | |
| Prematurity time 1 | 5 | 65 | 0 | 8 | 68 | 0 |
| Low birth weight time 1 | 6 | 64 | 0 | 5 | 71 | 0 |
| Ear infection at time 1 | 1 | 69 | 0 | 2 | 74 | 0 |
| Familial risk (someone in family) time 1 | 12 | 58 | 0 | 12 | 63 | 1 |
| Developmental disability time 1 | 0 | 69 | 1 | 0 | 76 | 0 |
| Visual or hearing impairment time 1 | 1 | 69 | 0 | 1 | 73 | 2 |
| Language concerns at time 1 | | | | | | |
| Hearing or communication concerns at time 1 | | 2 | 68 | 0 | 8 | 68 |
| Demographic factors at time 1 (total) | | | | | | |
| Maternal education time 1 | 9 | 61 | 0 | 3 | 73 | 0 |
| Household income time 1 | 18 | 52 | 0 | 16 | 60 | 0 |
| Language catch-up after concern at time 1 | 1 | 9 | 0 | 10 | 17 | 0 |

For this table, language not catching up after Concern at time 1 is categorized as 'with the risk factor' and language catching up is categorized as 'without the risk factor'.

TABLE 2 Number of children in each quartile for vocabulary and gestures at time 1 (group membership).

| Variable | 0–25th | 25–50th | 50–75th | 75–100th |
|---|---|---|---|---|
| Productive vocabulary | 39 | 30 | 39 | 38 |
| Receptive vocabulary | 33 | 32 | 32 | 49 |
| Gestures | 30 | 36 | 43 | 37 |

Physical/health factors:

2. Child sex/gender: operationalized as a binary variable (male female); male = score of 1
3. Prematurity: operationalized as a binary variable (premature/not premature): born before week 36 = score of 1
4. Low birth weight: operationalized as a binary variable (low/not low); weighing less than 5 lb. 8 oz. when born = score of 1
5. Ear infection: operationalized as answering yes to the question "has your child had an ear infection/ glue ear for longer than 3 months, 4 to 6 ear infections within a 6 month period, or another identified hearing problem?" = score of 1
6. Familial risk of language/literacy disorder: answering yes to "is there anyone in the immediate family with speech/language difficulty or dyslexia?" = score of 1
7. Developmental disability: answering yes to "does your child have a developmental disability?" = score of 1
8. Hearing or visual impairment: answering yes to "does your child have a hearing or visual impairment?" = score of 1

Demographic factors:

9. Maternal education: selecting "no formal qualifications" or "GCSE/O level/NVQ level 1 or 2" = score of 1
10. Household income: selecting "£0–£14,000" or "£14,000–£24,000" = score of 1

Although many studies look only at maternal education, we used both education and income categories as there is evidence that different operationalisations of SES can have different effects (see, e.g., De Cat, 2021). The cut off scores for maternal education and household income used above were designed to determine low SES status. For household income, families with income of around £22,800 per year were considered to have low income at the time of data collection (Department for Work and Pensions, 2019). Low maternal education was established as having no formal qualifications or GCSE/O level/ NBQ level 1 or 2. Previous research has shown that children of mothers with fewer than 12 years of education are at increased risk for persisting language impairment (Stanton-Chapman et al., 2002).

## 2.6. Language measures at time 1

### 2.6.1. Division by quartiles: group membership

Because we wanted to identify whether being in the bottom 25th percentile for vocabulary/gesture at time 1 would predict language-related disability at time 2, we divided the children into four groups based on vocabulary and gesture scores between 15 and 18 months using the UK-CDI norms. The UK-CDI norms were created using the entire UK-CDI Project sample and provide percentile cut-offs for productive and receptive vocabulary, and for gestures, for each month. Children were split into four groups based on percentiles: 0–25th, 25–50th, 50–75th, and above 75th percentiles. Each child was placed into one of the four groups separately for productive vocabulary, for receptive vocabulary and for gesture use.

### 2.6.2. Concern scores at time 2

Five questions asked about the child's likelihood of having a language-related disorder at time 2:

1. Parental concern for language development: answering yes to "have you ever worried that your child's speech was delayed compared to other children the same age?" = score of 1
2. Developmental disability: answering yes to "does your child have a developmental disability?" = score of 1
3. Diagnosis of language disorder: answering yes to "has your child been diagnosed with any of the following language disorders?" = score of 1
4. Hearing or visual impairment: answering yes to "does your child have a hearing or visual impairment?" = score of 1
5. Professional concern for language development. Identification by the Healthy Child Programme's 2 Year Review (the Two Year Check): answering yes to "did this programme identify any delays with your child's speech, language or communication abilities?" = score of 1

The parents' answers to these questions were used to calculate two overall scores for each child:

- Identified Disability: Children whose parents answered yes to any one of questions 2, 3 or 4 above were given a score of 1. Children whose parents answered no to questions 2, 3 and 4 were given a score of 0.
- Overall Concern: Children whose parents answered yes to any of the question above were given a concern score of 1. Those whose parents answered no to all of the questions above were given a concern score of 0.

Twenty children (13.70%) fit the criteria for having an Identified Disability (developmental disability =10, language disorder =1 and/or visual or hearing impairment = 12). Forty-nine children (33.56%) were identified as having language that was of Overall Concern at time 2. Note that DLD is estimated to affect approximately 7.58% of the population (Norbury et al., 2016).

Parents who answered yes to the question "have you ever worried that your child's speech was delayed compared to other children the same age?" were also asked a catch-up question ("Did your child's speech eventually catch up with that of other children the same age?"). Of the 37 parents who answered this question, 26 reported that their children's language had caught up with children the same age, and 11 reported that it had not caught up (7.53% of the full sample of 146 children).

## 2.7. Analysis strategy

All analyses were conducted using one-tailed tests, as all hypotheses are unidirectional hypotheses. All outliers were included, unless it was determined that the data point was due to experimenter or participant error. $Chi^2$ analyses were performed in R (R Core Team, 2017, R version 3.4.1) using R Studio (Version 1.0.153) using the CrossTable function as part of the gmodels package (Warnes et al., 2018). Logistic regressions were performed in R using the glm function as part of the pscl package (Jackman, 2010). Discriminant function analyses were run in SPSS Statistics 24.

Discriminant function analysis is a statistical technique used to determine how well predictor variables discriminate between two or more naturally occurring groups. Here we used it to determine which different combinations of risk factors gave us the best classification accuracy of children into our two outcome groups. For the first set of analyses the two groups were: children with an Identified Disability (1) and children without an Identified Disability (0) at time 2. For the second set of analyses the two groups were: children for whom concern about language had been expressed (1) and children for whom concern about language had not been expressed (0) by time 2. Discriminant function analysis yields an overall accuracy figure (how well the model performs at discrimination overall), and sensitivity and specificity values. The sensitivity value measures the ability of the model to correctly classify children who have an identified disability or for whom there is concern for language development (true positives). Specificity measures the ability of the model to correctly classify children who do not have an identified disability or for whom there has been no concern expressed about their language development (true negatives).

Sensitivity and specificity rates between 70 and 80% are deemed acceptable for diagnostic assessments (e.g., for autistic spectrum disorder; Bright Futures Steering Committee and Medical Home Initiatives for Children with Special Needs Project Advisory Committee, 2006). Therefore, we consider accuracy, sensitivity and specificity and values above 70% to be adequate in the present analyses. Discriminant function analyses also provides standardized canonical coefficients for each variable. These coefficients allow us to compare the weighted importance of each variable in predicting group membership.

# 3. Results

## 3.1. Descriptive statistics

Table 1 shows the number of children in the different risk factor categories at time 1 (15–18 months) and Table 2 details the number of children in each quartile group for productive vocabulary, receptive vocabulary and gestures at time 1 (15–18 months). Table 3 shows the number of children who had been diagnosed with an identified disability and/or whose language had raised concern at time 2.

We checked for collinearity in our predictor variables as this can affect the interpretability of regression models. We ran $Chi^2$ analyses between each pair of variables to check for associations, and then, for any two variables that yielded significant $Chi^2$ scores, we followed this up with a Cramer's V post-test to determine collinearity. Cramer's V provides an effect size where values vary between 0 and 1, with 0 indicating no collinearity and 1 indicating high collinearity. The $Chi^2$ analyses revealed that eight of the predictor variables were significantly associated (prematurity and low birth weight; low birth weight and family history of language delay or dyslexia; ear infection at time 1 and visual or hearing impairment at time 1; family history of language

10.3389/fpsyg.2023.1167810

delay or dyslexia and a visual or hearing impairment at time 1; family history of language delay or dyslexia and maternal education; visual or hearing impairment at time 1 and maternal education; maternal education and household income; productive vocabulary group and receptive vocabulary group; productive vocabulary and gesture group; receptive vocabulary and gesture group). However, none of these variables were highly collinear (all Cramer's V values below 0.70). Collinearity between developmental disability at time 1 and other variables could not be established because no parents reported developmental disability at time 1.

## 3.2. Predicting identified disability at time 2 from risk factors and language at time 1

First, we ran discriminant function analyses to assess the discriminatory ability of the risk factors and language group at time 1 to correctly classify children into two groups: Identified Disability or No Identified Disability.

We ran five analyses (see Table 4 for the overall results, and Table 5 for the standardized canonical coefficients for each variable in each model, which indicate the weighted importance of each variable in predicting group membership). The first analysis included only language group at time 1 (quartile groups for productive vocabulary, receptive vocabulary and gestures at time 1) to determine if this could predict group membership at time 2 (Identified Disability, No Identified Disability). This model failed to correctly classify children into their groups, $r = 0.13$, $\chi^2 = 2.36$, $df = 3$, $p < 0.50$. The accuracy of the model was high, 86.30%, and it had good specificity (100.00%), meaning that it did well in classifying children in with no identified disability. However, the sensitivity was poor at 0.00%, so the model did not do well in identifying the children in the Identified Disability group at time 2.

The second discriminant function analysis tested only the effect of health risk factors at time 1: child sex/gender, prematurity, low birth weight, ear infections, familial risk for speech or language impairment, developmental disability, and visual or hearing impairments. This model correctly classified children into their groups with an accuracy of 86.80%, $r = 0.35$, $\chi^2 = 1790$, $df = 6$, $p = 0.006$. Again, although the specificity of the complete model was excellent (98.40%) meaning it correctly classified almost all children in the No Identified Disability group, it did not do well in terms of correctly classifying children in the Identified Disability group (sensitivity = 10.50%).

The third discriminant function analysis tested only the effect of demographic factors: maternal education and household income. This model failed to correctly classified children into their groups, $r = 0.16$, $\chi^2 = 3.26$, $df = 2$, $p = 0.16$. Again, the accuracy of the model was good (86.30%), and the sensitivity was high (100.00%), sensitivity was poor (0.00%).

The fourth discriminant function analysis tested only the effect of parental concern for language development at time 1. This model failed to correctly classify children into their groups, with an accuracy of 86.30%, $r = 0.13$, $\chi^2 = 2.39$, $df = 1$, $p = 0.12$. Again, the sensitivity of this model was poor, 0.00%, therefore it did not do well in terms of correctly classifying children with an identified disability at time 2. The specificity of the model, however, was excellent (100.00%) as it correctly classified all children without an identified disability at time 2.

The fifth discriminant function analysis included all risk factors and language group at time 1 to determine if these variables together are better predictors of group membership than separately. The included risk factors were health factors (child sex/gender, prematurity, low birth weight, ear infections at time 1, familial risk for speech or language impairment, developmental disability, visual or hearing impairment), demographic factors (maternal education, household income), concern expressed at time 1 (hearing or communication concerns at time 1), and language groups at time 1 (quartile groups for productive and receptive vocabulary and for gestures). This model correctly classified children into these two groups with an accuracy rate of 86.10%, $r = 0.40$, $\chi^2 = 22.87$, $df = 12$, $p = 0.03$. Again, however, as with all the previous models, though specificity was good (96.80%), sensitivity was poor (15.80%). In sum,

TABLE 3 Number of children with and without each risk factor at time 2.

| Risk factor | Female | | | Male | | |
|---|---|---|---|---|---|---|
| | *n* with the risk factor | *n* without the risk factor | Missing data | *n* with the risk factor | *n* without the risk factor | Missing data |
| Identified Disability (answering "yes" to at least one of the three questions on diagnosis of developmental disability, diagnosis of language impairment, having a visual or hearing impairment at time 2) | 4 | 66 | 0 | 16 | 60 | 0 |
| Overall Concern at time 2 (answering "yes" to any of the 5 questions which denote concern) | 16 | 54 | 0 | 33 | 43 | 0 |
| Concern expressed by parent at time 2 | 10 | 60 | 0 | 27 | 49 | 0 |
| Concern expressed at 2 Year Review time 2 | 6 | 56 | 8 | 16 | 45 | 15 |
| Diagnosis of developmental disability time 2 | 3 | 67 | 0 | 7 | 69 | 0 |
| Visual or hearing impairment time 2 | 1 | 69 | 0 | 11 | 65 | 0 |
| Diagnosis of language disorder time 2 | 0 | 70 | 0 | 1 | 75 | 0 |

The total number of children with any one risk factor at time 2 may exceed the total number of children with overall concern at time 2. This is because some children experienced more than one risk factor.

10.3389/fpsyg.2023.1167810

TABLE 4 Results from the discriminant function analyses distinguishing between children with and without an identified disability at time 2.

| Variable | $r$ | $\chi 2$ | $n$ | df | $p$ | Accuracy | Sensitivity | Specificity |
|---|---|---|---|---|---|---|---|---|
| Language group at time 1 (quartile groups for vocabulary and gesture scores at time 1) | 0.13 | 2.36 | 146 | 3 | 0.50 | 86.30% | 0.00% | 100.00% |
| Health factors (sex/gender, prematurity, low birth weight, ear infection, visual or hearing impairment, family history, developmental disability) | 0.35 | 17.90 | 143 | 6 | 0.006 | 86.80% | 10.50% | 98.40% |
| Demographic factors (maternal education, family income) | 0.16 | 3.62 | 146 | 2 | 0.16 | 86.30% | 0.00% | 100.00% |
| Concern at time 1 | 0.13 | 2.39 | 146 | 1 | 0.12 | 86.30% | 0.00% | 100.00% |
| All variables | 0.40 | 22.87 | 143 | 12 | 0.03 | 86.10% | 15.80% | 96.80% |

all of the models had low sensitivity, and were thus unable to identify children in the Identified Disability group with reliable levels of accuracy.

## 3.3. Predicting overall concern for language at time 2 from risk factors and language at time 1

Since many children do not have a diagnosis of a language-related disorder by the time they are 4–6 years old, even if one is present, we ran these five analyses again with the broader category that included children with both an identified disability and children for whom a parent or professional had expressed 'concern' about their language development. For these analyses, children were split into two groups: Overall Concern and No Overall Concern. Here the sensitivity value measures the ability of the model to correctly classify children in the Overall Concern group (true positives) and specificity measure the model to correctly classify children in the No Overall Concern group (true negatives). The results from these five analyses can be seen in Tables 5, 6. As with the previous five analyses, the sensitivity of these models was very poor; they were unable to accurately classify children for whom there was concern about their language development.

## 3.4. Predicting catch-up in language development from risk factors recorded at time 1

In this section of analysis we tested whether our risk factors and language scores at time 1, when combined, could predict 'catch up ability' (i.e., could distinguish between children whose language had been of concern at some point in their development but whose difficulties resolved by time 2, and those whose language was still of concern). This analysis included only the subset of children whose parents answered yes to the question "Have you ever worried that your child's speech was delayed compared to other children the same age?" at time 2. For this analysis, the independent variables were all of the risk factors (child sex/gender, prematurity, low birth weight, ear infections at time 1, familial risk for speech of language impairment, developmental disability, visual or hearing impairment, maternal education, household income, concern at time 1), as well as vocabulary and gesture scores at time 1. The dependent variable was the answers to the language catch-up question at time 2 ("Did your child's speech eventually catch up to that of other children the same age?"). A total of 37 parents expressed concern for their children's language development at time 2. Of these 37, 26 reported that their children's language had caught up with children the same age, and 11 reported that it had not caught up.

We ran a discriminant function analysis to assess the discriminatory ability of our time 1 risk/language factors to correctly classify children whose language did (0) and did not (1) catch up by time 2. This model did not correctly classify children into their groups. While the accuracy of this model was good, 82.90%, it did not reach significance, $r = 0.57$, $\chi^2 = 10.66$, $df = 11$, $p = 0.47$. This result is reflected in the sensitivity of the model. The sensitivity of this model was poor, 54.50% meaning it did not do well at classifying children whose language did not catch up. The specificity of the model was good, 95.80%, meaning it did well in terms of correctly classifying children whose language did catch up. See Table 7 for the results of the discriminant function analysis and Table 8 for the standardized canonical coefficients for each variable.

## 3.5. Exploratory analyses

We ran three exploratory descriptive analyses to investigate possible reasons why the risk factors we identified were not sensitive enough to correctly classify children at time 2. We focussed on the concern measures, since few children had been diagnosed with a language-related disorder at time 2. Table 9 details the number and percentage of children in each Concern group (Overall Concern, No Overall Concern) with each risk factor. We can see from this table that the proportion of children with the risk factor is almost always bigger in the Overall Concern group than the No Overall Concern group. For example, if we consider the family history of language delay or dyslexia risk

TABLE 5 Standardized canonical discriminant function coefficients of each discriminant function analysis predicting overall and identified concern.

| Model | Variable | *r* (Overall Concern) | *r* (Identified Concern) |
|---|---|---|---|
| Language group | Productive vocabulary group | 1.07 | 0.66 |
| | Receptive vocabulary group | −0.18 | −0.36 |
| | Gesture group | 0.08 | 0.79 |
| | Sex/Gender | 0.69 | 0.67 |
| | Prematurity | −0.25 | −0.64 |
| | Low Birth Weight | 0.45 | 0.68 |
| | Ear Infection | 0.70 | 0.60 |
| | Family history | 0.22 | −0.11 |
| Health factors | Visual or hearing impairment | −0.15 | 0.16 |
| Demographic factors | Maternal education | −0.48 | −0.09 |
| | Household income | 1.02 | 1.02 |
| Concern at time 1 | Concern at time 1 | 1.00 | 1.00 |
| All variables | Sex/Gender | 0.36 | 0.57 |
| | Prematurity | 0.04 | −0.47 |
| | Low Birth Weight | 0.13 | 0.52 |
| | Ear Infection | 0.30 | 0.45 |
| | Visual or hearing impairment | 0.10 | 0.24 |
| | Family history | −0.01 | −0.22 |
| | Maternal education | −0.38 | −0.12 |
| | Household income | 0.66 | 0.46 |
| | Concern at time 1 | 0.08 | 0.17 |
| | Productive vocabulary group | −0.62 | −0.08 |
| | Receptive vocabulary group | 0.09 | 0.10 |
| | Gesture group | 0.09 | −0.13 |

factor, 20.83% of children in the Overall Concern group have that risk factor, compared to 14.43% in the No Overall Concern group. Thus, the expected differences in the prevalence and number of risk factors between groups is present in our sample. However, the differences are not big; for most risk factors, a substantial minority of children in the No Overall Concern group also have the risk factor.

Next, we looked at the number and proportion of children with and without overall concern in each of the language quartile groups at time 1 (Table 9). We can see from Table 9 why language and gesture scores at time 1 do not discriminate between groups at time 2. Again, although there are a greater proportion of children in the lowest quartiles who subsequently raise concerns than in the higher quartiles (e.g., 40.82% for 0–25th percentile vs. 12.24% in 75–100th percentile for productive vocabulary) the differences are not large or distinct enough to be discriminatory. A substantial minority of children in the higher quartiles go on to develop language that is of concern, and a substantial minority of children in the lower quartiles do not.

Finally, we created total risk factor scores for each child in the Overall Concern and No Overall Concern groups. The total number of risk factors was 10 (being a boy, being premature, having a low birth weight, ear infections, family history of language delay or dyslexia, having a developmental disability at time 1, having a visual or hearing impairment at time 1, hearing or communication concerns at time 1, low maternal education and low household income). The mean number of risk factors for children in the Overall Concern group was 1.71 (SD = 1.14) and the mean number of risk factors for children in the No Overall Concern group was 1.04 (SD = 1.09). Thus, once again, although the difference was in the expected direction, it was not large. In addition, the overlap in standard deviation of both groups was big, and the range was the same across the two groups (0–5). Therefore, while children in the Overall Concern group experience a larger number of risk factors overall, the differences are almost certainly not big or distinct enough to be discriminatory.

## 4. Discussion

The goal of the present study was to investigate if we could use a combination of risk factors, earlier vocabulary and gesture scores, as well as early parental concern about language at time 1 to predict (a) which children will have an identified language disability and (b) later concern for language development at time 2.

TABLE 6 Results from the discriminant function analyses distinguishing between children with and without overall concern for their language development at time 2.

| Variable | *r* | χ2 | *n* | df | *p* | Accuracy | Sensitivity | Specificity |
|---|---|---|---|---|---|---|---|---|
| Language group at time 1 (quartile groups for vocabulary and gesture scores at time 1) | 0.35 | 18.89 | 146 | 3 | <0.001 | 67.10% | 40.80% | 80.40% |
| Health factors (sex/gender, prematurity, low birth weight, ear infection, visual or hearing impairment, family history, developmental disability) | 0.33 | 15.40 | 143 | 6 | 0.02 | 70.80% | 14.90% | 97.90% |
| Demographic factors (maternal education, family income) | 0.25 | 9.48 | 146 | 2 | 0.009 | 71.20% | 34.70% | 89.70% |
| Concern at time 1 | 0.15 | 3.35 | 146 | 1 | 0.07 | 67.80% | 12.20% | 95.90% |
| All variables | 0.50 | 39.26 | 143 | 12 | <0.001 | 75.70% | 42.60% | 91.80% |

TABLE 7 Results of the discriminant function analysis distinguishing between children whose language did and did not catch up.

| Variable | *r* | χ2 | *n* | df | *p* | Accuracy | Sensitivity | Specificity |
|---|---|---|---|---|---|---|---|---|
| All variables | 0.57 | 10.66 | 35 | 11 | 0.47 | 82.90% | 54.50% | 95.80% |

Only 35 children (of the 37 whose language was of concern at some point) had data available for all variables in this analysis.

TABLE 8 Standardized canonical discriminant function coefficients for each variable predicting whether or not children's language caught up after concern was expressed.

| Model | Variable | r |
|---|---|---|
| All variables | Sex/Gender | −0.53 |
| | Prematurity | 0.69 |
| | Low birth weight | −0.17 |
| | Ear infection | −0.15 |
| | Visual or hearing impairment | 0.51 |
| | Family history | 0.18 |
| | Maternal education | NA[1] |
| | Household income | −0.13 |
| | Concern at time 1 | −0.63 |
| | Productive vocabulary group | −0.22 |
| | Receptive vocabulary group | −0.38 |
| | Gesture group | 1.03 |

[1]Only one child whose language did or did not catch up had low maternal education so it did not contribute to classification accuracy.

We used discriminant function analyses to examine if different combinations of health and demographic risk factors as well as earlier vocabulary and gesture scores, and concern measures, could discriminate children who had, at 4–6 years (time 2), either an identified language-related disability (Identified Disability) or for whom concern about language development had been registered by a parent or professional (Overall Concern). None of our models yielded successful predictors to identify either disability or concern.

We first examined the role of earlier language and gesture scores but these variables did not successfully discriminate between children (i.e., failed to identify both children with an identified disability and children in the Overall Concern groups). When we examine the number of children in each vocabulary and gesture quartile (Table 9), we see why: children in the two groups were distributed across all four quartiles with very little clustering at each end for each group. For example, only 19.59% and 13.40% of children in the overall Concern group at time 2 were in the bottom 0–25th and 25–50th, respectively, at time 1.

We then examined a number of health or demographic risk factors. As with early language skills above, these risk factors did not successfully identify children who had an identified disability or who were in the Overall Concern group. Again, our exploratory analyses shows why this was the case. For example, although the children in the Overall Concern group were reported at time 1 to have experienced a greater number of risk factors on average (1.71 vs. 1.04), there was considerable overlap (wide and overlapping standard deviations and ranges) in the number of risk factors in the Overall Concern/No Overall Concern groups. This means that no combination of health or demographic risk factors was discriminant enough to distinguish between these two groups of children. Furthermore, when we consider that there were a maximum of 10 risk factors and the most risk factors any one child experienced was 5, we can see that neither children with, nor children without, overall concern for their language development were exposed to a very high number of these risk factors. This result is consistent with previous research, which has shown that health and demographic risk factors are better at predicting individual differences than they are at predicting language impairment or concern for language development (Harrison and McLeod, 2010; Reilly et al., 2010). For example, Harrison and McLeod (2010) found health and demographic risk factors did not predict parental concern for vocabulary development or use of speech-language pathology services, with even a combination of these factors yielding poor levels of sensitivity.

TABLE 9 Number and percentage of children with each risk factor split by Overall Concern and No Overall Concern groups.

| Variable | Have risk factor | Overall Concern *n* (%) | No Overall Concern *n* (%) |
|---|---|---|---|
| Sex/Gender (male) | Yes | 33(67.35%) | 43(44.33%) |
| | No | 16(32.65%) | 54(55.67%) |
| Prematurity | Yes | 5(10.20%) | 8(8.25%) |
| | No | 44(89.80%) | 89(91.75%) |
| Low birth weight | Yes | 5(10.20%) | 6(6.19%) |
| | No | 44(89.80%) | 91(93.81%) |
| Ear infection at time 1 | Yes | 3(6.12%) | 0(0.00%) |
| | No | 46(93.88%) | 97(100.00%) |
| Family history | Yes | 10(20.83%) | 14(14.43%) |
| | No | 38(79.17%) | 83(85.57%) |
| Developmental disability at time 1 | Yes | 0(0.00%) | 0(0.00%) |
| | No | 48(100.00%) | 97(100.00%) |
| Visual impairment at time 1 | Yes | 1(2.13%) | 1(1.03%) |
| | No | 46(97.87%) | 96(98.97%) |
| Concern at time 1 | Yes | 6(12.24%) | 4(4.12%) |
| | No | 43(87.76%) | 93(95.88%) |
| Maternal education | Yes | 3(6.12%) | 9(9.28%) |
| | No | 46(93.88%) | 88(90.72%) |
| Family income | Yes | 18(36.73%) | 16(16.49%) |
| | No | 31(63.27%) | 81(83.51%) |
| Productive vocabulary at 15–18 months | 0–25th percentile | 20(40.82%) | 19(19.59%) |
| | 25–50th percentile | 17(34.69%) | 13(13.40%) |
| | 50–75th percentile | 6(12.24%) | 33(34.02%) |
| | 75th percentile and above | 6(12.24%) | 32(32.99%) |
| Receptive vocabulary at 15–18 months | 0–25th percentile | 14(28.57%) | 19(19.59%) |
| | 25–50th percentile | 13(26.53%) | 19(19.59%) |
| | 50–75th percentile | 12(24.49%) | 20(20.62%) |
| | 75th percentile and above | 10(20.41%) | 39(40.21%) |
| Gesture scores at 15–18 months | 0–25th percentile | 15(30.61%) | 15(15.46%) |
| | 25–50th percentile | 13(26.53%) | 23(23.71%) |
| | 50–75th percentile | 8(16.33%) | 35(36.08%) |
| | 75th percentile and above | 13(26.53%) | 24(24.74%) |

Even our full model, which included all risk factors, earlier vocabulary and gesture scores and early concern for language development at time 1, failed to identify individual children who had an identified disability or who were in the Overall Concern group at time 2. This may seem surprising given the wealth of evidence suggesting, for example, that some children who are late talkers in early childhood are more at risk of later language delay (Bishop and Adams, 1990; Rescorla, 2002). However, our findings are in line with much of the previous research that focusses on predicting *which* late talkers will develop language disorder. For example, Reilly et al. (2010) found that while risk factors were successful in predicting continuous language scores at 4 years (i.e., individual differences), they were unable to correctly classify children with DLD. Our findings are also in line with a recent systematic review and meta-analysis of screening tools for language disorder, which concluded that only a very small number (13.8%) of the 67 screening tools tested yielded good accuracy at identifying children with language disorder (So and To, 2022).

We then examined the stability of parental concern over time. Again, contrary to our predictions, overall concern at time 2 was not predicted by parental concern at time 1. Previous research has shown that parental concern for language delay can benefit clinical detection of language impairment (Glascoe, 1991; Glascoe and Dworkin, 1995), but this research has typically been carried out with older children. The findings here suggest that early concern for language development may not be as beneficial for predicting later problems. However, it is important to note that very few parents reported concern at time 1; only 6 parents of children who expressed concern at time 2 also expressed concern at time 1 (see Table 9). Therefore, we are hesitant to draw a firm conclusion that there is no relationship between parental concern at time 1 and Overall Concern at time 2.

Finally, we investigated if catch-up in language delay is associated with exposure to fewer risk factors and better scores on earlier vocabulary and gestures. In line with the results of the previous models, this model had poor sensitivity and did not correctly classify children whose language did not catch up (i.e., children whose language development was of the greatest concern).

One could argue that one reason for our overall pattern of results could be the narrowness of our sample in terms of demographic characteristics. Although the full UK-CDI sample was representative of the UK population, we found that parents from families with higher income and higher maternal education were more likely to respond to the follow-up questionnaire (see Supplementary materials https://osf.io/gvz3x/). This is a common problem when trying to collect representative samples (Reilly et al., 2010). However, we do not think that this explains our pattern of results. Just under 14% of our follow up sample had an identified disability, compared to 7.58% of all children in the UK population with DLD (Norbury et al., 2016). Thus, we would argue that we have data from a representative enough portion of the population, at least when it comes to the incidence of language-related disability, if not socio-economic status. In addition, there were no differences in the sample characteristics of those who did/did not take part in the follow up study in terms of the other risk factors. That said, it is important that low SES families are represented in studies, so future research should therefore make an increased effort to contact families represented in lower SES brackets. We may have been more successful at encouraging families to participate if we had personally contacted them, either *via* email or phone.

Another possible reason for the lack of predictiveness at the individual level could lie in children's changing environments between 15 and 18 months and 4–6 years having a substantial effect on their language development. Previous research has shown that, once children start attending playgroups and nurseries (typically at about 1–2 years of age in the UK), this has a substantial impact on their cognitive development (Turner, 1974; Sylva et al., 2011). Hence, it is plausible that these differences also impact on language development. The growing influence of peers and nursery staff on children contributes to their language diversification and growth, as well as to vocabulary growth and complexity of language. For example, children who hear little language in the home, and may develop slowly at first, might start to be looked after by grandparents or attend a nursery that promotes language development, and thus start to enhance their productive and receptive vocabulary and grammar. In other words, changes in children's environment over the preschool years may weaken the link between early and later language development.

Given that the profile of children with persisting language impairments into later childhood tends to be characterized by greater difficulties in syntax or pragmatics (Norbury and Bishop, 2002; Bishop, 2014) than in vocabulary (Bishop and Snowling, 2004; Leonard, 2014), it is worth to pay attention to underlying mechanisms. It is possible that the cognitive mechanisms that underpin language acquisition early in life (where the focus is on learning to use and interpret gestures and words) are not at the root of language impairment later in childhood (where development is more focussed on syntax and pragmatics). While some (e.g., Locke, 1997) propose that the development of lexicon and innate syntactic complexity emerge from separate linguistic mechanisms that activate in sequence, others suggest that more fundamental learning abilities including categorization and pattern recognition, joint attention, intention-reading and input analysis are emergent, and such joint factors drive the development of vocabulary and of grammatical constructions over time (e.g., De Ruiter et al., 2020). On both above approaches, there might be a different and changing relationship between vocabulary and syntax, but not one that is strong enough to predict later concern for language development. Alternatively, it may be possible that different (but partially overlapping) mechanisms underlie vocabulary and syntax acquisition, and, thus, that delays in vocabulary and syntax acquisition stem from different causes (see van der Lely et al., 1998, for a theory of specific language impairment based on this premise).

In line with this, one possible method for identifying which children go on to have persisting language impairments would be to examine the composition of late talker's vocabulary (Perry et al., 2022). Perry et al. (2022) found that late talking children produced significantly fewer shape-based nouns compared to typically developing children. It is possible that in addition to grouping children by quartiles, greater accuracy in predicting later language-related disabilities could be achieved by analyzing the composition of children's vocabulary when they are between 15 and 18 months old. However, this method would be limited to children who are already producing words at this age.

Furthermore, a shift in parental perception may also explain the independence of parental concern measures at time 1 and time 2; it is possible that the factors driving parental concern at time 1 and time 2 are not the same. For example, when children are 15–18 months old, a delay in productive vocabulary development is likely to be responsible for parental concern. However, by the time children are 4–6 years old, concern may focus on difficulties associated with syntax acquisition (Leonard, 1998; Bishop and Snowling, 2004; Leonard, 2014). Therefore, parental concern at time 1 and time 2 may stem from different sources.

One potential limitation is the use of parental report of developmental disability and the effect of recall bias. Relying on parental reports may not provide a completely accurate picture of their earlier concern for their children's language development. It is possible that a subsequent diagnosis of a developmental disability may result in parents misremembering their concern for their children's language development at an earlier timepoint. However, our analyses predicting Identified Disability controlled for any inflation in parental report of concern for language development. In addition, if there was an inflation in parental concern for language development due to a diagnosis of a developmental disability, it did not impact our ability to discriminate between children with and without concern for their language development.

In sum, throughout all of our analyses, there was an interesting overall picture: While we can predict reliably which children will go on to develop in a typical fashion, we fail to predict individual children with an identified disability, or with an overall concern for their language development. Given the number of robust correlations between early and later language, and between risk factor variables and language development in the literature, this may seem like a surprising result. However, our research has shown that such associations are not precise or accurate enough to enable us to identify individual children at risk of language disability.

## 5. Conclusion

The present findings shed light on the role of early language delays, and of health and demographic risk factors, in predicting later language outcomes. Such risk factors are currently recommended as a starting point when monitoring language delay in young children because the incidence of persisting language impairments is greater in children from families exposed to greater risks (Harrison and McLeod, 2010). However, we have found that these risk factors, either alone or when combined with early language measures, do not allow us to accurately predict identified disability or concern for children's language development over time. Thus, we cannot currently recommend that they be used as described above to screen individual children at risk for language impairment and further research needs to be carried out improve sensitivity.

## Data availability statement

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

## Ethics statement

The studies involving human participants were reviewed and approved by The University of Liverpool's Research Ethics Subcommittee for Non-Invasive Procedures (Institute Review Board protocol number: RETH000764). Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

## Author contributions

LJ and CR designed the follow-up study. CR, KA, and KM ran the original UK-CDI Project and contributed to the development of the manuscript. CR and KA contributed to the design of the follow-up questionnaire. JP contributed to the development of the manuscript. All authors contributed to the article and approved the submitted version.

## Funding

This study was completed as part of the first author's PhD. The funding for this PhD came from the ESRC International Centre for Language and Communicative Development [ES/L008955/1].

## Conflict of interest

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

## Editor's note

Maria-José Ezeizabarrena edited the article in collaboration with Melita Kovacevic, University of Zagreb, Zagreb, Croatia.

## Publisher's note

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

## Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2023.1167810/full#supplementary-material

## References

Alcock, K.J., Meints, K., Rowland, C. F., Christopher, A., Just, J., and Brelsford, J (2020). *The UK communicative development inventory: words and gestures*. Guildford J&R Press Ltd

All Party Parliamentary Group on Speech and Language Difficulties. (2013). The links between speech, language and communication needs and social disadvantage. Available at: https://www.rcslt.org/governments/docs/all_party_parliamentary_group_on_slcn_inquiry_report

Barre, N., Morgan, A., Doyle, L. W., and Anderson, P. J. (2011). Language abilities in children who were very preterm and/or very low birth weight: a meta-analysis. *J. Pediatr.* 158, 766–774. doi: 10.1016/j.jpeds.2010.10.032

Bishop, D. V. M. (2014). "Pragmatic language impairment: a correlate of SLI, a distinct subgroup, or part of the autistic continuum?" in *Speech and language impairments in children: causes, characteristics, intervention and outcome* (London: Psychology Press), 113–128.

Bishop, D. V. M., and Adams, C. (1990). A prospective study of the relationship between specific language impairment, phonological disorders and reading retardation. *J. Child Psychol. Psychiatry* 31, 1027–1050. doi: 10.1111/j.1469-7610.1990.tb00844.x

Bishop, D. V. M., Holt, G., Line, E., McDonald, D., McDonald, S., and Watt, H. (2012). Parental phonological memory contributes to prediction of outcome of late talkers from 20 months to 4 years: a longitudinal study of precursors of specific language impairment. *J. Neurodev. Disord.* 4, 1–12. doi: 10.1186/1866-1955-4-3

Bishop, D. V. M., and Snowling, M. J. (2004). Developmental dyslexia and specific language impairment: same of different? *Psychol. Bull.* 160, 858–886. doi: 10.1037/0033-2909.130.6.858

Bölte, S., Neufeld, J., Marschik, P. B., Williams, Z. J., Gallagher, L., and Lai, M. (2023). Sex and gender in neurodevelopmental conditions. *Nat. Rev. Neurol.* 19, 136–159. doi: 10.1038/s41582-023-00774-6

Bright Futures Steering Committee and Medical Home Initiatives for Children with Special Needs Project Advisory Committee (2006). Identifying infants and young children with developmental disorders in the medical home: an algorithm for developmental surveillance and screening. *Am. Acad. Pediatr.* 118, 405–420. doi: 10.1542/peds.2006-1231

Campbell, T. F., Dollaghan, C. A., Rockette, H. E., Paradise, J. L., Feldman, H. M., Shriberg, L. D., et al. (2003). Risk factors for speech delay of unknown origin in 3-year-old children. *Child Dev.* 74, 346–357. doi: 10.1111/1467-8624.7402002

Carroll, J. M., and Breadmore, H. L. (2017). Not all phonological awareness deficits are created equal: evidence from a comparison between children with Otitis Media and poor readers. *Dev. Sci.* 21, 1–12. doi: 10.1111/desc.12588

Dale, P. S., and Hayiou-Thomas, M. E. (2013). "Outcomes for late talkers. A twin study" in *Communication and language intervention series. Late talkers: language development, interventions, and outcomes*. eds. L. A. Rescorla and P. S. Dale (Baltimore: Paul H. Brookes Publishing Co.), 241–257.

Dale, P. S., Price, T. S., Bishop, D. V. M., and Plomin, R. (2003). Outcomes of early language delay I: predicting persistent and transient delay at 3 and 4 years. *J. Speech Lang. Hear. Res.* 46, 544–560. doi: 10.1044/1092-4388(2003/044)

De Cat, C. (2021). Socioeconomic status as a proxy for input quality in bilingual children? *Appl. Psycholinguist.* 42, 301–324. doi: 10.1017/S014271642000079X

De Ruiter, L. E., Lieven, E., Brandt, S., and Theakston, A. L. (2020). Interactions between givenness and clause order in children's processing of complex sentences. *Cognition* 198:104130. doi: 10.1016/j.cognition.2019.104130

Department for Work and Pensions. (2019). *Households below average income: an analysis of the UK income distribution: 1994/95–2014/15*. Available at: https://www.gov.uk/government/statistics/households-below-average-income-199495-to-201415

Department of Health. (2009). *Healthy child Programme: the two year review*. Available at: https://www.gov.uk/government/publications/healthy-child-programme-review-of-children-aged-2

Duff, F. J., Nation, K., Plunkett, K., and Bishop, D. V. M. (2015). Early prediction of language and literacy problems: is 18 months too early? *PeerJ* 3, 1–12. doi: 10.7717/peerj.1098

Eriksson, M., Marschik, P. B., Tulviste, T., Almgren, M., Pereira, M. F., Wehberg, S., et al. (2012). Differences between girls and boys in emerging language skills: evidence from 10 language communities. *Br. J. Develop. Psychol.* 30, 326–343. doi: 10.1111/j.2044-835x.2011.02042.x

Evans, G. W., Maxwell, L. E., and Hart, B. (1999). Parental language and verbal responsiveness to children in crowded homes. *Dev. Psychol.* 35, 1020–1023. doi: 10.1037/0012-1649.35.4.1020

Fan, S., Zhang, Y., Qin, J., Song, X., Wang, M., and Ma, J. (2021). Family environmental risk factors for developmental speech delay in children in northern China. *Sci. Rep.* 11, 1–7. doi: 10.1038/s41598-021-83554-w

Fisher, E. L. (2017). A systematic review and meta-analysis of predictors of expressive-language outcomes among late talkers. *J. Speech Lang. Hear. Res.* 60, 2935–2948. doi: 10.1044/2017_JSLHR-L-16-0310

Glascoe, F. P. (1991). Can clinical judgment detect children with speech-language problems? *Pediatrics* 87, 317–322. doi: 10.1542/peds.87.3.317

Glascoe, F. P., Altemeier, W. A., and MacLean, W. E. (1989). The importance of parents' concerns about their child's development. *Am. J. Dis. Child.* 143, 955–995.

Glascoe, F. P., and Dworkin, P. H. (1995). The role of parents in the detection of developmental and behavioural problems. *Pediatrics* 95, 829–836. doi: 10.1001/archpedi.1989.02150200115029

Grossheinrich, N., Schulte-Körne, G., Marschik, P. B., Kademann, S., von Suchodoletz, W., and Sachse, S. (2019). School-age outcomes of late-talking toddlers: long-term effects of an early lexical deficit. *Dev. Sci.* 22, 1–13. doi: 10.1111/desc.12826

Harrison, L. J., and McLeod, S. (2010). Risk and protective factors associated with speech and language impairment in a nationally representative sample of 4-to 5-year-old children. *J. Speech Lang. Hear. Res.* 53, 508–529. doi: 10.1044/1092-4388(2009/08-0086)

Henrichs, J., Rescorla, L., Schenk, J. J., Schmidt, H. G. G., Jaddoe, V. W. V., Hofman, A., et al. (2011). Examining continuity of early expressive vocabulary development: the generation R study. *J. Speech Lang. Hear. Res.* 54, 854–869. doi: 10.1044/1092-4388(2010/09-0255)854

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., and Lyons, T. (1991). Early vocabulary growth: relation to language input and gender. *Dev. Psychol.* 27, 236–248. doi: 10.1037/0012-1649.27.2.236

Jackman, S. (2010). *pscl: Classes and methods for R. developed in the political science computational laboratory*, Stanford University. Department of Political Science, Stanford University, Stanford, CA

Jansson-Verkasalo, E., Valkama, M., Vainionpää, L., Pääkkö, E., Ilkko, E., and Lehtihalmes, M. (2004). Language development in very low birth weight preterm children: a follow-up study. *Folia Phoniatr. Logop.* 56, 108–119. doi: 10.1159/000076062

Jin, F., Schjølberg, S., Wang, M. V., Eadie, P., Nes, R. B., Røysamb, E., et al. (2020). Predicting literacy skills at 8 years from preschool language trajectories: a population-based cohort study. *J. Speech Lang. Hear. Res.* 63, 2752–2762. doi: 10.1044/2020_JSLHR-19-00286

Kennedy, C. R., McCann, D. C., Campbell, M. J., Law, C. M., Mullee, M., Petrou, S., et al. (2006). Language ability after early detection of permanent childhood hearing impairment. *N. Engl. J. Med.* 354, 2131–2141. doi: 10.1056/NEJMoa054915

Kenyhercz, F., and Nagy, B. E. (2022). A new perspective: establishing developmental profiles of premature infants based on Bayley-III scores at age 2. *Appl. Neuropsychol. Child* 11, 125–132. doi: 10.1080/21622965.2020.1771338

Kovas, Y., Hayiou-Thomas, M. E., Oliver, B., Dale, P. S., Bishop, D. V., and Plomin, R. (2005). Genetic influences in different aspects of language development: the etiology of language skills in 4.5-year-old twins. *Child Dev.* 76, 632–651. doi: 10.1111/j.1467-8624.2005.00868.x

Leonard, L. B. (1998). *Children with specific language impairment*. MA: The MIT Press.

Leonard, L. B. (2014). *Children with specific language impairment* (*2nd*). Cambridge, MA: MIT Press.

Locke, J. L. (1997). A theory of neurolinguistic development. *Brain Lang.* 58, 265–326. doi: 10.1006/brln.1997.1791

Marschik, P. B., Einspieler, C., Garzarolli, B., and Prechtl, H. F. (2007). Events at early development: are they associated with early word production and neurodevelopmental abilities at the preschool age? *Early Hum. Dev.* 83, 107–114. doi: 10.1016/j.earlhumdev.2006.05.009

Norbury, C. F., and Bishop, D. V. M. (2002). Inferential processing and story recall in children with communication problems: a comparison of specific language impairment, pragmatic language impairment and high-functioning autism. *Int. J. Lang. Commun. Disor.* 37, 227–251. doi: 10.1080/13682820210136269

Norbury, C. F., Gooch, D., Wray, C., Baird, G., Charman, T., Simonoff, E., et al. (2016). The impact of nonverbal ability on prevalence and clinical presentation of language disorder: evidence from a population study. *J. Child Psychol. Psychiatry Allied Discip.* 57, 1247–1257. doi: 10.1111/jcpp.12573

Paul, R. (2000). "Predicting outcomes of early expressive language delay: ethical implications" in *Speech and language impairments in children: Causes, characteristics, intervention and outcome*. eds. D. V. M. Bishop and L. B. Leonard (Hove: Psychology Press), 195–209.

Perry, L. K., Kucker, S. C., Horst, J. S., and Samuelson, L. K. (2022). Late bloomer or language disorder? Differences in toddler vocabulary composition associated with long-term language outcomes. *Dev. Sci.* e13342, 1–16. doi: 10.1111/desc.13342

R Core Team (2017) *R: a language and environment for statistical computing*. Available at: https://www.R-project.org/

Reilly, S., Wake, M., Bavin, E. L., Prior, M., Williams, J., Bretherton, L., et al. (2007). Predicting language at 2 years of age: a prospective community study. *Pediatrics* 120, 1441–1449. doi: 10.1542/peds.2007-0045

Reilly, S., Wake, M., Ukoumunne, O. C., Bavin, E., Prior, M., Cini, E., et al. (2010). Predicting language outcomes at 4 years of age: findings from early language in Victoria study. *Pediatrics* 126, 1530–1537. doi: 10.1542/peds.2010-0254

Rescorla, L. (2002). Language and reading outcomes to age 9 in late-talking toddlers. *J. Speech Lang. Hear. Res.* 45, 360–371. doi: 10.1044/1092-4388(2002/028)

Rescorla, L. (2011). Late talkers: do good predictors of outcome exist? *Dev. Disabil. Res. Rev.* 17, 141–150. doi: 10.1002/ddrr.1108

Rice, M. L., Taylor, C. L., and Zubrick, S. R. (2008). Language outcomes of 7-year-old children with or without a history of late language emergence at 24 months. *J. Speech Lang. Hear. Res.* 51, 394–407. doi: 10.1044/1092-4388(2008/029)

Roos, E. M., and Ellis Weismer, S. (2008). Language outcomes of late talking toddlers at preschool and beyond. *Perspect. Lang. Learn. Educat.* 15, 119–126. doi: 10.1044/lle15.3.119

Rowe, M. L. (2018). Understanding socioeconomic differences in parents' speech to children. *Child Dev. Perspect.* 12, 122–127. doi: 10.1111/cdep.12271

Rowe, M. L., Pan, B. A., and Ayoub, C. (2009). Predictors of variation in maternal talk to children: a longitudinal study of low-income families. *Parenting* 5, 285–310. doi: 10.1207/s15327922par0503_3

Save the Children. (2015). *Ready to read: Closing the gap in early language skills so that every child in England can read well*. Available at: https://www.savethechildren.org.uk/content/dam/global/reports/education-and-child-protection/ready-to-read.pdf

Sim, F., Thompson, L., Marryat, L., Law, J., and Wilson, P. (2019). Preschool developmental concerns and adjustment in the early school years: evidence from a Scottish birth cohort. *Child Care Health Dev.* 45, 719–736. doi: 10.1111/cch.12695

So, K. K.To, C. K (2022). Systematic review and Meta-analysis of screening tools for language disorder. *Front. Pediatr.* 10, 1–21. doi: 10.3389/fped.2022.801220

Solgi, M., Feryn, A., Chavez, A. E., Wilson, L., King, M., Scholz, J., et al. (2022). Parents' concerns are associated with early intervention evaluation and eligibility outcomes. *J. Dev. Behav. Pediatr.* 43, e145–e152. doi: 10.1097/DBP.0000000000001003

Stanton-Chapman, T. L., Chapman, D. A., Bainbridge, N. L., and Scott, K. G. (2002). Identification of early risk factors for language impairment. *Res. Dev. Disabil.* 23. doi: 10.1016/s0891-4222(02)00141-5

Stolt, S., Haataja, L., Lapinleimu, H., and Lehtonen, L. (2009). The early lexical development and its predictive value to language skills at 2 years in very-low-birth-weight children. *J. Commun. Disord.* 42, 107–123. doi: 10.1016/j.jcomdis.2008.10.002

Sylva, K., Stein, A., Leach, P., Barnes, J., and Malmberg, L. E. (2011). Effects of early child-care on cognition, language, and task-related behaviours at 18 months: an English study. *Br. J. Dev. Psychol.* 29, 18–45. doi: 10.1348/026151010X533229

Turner, I. F. (1974). Cognitive effects of playgroup attendance. *Ir. J. Educ.* 8, 30–35.

van Der Lely, H. K. J., Rosen, S., and McClelland, A. (1998). Evidence for a grammar-specific deficit in children. *Curr. Biol.* 8, 1253–1258. doi: 10.1016/S0960-9822(07)00534-9

Van Noort-Van Der Spek, I. L., Franken, M.-C. J. P., and Weisglas-Kuperus, N. (2012). Language functions in preterm-born children: a systematic review and meta-analysis. *Pediatrics* 129, 745–754. doi: 10.1542/peds.2011-1728

Vernon-Feagans, L., Bratsch-Hines, M., Reynolds, E., and Willoughby, M. (2020). How early maternal language input varies by race and education and predicts later child language. *Child Dev.* 91, 1098–1115. doi: 10.1111/cdev.13281

Wallisch, A., Little, L., Dean, E., and Dunn, W. (2020). Parent concerns: differentiating developmental diagnoses in young children. *Res. Dev. Disabil.* 103:103684. doi: 10.1016/j.ridd.2020.103684

Warnes, G. R., Bolker, B., Lumley, T., Johnson, R. C. (2018). Package 'gmodels': various R programming tools for model fitting for R. Available at: https://CRAN.R-project.org/package=gmodels

Westerlund, M., Berglund, E., and Eriksson, M. (2006). Can severely language delayed 3-year-olds be identified at 18 months? Evaluation of a screening version of the MacArthur-Bates Communicative Development Inventories. *J. Speech Lang. Hear. Res.* 49, 237–247. doi: 10.1044/1092-4388(2006/020)

Winskel, H. (2006). The effects of an early history of otitis media on children's language and literacy skill development. *Br. J. Educ. Psychol.* 76, 727–744. doi: 10.1348/000709905X68312

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 19 June 2023
DOI 10.3389/fpsyg.2023.1110449

Check for updates

OPEN ACCESS

EDITED BY
Maria-José Ezeizabarrena,
University of the Basque Country UPV/EHU, Spain

REVIEWED BY
Miguel Perez-Pereira,
University of Santiago de Compostela, Spain
Miquel Serra,
University of Barcelona, Spain

*CORRESPONDENCE
Alejandra Auza B.
✉ alejandra.auza@yahoo.com

RECEIVED 28 November 2022
ACCEPTED 30 May 2023
PUBLISHED 19 June 2023

CITATION
Auza B. A, Murata C and Peñaloza C (2023) Predictive validity of a parental questionnaire for identifying children with developmental language disorders.
*Front. Psychol.* 14:1110449.
doi: 10.3389/fpsyg.2023.1110449



# Predictive validity of a parental questionnaire for identifying children with developmental language disorders

Alejandra Auza B.[1]*, Chiharu Murata[2] and Christian Peñaloza[3]

[1]Language and Cognition Laboratory, Hospital General Dr. Manuel Gea González, Mexico City, Mexico,
[2]Departamento de Metodología de la Investigación, Instituto Nacional de Pediatría, Mexico City, Mexico,
[3]Departamento de Fonoaudiología, Facultad de Medicina, Universidad de Chile, Santiago, Chile

**Background:** The underdiagnosis of developmental language disorder (DLD) in children is a serious problem in developing countries with limited resources. It has long been noted that the concerns parents have about their children's health and development are richly informative, and if this information can be used for diagnosis, it may provide a means to address the problem of underdiagnosis of DLD. This study aimed to quantify the utility of parental linguistic concern questions (PLCQ) on the identification of language disorders in monolingual Spanish-speaking children in Mexico. It also explored whether a combination of biological and environmental conditions questions (BECQ) might improve the performance of a screening test to identify DLD.

**Methods:** A total of 680 monolingual Mexican Spanish-speaking children and their parents from urban areas in Mexico participated in the study. The distribution of responses to questions about DLD concerns was compared between 185 children diagnosed with DLD and 495 control subjects, and multiple logistic regression analysis was performed to select questions with high predictivity, based on the Akaike information criterion. The diagnostic utility of the questions was assessed by receiver operating characteristic (ROC) curves, stratum-specific likelihood ratios (SSLRs), and changes in pretest and post-test probabilities of DLD. A similar procedure was used to explore whether adding BECQ would improve the diagnostic utility of questions about DLD concerns using data of 128 children.

**Results:** Four questions regarding parental linguistic concerns were found to be useful in identifying children with DLD. When all four concerns were present, the SSLR was 8.79, while it was only 0.27 when there were no concerns at all. The estimates of DLD probability increased from 0.12 to 0.55 at pretest and post-test. On the other hand, the BECQ did not perform as well as the PLCQ in identifying DLD, and the improvement in diagnostic performance it provided was limited to one question.

**Conclusion:** The parental questionnaire can be used as a screening tool to help in identifying children with DLD. The data presented in this study underscore the importance of considering linguistic parental concerns as part of the screening process. This is a realistic option to provide a solution to the current problem of underdiagnosis of DLD in Mexico.

KEYWORDS

**parental questionnaires, developmental language disorder (DLD), early identification, Spanish-speaking children, parental linguistic concerns**

## 1. Introduction

Parental questionnaires are tools that provide access to information about children's language and communication, based on the daily experiences of parents and sometimes teachers of young children. They have emerged as an alternative to the lack of standardized assessment tools in certain sociocultural contexts, such as migrant populations in countries where there are no established reference standards for language development (Restrepo, 1998). Screening for developmental language disorder (DLD) among children in speech and language clinics is a challenge when standardized tests for the monolingual or bilingual population are not available, and even more when there is a lack of knowledge on children's performance in the home language (Abutbul-Oz and Armon-Lotem, 2022).

Parental questionnaires have been used as a supplement to formal assessments since they provide a way to describe children's language skills (Restrepo, 1998). Parental questionnaires also enable professionals to understand the expectations that adults have regarding their children's skills, as well as their overall knowledge about child development. Additionally, they can help parents become more involved in their children's language development (Thal et al., 1999; Bishop and McDonald, 2009; Guiberson et al., 2011). However, this consideration must be examined within the context of Latin American families. While middle-class English speakers may report positive involvement in their children's school and therapeutic activities, it's important to understand how this may differ from the experiences of Latin American families. Moreover, Latin American parents often report not having the same attitude, particularly because they perceive it as an intrusion into the work of professionals and a challenge to their own authority (Rodriguez and Olswang, 2003).

In general, two types of PQ can be identified. The first type focuses on evaluating language development. In these assessments, parents are asked to report on the emergence of a set of communicative and linguistic resources in their children. The MacArthur-Bates Communicative Development Inventory (MB-CDI) (Fenson et al., 1993), Language Developmental Survey (LDS), and the Children's Communication Checklist-2 (CCC-2) (Bishop, 2003) belong to this category. Each one of them has adaptations for Spanish speakers and its use has increased as screening to identify children with language delays or difficulties at an early age. For example, Bishop and McDonald (2009) observed that the combined administration of the CCC-2 with language tests achieved a good degree of specificity, but not of sensitivity when discriminating between children with and without language disorders. However, the use of the CCC-2 always improved the results compared to the discriminating power of language tests alone.

The second type of questionnaire focuses on parental concerns regarding language development, and also collects information from the child's environment. An example is the PQ proposed by Restrepo (1998). She constructed this questionnaire to aid in the assessment of Spanish-speaking children, and when combined with the mean length of utterances (MLU) and the number of grammatical errors per clause, it can provide valuable information. Her questionnaire includes questions related to language milestones, biological, social, and linguistic concerns, which are answered with yes or no. However, there is no information available regarding the selection of the questions or whether the severity of reported difficulties is considered when parents respond to multiple questions.

The use of parental reports is based on the idea that they are equally sensitive to the formal assessment of a professional in the field to evaluate children's communicative and linguistic abilities in different cultural and linguistic contexts (Guiberson and Rodríguez, 2010). However, some authors have pointed out a discrepancy between children who qualify as having language difficulties based on low scores in standardized tests and those who are reported as having difficulties by their parents. Law et al. (2011) found that only a small percentage of children identified as having language difficulties through a series of standardized tests had been detected by teachers in a school setting and referred to language services. Tomblin et al. (1997) observed a similar pattern, noting that only 29% of children identified as having language impairment through formal testing had been previously identified by their parents or school services. This percentage increased only to 39% in the case of children with severe difficulties. Finally, in the study by Bishop and McDonald (2009), more than half of the children who obtained low scores on language tests had not been detected by their parents and teachers. An explanation for this discrepancy may lie in the fact that parental reports are completed without the assistance of a professional. This means that clinicians do not directly interview the parents, but instead parents complete the questionnaires themselves. However, other factors may be contributing to this discrepancy. For instance, parents may be more attentive to certain linguistic domains, such as speech-sound disorders, that are not evaluated in the tests. Additionally, parents and teachers may not be aware of difficulties in specific areas of language that are typically present between the ages of 4 and 16 (Caraveo-Anduaga et al., 2002). Therefore, more severe language problems, such as difficulties in understanding or producing grammar, may not be noticeable to parents or may be masked by other academic or social interaction issues. This raises the question of how parents interpret the questions on the questionnaires and compare them to what they observe in their child's daily life (Bishop and McDonald, 2009). The classification and terminology used to describe developmental language disorder

(DLD), previously referred to as specific language impairment (SLI), can be confusing for professionals and families of children with this condition due to the varying degrees of severity and persistent difficulties in structural language, which can result in functional, social, and educational problems (Serra, 2022).

On the other hand, many undetected children belong to socially vulnerable groups, which may indicate that tests should be sensitive to the social environment, or even certain cultural expectations about children, that could be influencing parents' judgment (Keegstra et al., 2007; Bishop, 2014).

When clear questions are provided, parents can be valuable resources for identifying children with language difficulties or supplementing formal assessments. However, the use of questionnaires, like any assessment tool, must take into account the social and cultural context of the population for which they are intended. The commonly considered predictive values are sensitivity, specificity, negative predictive value (NPV), and positive predictive value (PPV). The area under the curve (AUC), resulting from receiver operating characteristic (ROC) analysis, provides an estimate of the screening or diagnostic test's discriminative power. A recent systematic review analyzed screening tools for language, showing that the predictive validity data from all sample studies demonstrated a mean sensitivity of 77.7% and a PPV of 66.56%. This result indicates that screening tools for language are more effective and even achieve higher sensitivity, specificity, and negative predictive value than direct child assessment for language development (Sim et al., 2019).

Recently, a study was conducted to determine the effectiveness of a two-step procedure for identifying language difficulties in monolingual Mexican Spanish-speaking children (Auza et al., 2023). This procedure combined a grammatical screener with a short parental questionnaire (PQ). The results showed that both the grammatical screener and the PQ were effective in identifying children with developmental language disorder (DLD) between the ages of 4;0 and 6;11 years old. This was indicated by the stratum-specific likelihood ratios (SSLR) of the PQ, as well as the positive and negative likelihood ratios (LR+ and LR−) of a screening test called the "Tamiz de Problemas de Lenguaje" (TPL) (Auza et al., 2018). However, it is important to note that in this study, only eight linguistic concern questions were included in the questionnaire (see **Supplementary Appendix 1**, section I). The post-test probability for detecting children with DLD between the ages of 4;0 and 4;11 and between 5;0 and 5;11 was found to be 57% before administering the grammatical screener, and for children between 6;0 and 6;11, the post-test probability was 68%.

Another recent research sought to identify a set of factors associated with Spanish-speaking children with DLD (Peñaloza, 2018). To achieve this goal, 36 variables related to medical history, language development, and environmental factors were explored. These variables were selected based on a review of 60 articles on language development in Spanish and English. A questionnaire was constructed using these variables and piloted with 60 families in Mexico City and Querétaro, Mexico. In Peñaloza's (2018) study, the researchers investigated the association between environmental variables and the detection of DLD. They found that only eight variables differed statistically between children with and without DLD: these variables were: (1) the sex of the child, (2) the occurrence of motor and/or psychological difficulties during the first years of life, (3) the age of producing the first words, (4) the amount of time the child attended preschool, (5) the years of maternal education, (6) the years of paternal education, (7) the presence of a family history of speech or language problems, and (8) time children attended preschool, and years of maternal education (Auza et al., 2019).

Previous literature suggests that both biological and environmental factors may increase a child's risk of developing language difficulties. For instance, a child's risk may be affected by the quality of social and communicative interactions between the child and their parents or caregivers (Raviv et al., 2004; Bradley and Corwyn, 2005), family socio-educational level (Pelchat et al., 2003; Pan et al., 2004; Bornstein et al., 2007; Cabrera et al., 2007; Farkas et al., 2015), perinatal conditions, birth weight, premature delivery, parental education, environmental factors, sex of the children, and family history with DLD (Stromswold, 2001; Viding et al., 2003; Newbury et al., 2005, 2009; Chaimay et al., 2006; Bishop et al., 2012; Nudel and Newbury, 2013; Moriano-Gutiérrez et al., 2017) and the age the first words were produced (Prathanee et al., 2007). Therefore, based on an instrument that includes both questions about parental language concerns and biological and environmental conditions of the family and home, the primary objective of this study is to answer the following questions: 1. Do the eight questions in the parental linguistic concern questionnaire (PLCQ) help identify children with DLD? 2. Is it possible to further reduce the eight questions in the PLCQ? 3. If so, what is the diagnostic performance of the reduced PLCQ in screening children with DLD? Additionally, we attempted to answer an additional question: 4. Can a combination of biological and environmental questions improve the diagnostic accuracy of the PLCQ in identifying children with DLD? Our hypothesis is that the questionnaire can provide sufficient information and serve as a screening tool to help identify children with DLD. We predict that the parental concern questions, along with some biological and environmental questions, will improve the performance of the questionnaire.

## 2. Materials and methods

### 2.1. Participants

The study recruited a convenience sample of 680 monolingual Mexican Spanish-speaking children, with 240 4-year-olds, 225 5-year-olds, and 215 6-year-olds, from urban and suburban areas in four different locations in Mexico. The recruitment process involved contacting parents through schools and public health centers and inviting them to participate in the study, regardless of whether they were concerned about their child's language development. Parents were asked to sign an informed consent form, and children between 6;0 and 6;11 provided verbal assent to participate. The study was approved by the ethics and research committee of the Hospital General Dr. Manuel Gea González, Mexico City. The parents were also asked to complete the parental linguistic concern questionnaire (PLCQ).

### 2.2. Procedure for diagnosis of DLD

Each child underwent three individual evaluation sessions, each lasting approximately 20 min. During the first evaluation, an auditory screening was conducted, and subtests of the Kaufman Assessment Battery for Children 2 (KABC-II), a cognitive test, were administered (Kaufman and Kaufman, 2004), to rule out hearing or cognitive problems. During the second and third sessions, the linguistic tests and language sample were administered, including the morphosyntax subtest of the bilingual English-Spanish language test (BESA) (Peña et al., 2014); the grammatical subtests of the Spanish Clinical Evaluation of Language Fundamentals—Fourth Edition, Spanish (CELF-4) (Semel et al., 2006); the TPL screening test that identifies grammatical difficulties in Mexican Spanish-speaking children between four and six years of age with grammatical problems (Auza et al., 2018); and a language sample, from a retelling of one of the frog stories (Mayer, 1973, 1974; Mayer and Mayer, 1975), for obtaining the percentage of grammatical errors (percentage of ungrammaticality -PU-) per clause (Restrepo, 1998). According to their manuals, these tests demonstrate very good diagnostic accuracy: the BESA is reported to have 87.5% sensitivity and 100% specificity, and CELF-4 Spanish, has 96% sensitivity and 87% specificity. The TPL is reported to have a sensitivity and specificity of 90 and 83% for 4-year-olds, 90 and 84% for 5-year-olds, and of 94 and 92% for 6-year-olds. Children with typical language development (TLD) met the following criteria: (a) 4-year-old children scored above the cut score of 50 on the morphosyntax subtest of BESA; 5- and 6-year-old children scored within one standard deviation from the mean or above on grammatical subtests of the CELF-4 Spanish because it has higher specificity than BESA for this age range; (b) scored above or on the 16th percentile on the TPL screening test; (c) the PU per clause in the language sample was below 20% (Restrepo, 1998); and (d) the non-verbal IQ score was 80 or above on the KABC-II.

Children with DLD met the following criteria: (a) 4-year-old children scored at or below the cut score of 50 on the morphosyntax subtest of BESA; 5- and 6-year-old children scored below one standard deviation from the mean on the CELF-4 Spanish because it has higher sensitivity than BESA for this age range; (b) scored below the 16th percentile on the TPL screening test; (c) the PU per clause in a language sample was 20% or above; (d) the non-verbal IQ score was 80 or above on the KABC-II; and (e) two native Spanish-speaking speech-language pathologists (SLPs) with more than 15 years of experience confirmed the diagnosis, based on observations at the school setting and during the sessions.

In addition, the children's parents completed a parental questionnaire that asked about their language concerns, medical and language history, educational level of both the parents and child, relatives with a history of language difficulties, as well as the social and cultural activities that parents engaged in daily with their child (see **Supplementary Appendix 1**).

## 2.3. Instruments

The analyzed data were obtained from the administration of the questionnaire:

### 2.3.1. Parental questionnaire (PQ)

The questionnaire consisted of two parts. The first part, known as the parental linguistic concern questionnaire (PLCQ), included eight "Yes" or "No" questions adapted from Restrepo's (1998) original questions. In each question, parents had space to complete their responses with a brief description of what they observed: 1. Are you concerned about the language of your child? 2. Do other people have difficulty understanding your child? 3. Does your child talk as well as other children of the same age? 4. Does your child speak "funny" or "weird?" 5. Has a family member/teacher commented that your child talks little or talks poorly? 6. Does your child understand most of what is said to him/her? 7. Do you have to repeat what you say to your child more than to other children of the same age? 8. Compared to other children of the same age, does your child have difficulties understanding questions? For each question about their perception of their child's language, parents' responses were recorded. If the response indicated parental concern, it was labeled as "risk perception," which included "yes" responses to questions 1, 2, 4, 5, 7, and 8, as well as "no" responses to questions 3 and 6. Otherwise, the response was recorded as "no risk perception." The number of risk perceptions for each question was then calculated cumulatively.

The second part of the questionnaire is called the biological and environmental conditions questionnaire (BECQ). In this part, eleven questions were included: 9. the sex of the child; 10. motor problems; 11. neurological problems; 12. psychological problems; 13. the age of production of first words; 14. years of maternal education; 15. years of paternal education; 16. family history of language problems; 17. time children attended preschool; 18. time dedicated to social interaction with children (e.g., playing with toys, doing puzzles); 19. time dedicated to communicative interaction with children (e.g., reading books, talking about daily experiences), and 20. time spent on screens, which are variables that have been discussed in the literature that may influence DLD (Peñaloza, 2018). A complete Spanish version of the PQ is available in **Supplementary Appendix 1** as well as the short Spanish version in **Supplementary Appendix 2**.

## 2.4. Statistical analysis

The following procedure was used for the statistical analysis.

First, the characteristics of the two groups of children, DLD and TLD (henceforth called children from both clinical conditions), were described according to the age groups of 4-, 5-, and 6-year-olds. Differences in categorical data between both clinical conditions were tested by Pearson's $\chi^2$ test. Fisher's exact test and its extension methods were not used due to the design, in which the frequencies in the contingency table were not fixed (Kroonenberg and Verbeek, 2018). To determine differences in continuous data between groups, we performed Welch's $t$-test, where $P < 0.05$ was interpreted as statistically significant. Non-parametric tests, such as the Mann–Whitney U-test, were not employed as it is recognized that they may produce unreliable $P$-values if the assumption of homoscedasticity is violated (Grimes and Schulz, 2005). We provided 95% confidence intervals for odds ratios and AUC estimates. Effect sizes for continuous variables were calculated using Cohen's d, and effect sizes for categorical variables were expressed as Phi index or odds ratios. The interpretation of Cohen's d and Phi index was as follows: null $< 0.2$; $0.2 \leq$ small $< 0.5$; $0.5 \leq$ medium $< 0.8$; $0.8 \leq$ large for continuous variables, and null $< 0.1$; $0.1 \leq$ small $< 0.3$; $0.3 \leq$ medium $< 0.5$; $0.5 \leq$ large for categorical variables (Cohen, 1988).

The distribution of responses to the eight questions in the PLCQ, the first part of the PQ, was compared between the two clinical conditions. Following this comparison, the independent and dependent variables were interchanged, and multiple logistic regression models were created to identify variables that distinguish between the clinical conditions (Knottnerus et al., 2008). Variable selection for the logistic regression models was conducted using a stepwise method based on the Akaike information criterion (AIC). As indicated in the section "3. Results," statistical analysis using the overall sample of 4–6-year-olds ($n = 680$) allowed us to reduce the number of questions from 8 to 4. The questions thus obtained were used as a four-item PLCQ, a screening tool that yields a score from 0 to 4 according to the number of questions selected by each respondent. The likelihood ratio (LR) is the ratio of the "proportion of true positives" to the "proportion of false positives" (LR+), or the "proportion of false negatives" to the "proportion of true negatives" (LR−). Stratum-specific likelihood ratio (SSLR) was calculated for each of the five strata corresponding to these scores. The weights derived from the LR and SSLR were used to determine the change in pretest and posttest probabilities. The same analysis was conducted for the three age groups. Furthermore, the data of 128 children with no missing responses in the 11 questions of the BECQ, the second part of the PQ, were analyzed to investigate if the biological and environmental information improves the diagnostic performance of the PLCQ. The distribution of the 11 variables was compared between the clinical conditions, and a multiple logistic regression model was constructed to identify questions useful in identifying children with DLD, by interchanging independent and dependent variables. The utility of a diagnostic test can be evaluated by how much the results of the test change the probability of a particular condition expected in an individual, such as the presence or absence of a condition or a property aimed at diagnosing a condition. The LR approach to diagnostic test utility studies uses LR and SSLR as algebraic weighting factors based on Bayes' theorem to update pretest probabilities into posttest probabilities. Likelihood ratios are typically interpreted as follows: an LR of 10 or greater (or its inverse less than 0.1) indicates a large change between pretest and posttest probabilities; an LR of 5–10 (or its inverse 0.2–0.1) indicates a moderate change between pretest and posttest probabilities; an LR of 2–5 (or its inverse 0.5–0.2) indicates a small change between pretest and posttest probabilities; an LR less than 2 (or its reciprocal 0.5 or greater) indicates no change between pretest and posttest probabilities.

Statistical analysis was performed using the free software environment r (R Core Team, 2021); r was used with RStudio (R Studio Team, 2020) and the effsize package (Torchiano, 2020).

## 3. Results

Table 1 summarizes the sociodemographic and developmental characteristics of children with DLD and TLD across three age groups (4-, 5-, and 6-year-olds) and compares them. The distribution of sex and age was similar between both clinical conditions across all age groups, with almost zero effect sizes. Differences in maternal education were found between DLD and TLD groups in the 5- and 6-year-old age groups, with statistically significant or nearly significant $p$-values, but effect sizes were small for both. Statistically significant differences were observed between clinical conditions in the KABC-II scores in the 4- and 6-year-old age groups, with small and medium effect sizes, respectively. However, these scores were within the normal range for both clinical conditions, and their standard deviations were expected. In contrast, large differences were observed in language parameters measured by BESA (age 4) or CELF-4 (ages 5 and 6) and the PU across all age groups, as expected.

To determine whether the eight questions in the PLCQ are effective in identifying children with DLD, we compared the proportion of parental concern by examining the number of positive responses to each question. As presented in **Table 2**, parents of children with DLD expressed a higher percentage of concern for all questions, although the effect size varied depending on the question. Questions 1, 2, and 5 showed a "medium" effect size, while questions 3, 4, 7, and 8 had a "small" effect size. Question 6, on the other hand, had a null effect size.

Next, a multiple logistic regression model was constructed with the eight questions in the PLCQ as explanatory variables, and clinical conditions (DLD/TLD) as the criterion variable, to identify the questions that contribute to the prediction of DLD. **Table 3** shows four questions (1, 2, 5, and 8) selected by the best fit logistic regression model (for groups 4- to 6-year-olds, see **Supplementary Appendix 3**).

In addition, ROC curves were generated based on the same multiple logistic regression model. In this, the area under the curve (AUC) was 0.795 (95% CI: 0.753, 0.831). By age, the AUC was 0.737 (95% CI: 0.664, 0.799); 0.881 (95% CI: 0.828, 0.919); 0.852 (95% CI: 0.774, 0.906) for the 4, 5 and 6-year-old samples, respectively, all indicating satisfactory diagnostic performance (See **Figure 1**).

**Table 4** shows the distribution of absolute frequencies of the number of positive responses of parents of DLD and TLD children to the four questions in the PLCQ for all 4- to 6-year-olds ($n = 680$). The SSLR was obtained, as well as the change in the pretest to posttest DLD probability. Here we used 0.12, an estimate of the prevalence of DLD among 4- to 6-year-olds in Mexico obtained in a study in preparation (Auza et al., 2018). This estimate was used as the pretest probability of administering the PLCQ; the posttest probability with SSLR was calculated as a weight factor. The results showed that if a parent reported two out of four concerns, the posttest probability was almost the same as the pretest probability. However, if the parent reported three out of four concerns, the probability of the child having DLD almost tripled, increasing from 0.12 to 0.35. If all four concerns were present, the posttest probability increased more than 4.5 times, from 0.12 to 0.55. On the other hand, the pretest probability of DLD decreased three times to 0.09 when only one concern was reported, and one-third to 0.04 when no concerns were reported. The above analysis was also conducted for each age group of 4-, 5-, and 6-year-olds, with similar results among them (Table 6 in **Supplementary Appendix 3**).

Data from the BECQ questionnaire were available for 128 cases, representing 19% of the overall sample of 4- to 6-year-olds, 36 (15%) for the 4-year-old group, 45 (20%) for the 5-year-old group, and 47 (22%) for the 6-year-old group. Based on the small sample size and unstable results within each age group, we have presented the overall results from the 128 children. However, for additional information, results by age groups can be found in Table 7 in **Supplementary Appendix 3**. However, the results of the age-specific analysis are also included in the same table for reference. In both clinical conditions, the distribution of eight questions in the PLCQ in the subgroups was similar to the results obtained from the overall sample. *P*-values were larger in the analysis performed in subgroups, which can be expected *a priori*, because of the reduced power in the statistical tests when reducing the sample size. On the other hand, effect sizes for the eight questions showed a similar pattern between the overall sample and the subgroup of 128 children.

In a multiple logistic regression model for assessing the predictive accuracy of DLD using PLCQ, the following questions were selected: 1, 2, 5, and 8. These questions were found to contribute to the model that best meets the AIC criteria. Its performance was 0.804 (95% confidence interval: 0.683, 0.886) on the AUC-ROC curve. On the other hand, when questions 1, 2, 5, and 8, which comprise the four questions in the PLCQ obtained from the overall sample of 680 subjects, were administered to the same subgroup, the DLD diagnostic performance was 0.808 (95% confidence interval: 0.690, 0.888) on the AUC-ROC curve, almost equal to the best-fit model (Table 8 in **Supplementary Appendix 3** and **Figure 1**). Based on this equivalence, the subgroup of 128 participants was explored to observe whether the addition of the BECQ questions to the PLCQ might improve the performance of the parental questionnaire. To evaluate its performance, first,

TABLE 1 Demographic, sociocultural, and clinical characteristics of children with DLD and TLD in each of the three age groups.

| Variable | DLD | TLD | ES | *P*-value |
|---|---|---|---|---|
| 4-year-old [*n* = 240; DLD: *n* = 84 (35%), TLD: *n* = 156 (65%)] | | | | |
| Demographic and sociocultural | | | | |
| Age (month) | 53.0 (3.7) | 53.2 (3.2) | $0.06^{N}$ | 0.200 |
| Sex (female) | 35 (42%) | 77 (49%) | $0.07^{N}$ | 0.255 |
| Maternal education (years) ($n_{DLD}$ = 80; $n_{TLD}$ = 150) | 11.7 (3.8) | 12.3 (4.1) | $0.15^{N}$ | 0.235 |
| Cognitive and language development | | | | |
| KABC-II (score) ($n_{DLD}$ = 77; $n_{TLD}$ = 123) | 100.3 (9.6) | 104.3 (10.6) | $0.39^{S}$ | 0.007 |
| BESA (%) | 50.8 (19.0) | 83.8 (10.1) | $2.51^{L}$ | < 0.001 |
| PU (%) | 35.2 (22.5) | 13.6 (13.0) | $1.37^{L}$ | < 0.001 |
| 5-year-old [*n* = 225; DLD: *n* = 59 (26%), TLD: *n* = 166 (74%)] | | | | |
| Demographic and sociocultural | | | | |
| Age (month) | 65.4 (3.3) | 65.7 (3.5) | $0.09^{N}$ | 0.587 |
| Sex (female) | 21 (36%) | 67 (40%) | $0.04^{N}$ | 0.519 |
| Maternal education (years) ($n_{DLD}$ = 56; $n_{TLD}$ = 162) | 10.9 (3.6) | 12.4 (4.0) | $0.38^{S}$ | 0.012 |
| Cognitive and language development | | | | |
| KABC-II (score) ($n_{DLD}$ = 55; $n_{TLD}$ = 124) | 100.3 (9.6) | 101.3 (11.0) | $0.09^{N}$ | 0.520 |
| CELF-SR (score: 1–18) | 8.6 (2.9) | 13.1 (2.2) | $1.91^{L}$ | < 0.001 |
| CELF-WS (score: 1–18) | 7.6 (2.9) | 12.7 (2.6) | $1.92^{L}$ | < 0.001 |
| PU (%) | 41.3 (22.2) | 12.8 (9.8) | $2.24^{L}$ | < 0.001 |
| 6-year-old [*n* = 215; DLD: *n* = 42 (19%), TLD: *n* = 173 (80%)] | | | | |
| Demographic and sociocultural | | | | |
| Age (month) | 77.0 (3.4) | 77.5 (3.8) | $0.13^{N}$ | 0.459 |
| Sex (female) | 13 (31%) | 74 (43%) | $0.10^{N}$ | 0.161 |
| Maternal education (years) ($n_{DLD}$ = 41; $n_{TLD}$ = 164) | 9.5 (4.1) | 10.9 (3.8) | $0.36^{S}$ | 0.054 |
| Cognitive and language development | | | | |
| KABC-II (score) ($n_{DLD}$ = 42; $n_{TLD}$ = 125) | 98.4 (10.1) | 104.5 (10.7) | $0.58^{M}$ | 0.001 |
| CELF-SR (score: 1–18) | 7.4 (2.8) | 12.6 (2.3) | $2.11^{L}$ | < 0.001 |
| CELF-WS (score: 1–18) | 6.7 (3.0) | 12.2 (2.2) | $2.30^{L}$ | < 0.001 |
| PU (%) | 44.0 (24.4) | 12.7 (9.6) | $2.29^{L}$ | < 0.001 |

Data were summarized by mean (standard deviation), except the variable "sex" reported in the number of cases (percentage). ES, effect size.

it was examined whether there were significant differences in the distribution between the clinical groups for each of the eleven questions in BECQ, as we did for the eight questions in the PLCQ. We found moderate or large effect sizes for only two questions, such as 17. time children attended preschool, and 19. communicative interaction with children. Statistically significant differences between DLD and TLD groups were detected for questions 10. neurological and/or psychological problems, 17, and 19. The effect sizes and *P*-values were small (Phi = 0.20), $P = 0.023$ for question 2; large (Cohen's $d = 0.20$), $P < 0.001$ for question 17; and medium (Cohen's $d = 0.54$), $P = 0.023$, for question 19 (Table 9 in **Supplementary Appendix 3**). Second, a multiple logistic regression model was constructed with the BECQ questions as explanatory variables and DLD (or TLD) as the criterion variable, and the best-fit model was searched for, following the AIC criteria; questions 10, 16, 17, and 19 were selected as explanatory variables. However, in this model, the association of questions 10, 16, and 19 with the criterion variables was very low compared to the association shown in question 17. Given so, a logistic regression model with only question 17 as the explanatory variable was created. Then, ROC curves were constructed based on these models. As a result, ROC-AUC was 0.767 (95% CI: 0.634, 0.863) in the best-fit model with four variables. In the single-variable model with only question 17, the result was 0.753 (95% CI: 0.640, 0.840), indicating that both showed almost equal performance. Therefore, based on the parsimony principle, only question 17. time children attended preschool was used as an explanatory variable in the BECQ (Table 10 in **Supplementary Appendix 3**).

Finally, the predicted probabilities of DLD were obtained as shown in Table 5. Very few cases in this subgroup had a PLCQ score of 4 (2 in the group of DLD and 3 in the group of TLD). Therefore, this stratum of 4 was combined with that of score 3. As a result, SSLRs of 0.653, 0.166, 0.079, and 0.019 were calculated for the score 3–4, score 2, score 1, and score 0 strata, respectively. Regarding the BECQ questions, the only explanatory variable was 17. time children attended preschool. The variable showed a likelihood ratio of 1.829 below the median and 0.525 above the median. When setting the pretest probability of the PLCQ at 0.13, which represents the estimated prevalence of DLD among children aged 4–6 in Mexico, the posttest probability of the PLCQ varied from 0.37 to 0.02, depending on the PLCQ score. Furthermore, the posttest probability of the PLCQ was higher than that of the BECQ. When weighted by likelihood ratios, as pretest probabilities (i.e., the binary preschool enrollment years), the posttest probabilities were 0.52 or 0.24, 0.22 or 0.07, 0.12 or 0.04, and 0.03 or 0.01 for BECQ pretest probabilities of 0.37, 0.13, 0.07, and 0.02, respectively.

## 4. Discussion

The purpose of this paper was to analyze whether a parental language concern questionnaire can help identify children with developmental language disorder. Our hypothesis was confirmed, as the questionnaire can provide sufficient information and be used as a screening tool to help identify children with DLD. The data presented in this study emphasize the importance of considering parental linguistic concerns as part of a screening process. As previously stated, parents can be valuable allies in obtaining reliable information about children, as they are attuned to their communicative and linguistic needs and difficulties, regardless of their cultural and linguistic backgrounds (Thal et al., 2000; Bishop and McDonald, 2009; Guiberson and Rodríguez, 2010; Peñaloza et al., 2021). This is particularly significant because the absence of suitable screening and assessment tools for certain underrepresented populations has resulted in inaccurate under-diagnosis of children with DLD. Regarding more qualitative information on which questions in the PLCQ have the best predictive value for identifying children with DLD, our hypothesis was also confirmed. We observed that specific questions in the PLCQ better improved the performance of this questionnaire. Some questions, such as your child speaks as well as other children of the same age? and Do you have to repeat a question to your child several times in order for him/her to understand it? were removed from the linguistic parental concern set of questions since they did not contribute to identifying children with DLD. An interesting outcome is that by using a smaller number of more sensitive screening questions, many children at risk of having DLD can be identified, even without administering any tests to them yet. Our previous study (Auza et al., 2023) showed that a combination of a parental questionnaire and a screening test could satisfactorily identify children with DLD. Administering these questions to families was useful as a screening test for

TABLE 2 Eight questions in the PLCQ in both clinical conditions.

| Questions in PLCQ (*n* = 680) | DLD | TLD | ES | *P*-value |
|---|---|---|---|---|
| 1. Are you concerned about the way your child talks? | 139 (75%) | 183 (37%) | $0.34^{M}$ | < 0.001 |
| 2. Do other people have difficulty understanding your child? | 120 (65%) | 94 (19%) | $0.44^{M}$ | < 0.001 |
| 3. Does your child talk as well as other children of the same age? | 105 (57%) | 180 (36%) | $0.18^{S}$ | < 0.001 |
| 4. Does your child speak "funny" or "weird?" | 90 (49%) | 101 (20%) | $0.28^{S}$ | < 0.001 |
| 5. Has a family member/teacher commented that your child talks little or talks poorly? | 118 (64%) | 92 (19%) | $0.44^{M}$ | < 0.001 |
| 6. Does your child understand most of what is said to him/her? | 69 (37%) | 147 (30%) | $0.07^{N}$ | 0.058 |
| 7. Do you have to repeat what you say to your child more than to other children of the same age? | 75 (41%) | 80 (16%) | $0.26^{S}$ | < 0.001 |
| 8. Compared to other children of the same age, does your child have difficulties understanding questions? | 61 (33%) | 47 (9%) | $0.29^{S}$ | < 0.001 |

Developmental language disorder$_{DLD}$ and typical language development$_{TLD}$: sample size of both clinical groups, respectively. Data were summarized by the number of cases (percentage). ES, effect size.

TABLE 3 The best fit logistic regression model with four questions in the parental questionnaire.

| Model terms | β (SE) | $\chi^2$ | *P*-value | OR (95%CI) |
|---|---|---|---|---|
| Global: 4 to 6-year-old (*n* = 680) | | | | |
| Best fit model: χ2 (d.f. 4) = 172.43, *P* < 0.001, AIC = 633.63, AUC = 0.795 | | | | |
| Intercept | −2.26 (0.17) | 172.57 | < 0.001 | |
| 1. Are you concerned about the way your child talks? | 0.48 (0.25) | 3.66 | 0.056 | 1.61 (0.99, 2.62) |
| 2. Do other people have difficulty understanding your child? | 1.06 (0.25) | 18.16 | < 0.001 | 2.90 (1.78, 4.73) |
| 5. Has a family member/teacher commented that your child talks little or talks poorly? | 1.17 (0.23) | 25.63 | < 0.001 | 3.21 (2.04, 5.05) |
| 8. Compared to other children of the same age, does your child have difficulties understanding questions? | 0.64 (0.26) | 6.20 | 0.013 | 1.90 (1.15, 3.14) |

β (SE): regression coefficient (standard error). χ2, Wald's χ2; OR (95% CI), odds ratio (95% confidence interval).

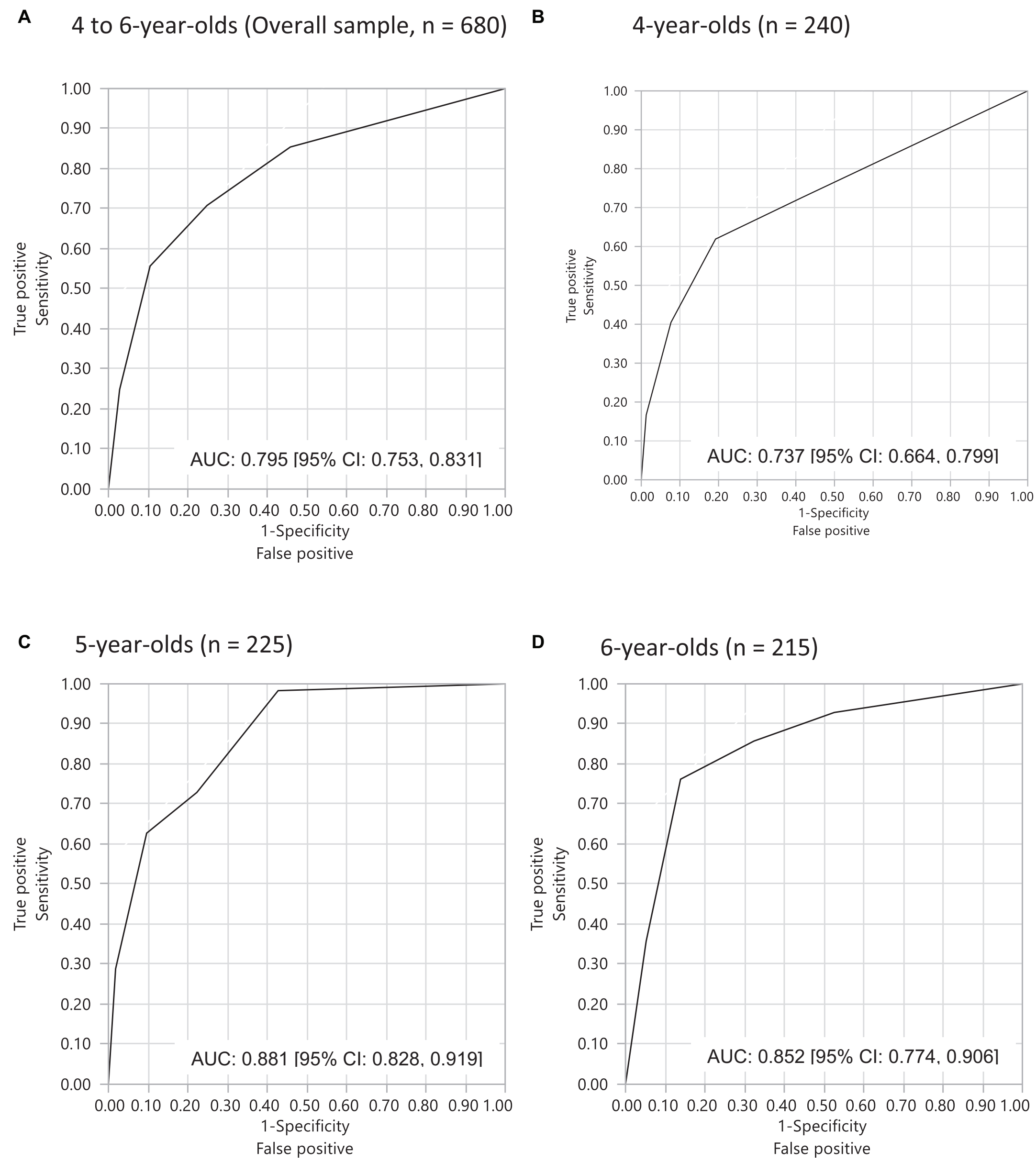


FIGURE 1
ROC (receiver operating characteristic) curves showing DLD diagnostic accuracy of the PLCQ based on four questions: **(A)** 4- to 6-years-olds (overall sample, *n* = 680); **(B)** 4-year-olds (*n* = 240); **(C)** 5-year-olds (*n* = 225); **(D)** 6-year-olds (*n* = 215).

identifying DLD in 4-, 5-, and 6-year-olds, as evidenced by the SSLR of the questionnaire. Additionally, our study demonstrated which questions were more helpful and effective in identification. Questions such as Does your child talk as well as other children of the same age? may not be accurate since parents may not always be aware of linguistic developmental milestones and therefore do not have a point of comparison. Similarly, if we ask them Does your child speak “funny” or “weird?” Parents may not always consider that speaking differently, strangely, or funny does not necessarily imply a language problem. This may be due to their limited knowledge of morphosyntactic milestones. They may not be aware of when children start to combine words (emergence of syntax), when they usually start using functional words, and so on. Therefore, parental linguistic concerns during these ages may also be focused on other aspects beyond grammar. Other questions about comparison with other children, such as Do you have to repeat what you say to your child more than to other children of the same age? may be related to parents’ interpretation of their child’s behavior. Parents often understood this question as being distracted or disobedient to commands rather than having difficulty understanding language. Furthermore, statistical analyses indicated that four questions about parental linguistic concern are sufficient as a screening tool, especially in contexts where language pathologists and/or language tests may not be readily available in large clinical or educational settings. When three or more concern questions are obtained in the questionnaire, parents should be encouraged to seek evaluation by a clinician to confirm the diagnosis, as the probability of having DLD increases threefold with these concern questions. Conversely, if there is only one concern or none, it is acceptable to lower suspicions about having DLD.

Our last hypothesis was partially confirmed. We expected that including more biological and environmental questions in the questionnaire would improve its performance, but our results showed only a slight improvement with the inclusion of one question: time children attended preschool. A prolonged and consistent stay in preschool may work as a critical mass in providing children with linguistic tools, such as greater grammatical complexity, as they are exposed to more interactions and varied discursive practices. However, our results only showed a slight improvement in performance with the inclusion of one additional question. Therefore, our study may not provide conclusive evidence on whether other biological or environmental factors, such as the age of onset of first words and family heritability, may also play a significant role, as suggested by several previous studies (Pelchat et al., 2003; Raviv et al., 2004; Bornstein et al., 2007; Farkas et al., 2015). Our study has demonstrated that communicative interaction between parents and their children is a significant factor at the ages of

**TABLE 4** Stratum-specific likelihood ratio (SSLR) and the change from pretest probability to posttest probability of DLD in the three age groups with four questions in PLCQ.

| Stratum | DLD | TLD | SSLR | Pretest probability* | Pretest odds | Posttest odds | Posttest probability |
|---|---|---|---|---|---|---|---|
| Global, 4 to 6-year-old (*n* = 680) | | | | | | | |
| 4 | 46 | 14 | 8.792 | 0.12 | 0.136 | 1.199 | 0.55 |
| 3 | 57 | 38 | 4.014 | | | 0.547 | 0.35 |
| 2 | 28 | 71 | 1.055 | | | 0.144 | 0.13 |
| 1 | 27 | 104 | 0.695 | | | 0.095 | 0.09 |
| 0 | 27 | 268 | 0.270 | | | 0.037 | 0.04 |
| Total | 185 | 495 | | | | | |

SSLR, stratum-specific likelihood ratio.
*For the pretest probabilities, we used estimates of the prevalence of DLD in Mexican children aged 4 to 6 years obtained from another study [Auza, A., Murata, C., and Méndez, I. (under review). Prevalence of developmental language disorders in Mexico. *Semin. Speech Lang.*].

**TABLE 5** DLD probabilities at the pretest and two posttests combined with four questions in the PLCQ and (time children attended preschool) data.

| Prob1 | Odds 1 | PLCQ4 | DLD (*n* = 21) | TLD (*n* = 41) | SSLR | Odds 2 | Prob 2 | Preschool > median | LR | Odds 3 | Prob 3 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.12 | 0.136 | 3–4 | 12 | 14 | 4.367 | 0.596 | 0.37 | No | 1.829 | 1.194 | 0.52 |
| | | 2 | 5 | 23 | 1.108 | 0.151 | 0.13 | No | 1.829 | 0.303 | 0.22 |
| | | 1 | 3 | 29 | 0.527 | 0.072 | 0.07 | No | 1.829 | 0.144 | 0.12 |
| | | 0 | 1 | 41 | 0.124 | 0.017 | 0.02 | No | 1.829 | 0.034 | 0.03 |
| | | | | | | | | Yes | 0.525 | 0.342 | 0.24 |
| | | | | | | | | Yes | 0.525 | 0.087 | 0.07 |
| | | | | | | | | Yes | 0.525 | 0.041 | 0.04 |
| | | | | | | | | Yes | 0.525 | 0.010 | 0.01 |

Prob 1: estimated prevalence of DLD in Mexico used as the pretest probability for DLD prior to the application of four questions in the PLCQ. Odds 1: pretest odds calculated from pretest probability. PLCQ4: 4 stratums of SSLR based on the results of the four questions in the PLCQ. Odds 2: posttest odds for DLD according to the result of four questions on PLCQ weighted by the SSLR. Prob 2: posttest probability calculated from the posttest odds. Preschool: (time children attended preschool) dichotomized into ≤ and > the median within age groups (for 4-, 5-, and 6-year-olds, 1.5, 2.5, and 3.5 years, respectively). LR: likelihood ratios for children ≤ median or > median for (time children attended preschool). Odds 3: posttest odds for DLD according to (time children attended preschool) weighted by the LR. Prob 3: posttest probability calculated after the combination of 4 questions on PLCQ and (time children attended preschool).

four and six. We have observed medium and large effect sizes, respectively, which suggest that a lower level of communicative interaction may be associated with the diagnosis of DLD. Surprisingly, we have not found any association between low maternal education or age of onset of first words and DLD, possibly due to the limited statistical power in the small subgroup analyzed. However, it is worth noting that the existence of associations does not necessarily imply a predictive value in the parental questionnaire. For instance, previous research has shown that maternal education is linked to DLD in several studies (Hart and Risley, 1995; Hoff, 2003). This association may have an impact on language development, particularly on vocabulary acquisition, but its influence on morphosyntax may be marginal (Abutbul-Oz and Armon-Lotem, 2022). Although most of the biological and environmental questions did not improve the performance of the parental questionnaire, it is still important to identify which variables showed an association and which ones may become associated with a larger sample in future studies. For instance, time spent on screens did not demonstrate a significant association in our analysis, but it had a small effect in the general analysis of the 128 children and a large effect in the 6-year-old group. This suggests that, by increasing the statistical power of our data in a future study, this variable may contribute to predicting which children are at risk of developing DLD. Other variables related to social interaction are also worth considering in future research, even in younger children. For example, it would be interesting to explore the hours parents spend playing with their children, the frequency of doing homework together, and the frequency of shared playtime.

Although several studies have reported associations between various variables and language disorders, few have examined whether these associations improve the predictive validity of parental questionnaires. Our study has contributed to this issue by analyzing some factors that do not enhance the performance of the parental questionnaire, despite being known to influence language disorders. Therefore, we can conclude that using a reduced number of sensitive parental linguistic concern questions is a reliable method for identifying children with language disorders, particularly in settings where standardized assessment instruments or special education services are scarce. Future research could explore the identification of younger children who are at risk of developing DLD, including work with 3-year-olds. Additionally, we believe that these questions are worth exploring in other cultures facing similar issues, in order to generalize our results.

## 5. Limitations of the study

Our study has some limitations that should be noted. Firstly, our findings about the usefulness of some of the biological and environmental questions are inconclusive, as they might not have been clear to some parents, resulting in incomplete answers. Hence, other biological and environmental factors could potentially play a significant role in identifying children with DLD, when answers are not omitted. With larger sample sizes, we may be able to identify additional factors that contribute to the identification of more children with DLD. To address this issue, future studies should ensure that parents answer all questions in the questionnaire or provide improved explanations on how to complete it. In an ideal situation, a clinician should administer the questionnaire to parents, ensuring that all questions are answered accurately. Furthermore, it is important to acknowledge that parents, especially those with lower levels of education, may tend to underestimate certain aspects of their child's language development, such as grammar use (Caraveo-Anduaga et al., 2002; Keegstra et al., 2007; McLeod and Harrison, 2009). Additionally, it is worth noting that even when there is a family history of language impairments, parents may tend to downplay its significance. Additionally, many parents from our sample report short linguistic and social interactions with their children, that may not be sufficient for optimal language development as reported previously in the literature.

Despite the reduced sample size, our study provides valuable insights for identifying monolingual children with DLD using parental linguistic concern questions.

## Data availability statement

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

## Ethics statement

The studies involving human participants were reviewed and approved by the Comité de Ética en Investigación, Hospital General Dr. Manuel Gea González. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

## Author contributions

AAB, CM, and CP contributed equally to the design and implementation of the research, to the analysis of the results, and to the writing of the manuscript. AAB and CP were in charge of data collection and coding. CM supervised, supported, and processed statistical analysis. AAB guided this research project from beginning to end. All authors contributed to the article and approved the submitted version.

## Funding

The research was funded by the Ciencia Básica Program 2013, CONACyT, contract no. 0220634.

## Acknowledgments

We thank Anne Yrigoyen for the English editing of the manuscript.

## Conflict of interest

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

## Editor's note

Maria-José Ezeizabarrena edited the article in collaboration with Melita Kovacevic, University of Zagreb, Zagreb, Croatia.

## Publisher's note



## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2023.1110449/full#supplementary-material

## References

Abutbul-Oz, H., and Armon-Lotem, S. (2022). Parent questionnaires in screening for developmental language disorder among bilingual children in speech and language clinics. *Front. Educ.* 7:846111. doi: 10.3389/feduc.2022.846111

Auza, A., Murata, C., Márquez, M. E., and Morgan, G. (2018). *Tamiz de problemas de lenguaje TPL*. Ciudad de México: Manual Moderno.

Auza, A., Peñaloza, C., and Murata, C. (2019). "The influence of maternal education on the linguistic abilities of monolingual Spanish-Speaking children with and without Specific Language Impairment," in *Atypical language development in Romance languages*, eds E. Aguilar-Mediavilla, L. Buil-Legaz, R. López-Penadés, V. A. Sanchez-Azanza, D. Adrover-Roig, et al. (Amsterdam: John Benjamins Publishing Company).

Auza, B. A., Murata, C., and Peñaloza, C. (2023). Early detection of Spanish-speaking children with developmental language disorders: concurrent validity of a short questionnaire and a screening test. *J. Commun. Disord.* 104:106339. doi: 10.1016/j.jcomdis.2023.106339

Bishop, D. (2014). Ten questions about terminology for children with unexplained language problems. *Int. J. Lang. Commun. Disord.* 49, 381–415. doi: 10.1111/1460-6984.12101

Bishop, D. V. M. (2003). *The children's communication checklist - 2*. London: Psychological Corporation.

Bishop, D., and McDonald, D. (2009). Identifying language impairment in children: combining language test scores with parental report. *Int. J. Lang. Commun. Disord.* 44, 600–615. doi: 10.1080/13682820802259662

Bishop, D., Holt, G., Line, E., McDonald, D., McDonald, S., and Watt, H. (2012). Parental phonological memory contributes to prediction of outcome of late talkers from 20 months to 4 years: a longitudinal study of precursors of specific language impairment. *J. Neurodev. Disord.* 4:3. doi: 10.1186/1866-1955-4-3

Bornstein, M., Hendricks, C., Haynes, O., and Painter, K. (2007). Maternal sensitivity and child responsiveness: associations with social context, maternal characteristics, and child characteristics in a multivariate analysis. *Infancy* 12, 189–223. doi: 10.1111/j.1532-7078.2007.tb00240.x

Bradley, R., and Corwyn, R. (2005). Caring for children around the world: a view from HOME. *Int. J. Behav. Dev.* 29, 468–478. doi: 10.1177/01650250500146925

Cabrera, N., Shannon, J., and Tamis-LeMonda, C. (2007). Father's influence on their children's cognitive and emotional development: from toddlers to pre-k. *Appl. Dev. Sci.* 11, 208–213. doi: 10.1080/10888690701762100

Caraveo-Anduaga, J., Colmenares-Bermúdez, E., and Martínez-Vélez, N. (2002). [Mental symptoms perceptions of healthcare needs, and health care seeking behaviors, among children and adolescents in Mexico City]. *Salud Publica Mex.* 44, 492–498.

Chaimay, B., Thinkhamrop, B., and Thinkhamrop, J. (2006). Risk factors associated with language development problems in childhood–a literature review. *J. Med. Assoc. Thai.* 89, 1080–1086.

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

Farkas, C., Carvacho, C., Galleguillos, F., Montoya, F., León, F., Santelices, M. P., et al. (2015). Estudio comparativo *de* la sensibilidad entre madres y personal educativo en interacción con niños y niñas de un año de edad. *Perfiles Educ.* 37, 16–33.

Fenson, L., Dale, P., Reznick, J. S., Thal, D., Bates, E., Hartung, J., et al. (1993). *Technical manual for the Macarthur communicative development inventory*. San Diego, CA: San Diego State University.

Grimes, D., and Schulz, K. (2005). Refining clinical diagnosis with likelihood ratios. *Lancet* 365, 1500–1505. doi: 10.1016/S0140-6736(05)66422-7

Guiberson, M., and Rodríguez, B. (2010). Measurement properties and classification accuracy of two Spanish parent surveys of language development for preschool-age children. *Am. J. Speech Lang. Pathol.* 19, 225–237. doi: 10.1044/1058-0360(2010/09-0058

Guiberson, M., Rodríguez, B., and Dale, P. (2011). Classification accuracy of brief parent report measures of language development in Spanish-speaking toddlers. *Lang. Speech Hear. Serv. Sch.* 42, 536–549. doi: 10.1044/0161-1461(2011/10-0076

Hart, B., and Risley, T. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore, MD: Paul H. Brookes.

Hoff, E. (2003). The specificity of environmental influence: socioeconomic status affects early vocabulary development via maternal speech. *Child Dev.* 74, 1368–1378. doi: 10.1111/1467-8624.00612

Kaufman, A. S., and Kaufman, N. L. (2004). *Kaufman assessment battery for children*, Second Edn. Circle Pines, MN: American Guidance Service.

Keegstra, A., Knijff, W., Post, W., and Goorhuis-Brouwer, S. (2007). Children with language problems in a speech and hearing clinic: background variables and extent of language problems. *Int. J. Pediatr. Otorhinolaryngol.* 71, 815–821. doi: 10.1016/j.ijporl.2007.02.001

Knottnerus, J. A., Buntinx, F., and van Weel, C. (2008). "General introduction: evaluation of diagnostic procedures," in *The evidence base of clinical diagnosis*, eds J. A. Knottnerus and F. Buntinx (Oxford: Wiley-Blackwell), 1–19. doi: 10.1002/9781444300574.ch1

Kroonenberg, P. M., and Verbeek, A. (2018). The tale of Cochran's rule: my contingency table has so many expected values smaller than 5, what am i to do?. *Am. Stat.* 72, 175–183. doi: 10.1080/00031305.2017.1286260

Law, J., McBean, K., and Rush, R. (2011). Communication skills in a population of primary school-aged children raised in an area of pronounced social disadvantage. *Int. J. Lang. Commun. Disord.* 46, 657–664. doi: 10.1111/j.1460-6984.2011.00036.x

Mayer, M. (1973). *Frog on his own*. New York, NY: Penguin Young Readers Group.

Mayer, M. (1974). *Frog goes to dinner*. New York, NY: Penguin Young Readers Group.

Mayer, M., and Mayer, M. (1975). *One frog too many*. New York, NY: Penguin Young Readers Group.

McLeod, S., and Harrison, L. J. (2009). Epidemiology of speech and language impairment in a nationally representative sample of 4- to 5-year-old children. *J. Speech Lang. Hear. Res.* 52, 1213–1229. doi: 10.1044/1092-4388(2009/08-0085)

Moriano-Gutiérrez, A., Colomer-Revuelta, J., Sanjuan, J., and Carot-Sierra, J. (2017). [Environmental and genetic variables related with alterations in language acquisition in early childhood]. *Rev. Neurol.* 64, 31–37.

Newbury, D., Bishop, D., and Monaco, A. (2005). Genetic influences on language impairment and phonological short-term memory. *Trends Cogn. Sci.* 9, 528–534. doi: 10.1016/j.tics.2005.09.002

Newbury, D., Winchester, L., Addis, L., Paracchini, S., Buckingham, L., Clark, A., et al. (2009). CMIP and ATP2C2 modulate phonological short-term memory in language impairment. *Am. J. Hum. Genet.* 85, 264–272. doi: 10.1016/j.ajhg.2009.07.004

Nudel, R., and Newbury, D. (2013). FOXP2. *Wiley Interdiscip. Rev. Cogn. Sci.* 4, 547–560. doi: 10.1002/wcs.1247

Pan, B. A., Rowe, M. L., Spier, E., and Tamis-LeMonda, C. (2004). Measuring productive vocabulary of toddlers in low-income families: concurrent and predictive validity of three sources of data. *J. Child Lang.* 31, 587–608. doi: 10.1017/s0305000904006270

Pelchat, D., Bisson, J., Bois, C., and Saucier, J.-F. (2003). The effects of early relational antecedents and other factors on parental sensitivity of mothers and fathers. *Infant Child Dev.* 12, 27–51. doi: 10.1002/icd.335

Peña, E. D., Gutierrez-Clellen, V., Iglesias, A., Goldstein, B., and Bedore, L. M. (2014). *Bilingual English-Spanish assessment manual*. San Rafael, CA: AR-Clinical Publications.

Peñaloza, C. (2018). *Habilidades lingüístico-discursivas y características familiares y socioambientales de niños preescolares de Ciudad de México [tesis doctoral]*. Mexico: Universidad Nacional Autónoma de México.

Peñaloza, C., Auza, A., and Murata, C. (2021). "Parental concern in typical and atypical language acquisition of Monolingual Spanish-speaking children in adverse social conditions," in *Child and adolescent development in risky adverse contexts: a Latin American perspective*, ed. B. Barcelata (Berlin: Springer).

Prathanee, B., Thinkhamrop, B., and Dechongkit, S. (2007). Factors associated with specific language impairment and later language development during early life: a literature review. *Clin. Pediatr.* 46, 22–29. doi: 10.1177/0009922806297153

R Core Team (2021). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

R Studio Team (2020). *RStudio: integrated development for R*. Boston, MA: RStudio Inc.

Raviv, T., Kessenich, M., and Morrison, F. J. (2004). A mediational model of the association between socioeconomic status and three-year-old language abilities: the role of parenting factors. *Early Child. Res. Q.* 19, 528–547. doi: 10.1016/j.ecresq.2004.10.007

Restrepo, M. (1998). Identifiers of predominantly Spanish-speaking children with language impairment. *J. Speech Lang. Hear. Res.* 41, 1398–1411. doi: 10.1044/jslhr.4106.1398

Rodriguez, B., and Olswang, L. (2003). Mexican-American and Anglo-American mothers' beliefs and values about child rearing, education, and language impairment. *Am. J. Speech Lang. Pathol.* 12, 452–462. doi: 10.1044/1058-0360(2003/091

Semel, E., Wiig, E., and Secord, W. (2006). *Clinical evaluation of language fundamentals*, Fourth Edn. San Antonio, TX: Psycorp.

Serra, M. (2022). Estudio del consenso en la nosología y la terminología del trastorno específico del lenguaje (TEL) en lengua catalana (Cataluña y Mallorca) con metodología Delphi. *Rev. Logopedia Foniatr. Audiol.* 42, 41–57. doi: 10.1016/j.rlfa.2020.07.001

Sim, F., Thompson, L., Marryat, L., Ramparsad, N., and Wilson, P. (2019). Predictive validity of preschool screening tools for language and behavioural difficulties: a PRISMA systematic review. *PLoS One* 14:e0211409. doi: 10.1371/journal.pone.0211409

Stromswold, K. (2001). Why 'identical twins linguistically identical? Genetic, prenatal and postnatal factors. *Cognition* 101, 333–384. doi: 10.1016/j.cognition.2006.04.007

Thal, D., Jackson-Maldonado, D., and Acosta, D. (2000). Validity of a parent-report measure of vocabulary and grammar for Spanish-speaking toddlers. *J. Speech Lang. Hear. Res.* 43, 1087–1100. doi: 10.1044/jslhr.4305.1087

Thal, D., O'Hanlon, L., Clemmons, M., and Fralin, L. (1999). Validity of a parent report measure of vocabulary and syntax for preschool children with language impairment. *J. Speech Lang. Hear. Res.* 42, 482–496. doi: 10.1044/jslhr.4202.482

Tomblin, J., Records, N., Buckwalter, P., Zhang, X., Smith, E., and O'Brien, M. (1997). Prevalence of specific language impairment in kindergarten children. *J. Speech Lang. Hear. Res.* 40, 1245–1260. doi: 10.1044/jslhr.4006.1245

Torchiano, M. (2020). *effsize: Efficient Effect Size Computation*. Available online at: https://CRAN.R-project.org/package=effsize (accessed February 06, 2023).

Viding, E., Price, T., Spinath, F., Bishop, D., Dale, P., and Plomin, R. (2003). Genetic and environmental mediation of the relationship between language and nonverbal impairment in 4-year-old twins. *J. Speech Lang. Hear. Res.* 46, 1271–1282. doi: 10.1044/1092-4388(2003/099

frontiers | Frontiers in Psychology

TYPE Mini Review
PUBLISHED 20 June 2023
DOI 10.3389/fpsyg.2023.1169775

Check for updates

OPEN ACCESS

EDITED BY
Maria-José Ezeizabarrena,
University of the Basque Country (UPV/EHU), Spain

REVIEWED BY
Caroline Junge,
Utrecht University, Netherlands
Julie Elizabeth Dockrell,
University College London, United Kingdom

*CORRESPONDENCE
Diane Pesco
diane.pesco@concordia.ca

†These authors share first authorship

RECEIVED 20 February 2023
ACCEPTED 30 May 2023
PUBLISHED 20 June 2023

CITATION
Pesco D and O'Neill DK (2023) Assessing pragmatics in early childhood with the Language Use Inventory across seven languages.
*Front. Psychol.* 14:1169775.
doi: 10.3389/fpsyg.2023.1169775



# Assessing pragmatics in early childhood with the Language Use Inventory across seven languages

Diane Pesco[1]*† and Daniela K. O'Neill[2]†

[1]Department of Education, Concordia University, Montreal, QC, Canada, [2]Department of Psychology, University of Waterloo, Waterloo, ON, Canada

The Language Use Inventory (LUI) is a parent-report measure of the pragmatic functions of young children's language, standardized and norm-referenced in English (Canada) for children aged 18–47 months. The unique focus of the LUI, along with its appeal to parents, reliability and validity, and usefulness in both research and clinical contexts has prompted research teams globally to translate and adapt it to other languages. In this review, we describe the original LUI's key features and report on processes used by seven different research teams to translate and adapt it to Arabic, French, Italian, Mandarin, Norwegian, Polish, and Portuguese. We also review data from the studies of the seven translated versions, which indicate that all the LUI versions were reliable and sensitive to developmental changes. The review demonstrates that the LUI, informed by a social-cognitive and functional approach to language development, captures growth in children's language use across a range of linguistic and cultural contexts, and as such, can serve as a valuable tool for clinical and research purposes.

KEYWORDS

**language development, pragmatics, social communication, parent report, Language Use Inventory, LUI, language assessment, cross-linguistic**

## Introduction

This mini-review considers seven studies aimed at translating and adapting the original English version of the Language Use Inventory (LUI; O'Neill, 2009) – a parent-report measure – to Arabic-Saudi Najdi dialect (AlKadhi, 2015), Canadian French (Pesco and O'Neill, 2016), Italian (Longobardi et al., 2017, 2021), Mandarin Chinese (Qian et al., 2022), Norwegian (Helland and Møllerhaug, 2020), Polish (Bialecka-Pikul et al., 2019), and European Portuguese (Guimarães et al., 2013; Guimarães and Cruz-Santos, 2020). The LUI is a standardized and norm-referenced parent questionnaire designed to specifically assess pragmatics for children aged 18–47 months old. The LUI asks parents about how their child is using language, including for what purposes, types of questions and comments, and how they adapt their communication to context (O'Neill, 2007, 2009). Its completion by parents is grounded on the premise that they are ideally suited to observe these early abilities in diverse contexts, and on evidence that parents are accurate reporters of children's language production (Fenson et al., 2007) and that their reports are comparable in accuracy to screenings for language disorder carried out by trained examiners (So and To, 2022). Pragmatics is also referred to in the literature as social communication (e.g., Dillon et al., 2021) and implicated in the category of social (pragmatic) communication disorder (SPCD) introduced in the last edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5; American Psychiatric Association [APA], 2013). As the use and differentiation of these terms is still evolving, we will continue to use the term pragmatics here.

We begin with key information about the original LUI (O'Neill, 2009) to set the stage for our review of seven studies of the LUI's development in other languages, in terms of (a) the procedures used to translate and adapt the LUI to specific linguistic and cultural contexts, (b) the reliability of the translations, and (c) the translations' developmental sensitivity, drawn from patterns of children's LUI scores by age and gender across the translations.

# The Language Use Inventory

## Design of the LUI

The LUI is organized to three parts and 14 subscales, as shown in **Table 1**. Part 1, comprised of two subscales (A and B), focuses on children's gestures. While Part 1 does not address language use *per se* and consequently does not figure into the LUI Total Score, it allows parents an opportunity to respond affirmatively even if a child produces only a few words. Parts 2 and 3, which comprise the LUI Total Score, ask parents to reflect on how their child uses language in daily life (e.g., "Your child's requests for help") and roughly follow a developmental sequence (i.e., Part 2 asks about pragmatic functions realized through words and early word combinations, while Part 3 focuses on longer sentences). Across Parts 2 and 3, 10 subscales focus on children's language use (161 items comprising the LUI Total Score), while two unscored subscales (E and L) survey children's interests in play and conversation via open-ended questions. Norms for the original LUI with English-speaking children are monthly and were derived from 3,563 children residing in Canada (O'Neill, 2009).

For the 10 scored subscales, parents are asked about a particular use of language and, for most subscale items, are provided with examples of what a child might say. For example, one item asking whether a child expresses a desire to do something on their own is accompanied by the examples "I want to do it" and "Me do it." The examples are intended to support parents' understanding of the questions and also make clear to parents that variations in the form of children's utterances are allowed. Parents typically find the LUI easy to complete, likely due to its intentionally simple format (mainly yes/no questions) and focus on parents' recent observations – factors that enhance the reliability of parent reports (Fenson et al., 2007). The LUI also avoids probing language use influenced by social and cultural conventions of politeness and/or appropriateness but prone to more variation across cultures such as saying "please" or "bye-bye" (Pesco and O'Neill, 2012; O'Neill, 2014). Instead, it emphasizes language use driven by advances in children's social cognition (O'Neill, 2007), such as children's growing awareness of their own and others' mental states, differences that may exist between them, and how children may need to adapt their communication as a result.

The LUI's directions to parents also make clear that they can respond affirmatively to an item regardless of language (or other communicative means such as sign language) used by the child. Additionally, the LUI (in English) includes a question asking parents to estimate how much of the time their child is regularly exposed to a language other than English (range 0–100%). Only children whose parents reported they were exposed to English 80% or more of their waking hours were included in the norming sample (O'Neill, 2009). Thus, a parent's estimate can be considered by clinicians in deciding whether to apply the LUI's norms or report results only descriptively.

## Psychometric properties of the LUI and use in clinical practice and research

The LUI has strong discriminative, predictive, and concurrent validity. To elaborate, O'Neill (2007) showed that the LUI classified children into two groups – language delay or impairment based on clinical assessments by speech-language pathologists versus typically developing – with over 95% accuracy (i.e., sensitivity and specificity were each 95.9%). Pesco and O'Neill (2012) examined the LUI's predictive validity with 348 children from the LUI norming sample. It was found that for children assessed with the LUI between the ages of 24 and 47 months, LUI scores predicted their language and communication skills at ages 5–6 years (*M* age 5;6) assessed via a protocol that included standardized language measures and clinical history. The values were 81% for sensitivity and 93% for specificity, despite a time interval of up to 3 years between the LUI and follow-up measures at ages 5–6, a factor known to lower these values (So and To, 2022). Additionally, children who scored below an empirically derived cut-off on the LUI at ages 24–47 months were 27 times more likely to display language difficulties at ages 5–6 than children scoring above the cut-off. Children's LUI scores also concur with direct measures such as the Communication and Symbolic Behavior Scales (O'Neill, 2009), observations of language use in laboratory settings (Abbot-Smith et al., 2015), and an SLP- and parent-report measure, the Functional Communication Classification System (Caynes et al., 2021). The LUI also has high test-retest reliability (O'Neill, 2007, 2009).

The LUI's unique focus on language use in daily life, design features, and psychometric properties have led to its wide use globally and to its recommendation as a benchmark measure of pragmatics (e.g., Tager-Flusberg et al., 2009). Researchers working with both the English and translated versions have also found the LUI to be a highly valuable tool to describe strengths and weaknesses in pragmatics among diverse groups of children, such as children with autism (Qian et al., 2022) and their siblings (Miller et al., 2015), children who have experienced neglect (Di Sante et al., 2019), and children with complex disabilities (Foster-Cohen and van Bysterveldt, 2016), amongst others. It has also been used in intervention to set goals for children and monitor their progress (Foster-Cohen and van Bysterveldt, 2016).

While there is continued discussion of whether routine, universal screening of early language is advisable (see Sansavini et al., 2021; So and To, 2022), the discriminative and predictive validity studies of the LUI described above provide support for screening and monitoring children, particularly when a concern about pragmatic language use is present (see also Miller et al., 2015; Conti et al., 2020). Additionally, the LUI's internal reliability and sensitivity to growth in children's language use are each high; data

TABLE 1 Cronbach's alpha coefficients by LUI part and subscale (studies ordered by sample size).

| Study | English *N* = 3,563 (norming) | Portuguese *N* = 1,555[a] (norming) | Polish *N* = 256 | French *N* = 242 | Italian *N* = 190[b] | Mandarin *N* = 177 | Norwegian *N* = 139 | Arabic *N* = 134 |
|---|---|---|---|---|---|---|---|---|
| Age or age range of children | 18–47 months | 18–47 months | 32 months[c] | 18–47 months | 18–47 months | 18–47 months | 18–47 months | 24–41 months |
| LUI parts and subscales[d] | | | | | | | | |
| Part 1 Gestures | 0.88 | 0.88 | 0.83 | 0.86 | 0.87 | 0.90 | 0.86 | – |
| A Asking for something | 0.89 | 0.88 | 0.84 | 0.86 | 0.86 | 0.92 | 0.88 | – |
| B Getting someone to notice something[e] | 0.53 | 0.38 | 0.49 | 0.52 | 0.63 | 0.74 | 0.74 | – |
| Part 2 Words | 0.95 | 0.94 | 0.89 | 0.94 | 0.94 | 0.94 | 0.94 | 0.93 |
| C Types of words | 0.93 | 0.93 | 0.88 | 0.93 | 0.94 | 0.94 | 0.93 | – |
| D Requests for help | 0.88 | 0.80 | 0.64 | 0.76 | 0.74 | 0.71 | 0.77 | – |
| Part 3 Longer Sentences | 0.99 | 0.99 | 0.98 | 0.99 | 0.99 | 0.99 | 0.99 | 0.98 |
| F Getting someone to notice something[e] | 0.81 | 0.76 | 0.69 | 0.73 | 0.76 | 0.75 | 0.81 | – |
| G Questions/comments – things | 0.90 | 0.91 | 0.85 | 0.90 | 0.92 | 0.94 | 0.88 | – |
| H Questions/comments – self/others | 0.98 | 0.98 | 0.95 | 0.98 | 0.97 | 0.98 | 0.98 | – |
| I Talk in activities with others | 0.93 | 0.93 | 0.85 | 0.92 | 0.91 | 0.93 | 0.93 | – |
| J Teasing/sense of humour[d] | 0.79 | 0.75 | 0.59 | 0.78 | 0.73 | 0.83 | 0.78 | – |
| K Interest in words and language | 0.86 | 0.86 | 0.78 | 0.86 | 0.81 | 0.92 | 0.84 | – |
| M Adapting conversation to others | 0.93 | 0.93 | 0.85 | 0.93 | 0.89 | 0.95 | 0.91 | – |
| N Building longer sentences and stories | 0.98 | 0.97 | 0.96 | 0.98 | 0.96 | 0.98 | 0.97 | – |
| LUI Total Score, Parts 2 and 3 | 0.99 | – | 0.85 | – | – | 0.99 | 0.99 | – |

En-dashes, not reported in publication.

[a] Guimarães and Cruz-Santos (2020).

[b] Longobardi et al. (2017).

[c] For the LUI-Polish longitudinal study, alpha coefficients were calculated at 20, 32, and 44 months (see p. 2325 of their article); we took the midpoint.

[d] For a description of parts and subscales, see section "Design of the LUI." Note that E and L are unscored subscales.

[e] As noted in the text, subscale B has low variance and only 2 items, contributing to its lower alpha values; F and J also have relatively small numbers of items (6 and 5, respectively) compared to the other subscales.



10.3389/fpsyg.2023.1169775

relevant to these features are reported in later sections where they serve as a comparison point for findings from studies of the seven LUI translations.

## Research on translations of the LUI into other languages

The LUI's assets have prompted researchers globally to translate it into other languages. There are, at the time of writing, 16 translations completed or in-progress according to the publisher's website.[1] Our review focuses on the translations of the LUI to Arabic-Saudi Najdi dialect (AlKadhi, 2015), Canadian French (Pesco and O'Neill, 2016), Italian (Longobardi et al., 2017), Mandarin Chinese (Qian et al., 2022), Norwegian (Helland and Møllerhaug, 2020), Polish (Bialecka-Pikul et al., 2019), and European Portuguese (Guimarães et al., 2013; Guimarães and Cruz-Santos, 2020), seven languages for which authors have disseminated their research findings. We report on these studies next, addressing in turn the procedures for translating and adapting the LUI; the reliability of the translated/adapted versions; and the patterns observed across language in the children's LUI scores overall and for boys and girls separately, given sex differences noted in the original English LUI norming study that led to separate norms.

### Translation and adaptation processes

A number of procedures were used across the studies to ensure the translated LUI was consistent with the original measure yet adapted as needed to be appropriate to the linguistic and cultural context. First, all seven studies involved translations of the instructions to parents and all items from English to the target language (i.e., *forward translation*). The examples of what a child might say to realize a particular pragmatic function were also translated to reflect children's utterances in the target language. The forward translations for all the LUI translations were carried out by native speakers of the target language with expertise in child language (either the principal researchers or research assistants), and were then reviewed by expert panels. Among the members of these expert panels were translators, other research team members, consulting researchers from relevant fields (e.g., linguists), and speech-language pathologists. For the LUI-Arabic (AlKadhi, 2015), LUI-Portuguese (Guimarães et al., 2013; Guimarães and Cruz-Santos, 2020), LUI-Italian (Longobardi et al., 2017), and LUI-Norwegian (Helland and Møllerhaug, 2020), *back translation* (i.e., translating material back to the original language to check for equivalence) followed. The remaining teams used only forward translation to avoid the risk of overly literal back translations and confounding true differences in meaning between the translation and original with differential quality of the forward and back translations (see Qian et al., 2022 for sources recommending this approach).

A third procedure reported in the studies relates to the instructions to parents for completing the LUI. In the seven studies we reviewed, these remained very close to the original. Occasionally, a team deemed it necessary to adapt the instructions. Bialecka-Pikul et al. (2019) explained that it is common for children acquiring Polish – a morphologically rich language – to truncate multimorphemic or multisyllabic words. The research team thus added instructions to the LUI-Polish to encourage parents to consider truncated forms as words (p. 2322). For the LUI-Arabic, the instructions to parents were intentionally written in Modern Standard Arabic, while the examples of what children might say were provided in the Saudi Najdi dialect. Other teams reported no or only minor adjustments to the instructions or the items. The (free) license from the publisher to translate the LUI does ask researchers to report all changes, however minor, and the reason for these, to allow readers and users to understand just what was changed from the original English LUI, and why.

Parents were also invited by some research teams to review the LUI translations for any final revisions needed. Guimarães et al. (2013) engaged ten parents in a think-aloud session to obtain their feedback on the wording of items and examples of children's language use for the LUI-Portuguese items and Pesco and O'Neill (2016) invited three Canadian (Quebec) mothers with varying educational levels to complete the LUI-French and comment on whether it was clear, thorough (i.e., covered their child's language uses adequately) and easy to complete. AlKadhi (2015), Longobardi et al. (2021), and Qian et al. (2022) also engaged Saudi, Italian, and Chinese mothers (respectively) in a similar process.

### Reliability

For all versions, data were obtained by having parents of children of different ages complete the LUI in the target language. The ages of the children varied, with some of the seven teams including parents of children from 18 to 47 months of age as for the norming study of the original LUI, and other teams sampling children only at selected ages within this period (see **Table 1** and section "Developmental sensitivity"). The decision to sample only at selected ages was mainly due to limitations in resources available to the research teams for these initial studies and/or consideration of the ages of child health checks or immunization schedules in the country of interest. As **Table 1** shows, the size of the samples was also dramatically different, with as many as 1,555 participants in a norming study of the LUI-Portuguese and as few as 134 participants in the pilot study of the LUI-Arabic.

To test for internal reliability, the seven research teams uniformly calculated Cronbach's alpha. The values are reported in **Table 1** by study, for the LUI's three parts and subscales. The coefficients for Part 1 (the gesture subscales) range from 0.83 to 0.90 but were lower for subscale B, likely as it is comprised of just two items related to pointing and showing gestures whose use remains high at all ages (i.e., have low variance), influencing the alpha value. For Parts 2 and 3 (comprising the 10 expressive subscales used to calculate a child's LUI Total Score), the coefficients range from 0.93 to 0.99, with one exception (LUI-Polish Part 2 = 0.89). Thus, the parts of the LUI demonstrated very good to excellent internal reliability. The reliability analyses also showed that there was little need to change items, resulting in maximum LUI Total Scores that are equivalent to the original (161) or differ at most by 2. The need

1 https://languageuseinventory.com/translations

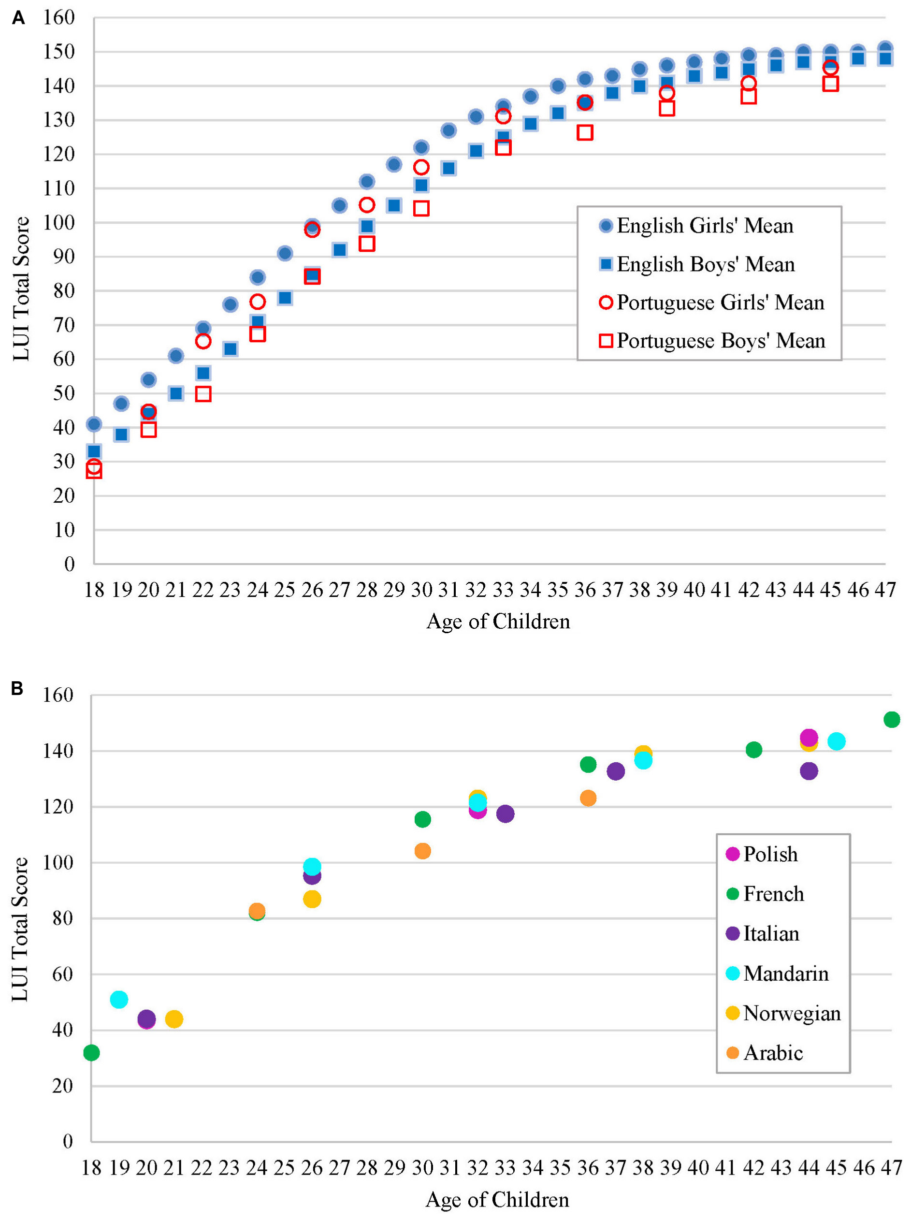


FIGURE 1
**(A)** Mean LUI Total Score by age and sex for the LUI (*N* = 3,563) and the LUI-Portuguese (*N* = 1,555) norming studies. **(B)** Mean LUI Total Score by age for six additional languages. For the LUI-Polish and LUI-French, means are for exact ages; for the remainder, the mean of the relevant age band is plotted.

for so few changes suggests that the uses of language addressed on the LUI are present in early childhood across the diverse linguistic and cultural contexts studied. This unanticipated finding could be partly attributed to the LUI's focus on uses of language that reflect children's developing social cognition rather than social and cultural conventions, as noted in the section "Design of the LUI."

## Developmental sensitivity

For the LUI-Portuguese, Guimarães and Cruz-Santos (2020) reported data for 1,555 children from 18 to 47 months old at 2-month intervals (18, 20 months, etc.). Their data thus most resemble the English LUI norming data (O'Neill, 2009) where data were collected at every month. Figure 1A shows how children's LUI Total Score compared across the two languages. It reveals a clear rise in scores at the younger ages for the two languages and for both boys and girls, followed by a more gradual rise or leveling off of scores at the older ages. The girls' and boys' means across the two languages are strikingly similar. Furthermore, higher scores for girls at some ages were observed in both samples and led to separate norms for boys and girls.

Of the six remaining studies, two reported data at selected ages, rather than at every month: namely at 18, 24, 30, 36, 42, and 47 months for the LUI-French, and at ages 20, 32, and 44 months for the LUI-Polish. The four other studies combined data from children in 6-month age bands, namely 18–23, 24–29, 30–35, 36–41, and 42–47 months for the LUI-Italian, LUI-Mandarin, and LUI-Norwegian, and from only the three middle age bands (i.e., 24–29, 30–35, 36–41 months) for the LUI-Arabic. Figure 1B shows the

LUI Total Scores by age for these six translations. The data show that in the six studies, as in the two norming studies presented in **Figure 1A**, the LUI effectively captured developmental change in the 18–47-month period. First, growth in the LUI Total Score was observed, as evidenced by significant main effects for age in all the studies. Second, the reported *post hoc* comparisons of LUI Total Scores and/or of the LUI subscales showed that older children, on the whole, had significantly higher scores than younger children. These findings, drawn from the six studies with cross-sectional designs but also the one longitudinal study of the LUI-Polish, suggest that the LUI is sensitive to development. Furthermore, although the scores of older children at "near" ages (e.g., 36 vs. 42 months) were not always significantly different, they were in the expected direction (i.e., higher amongst the older children). It is also important to keep in mind that for children with difficulties in language or pragmatics, the onset of skills and subsequent growth in skills is likely to appear later than for typically developing children, thus resulting in significant differences even at the later ages.

Sex differences in the LUI Total Score observed in the norming studies for the original LUI and LUI-Portuguese were also found for the French, Italian, and Mandarin versions of the LUI: girls scored higher than boys, particularly at younger ages, while in the Polish longitudinal study, girls scored higher at both 30 and 44 months. Boys scored significantly higher than girls on only the LUI-Italian at older ages (i.e., in the 36–41 and 42–47 month groups). On the LUI-Norwegian and the LUI-Arabic, there were no significant differences between boys and girls on the LUI Total Score at any age, possibly due to the relatively smaller sample sizes per age group compared to the other studies.

## Discussion

This mini-review presents a first look at research conducted over the last decade on the development of translations of the LUI, a parent report measure of pragmatics, into seven different languages. By using forward translation and gathering feedback from multiple parties on the translation quality and the instructions for parents, the researchers developed versions that retain the original's appeal to parents, reflect linguistic and cultural differences, and show comparable and high internal reliability across the studies. The results attest to the value of parent report and of the translations of the LUI in investigating early pragmatic abilities across many different languages.

The studies revealed an intriguing similarity across the different cultural and linguistic contexts in terms of growth in scores with age. Part of the reason for the similarity may be that the LUI, by design, avoids conventions that are known to be culturally specific (e.g., politeness markers) and judgments of "appropriateness," and thus may reveal pragmatic functions that develop in early childhood across a wide range of contexts and regardless of the language a child is acquiring. The developmental sensitivity of each translation does not mean, however, that the pragmatic functions assessed by the LUI will emerge at precisely the same age across contexts (as one can see in **Figure 1A**, for example, the LUI-Portuguese scores appear slightly lower at most ages than the scores of English-speaking children in the LUI norming sample). Due to differences in the nature and size of the samples, and possibly environmental influences, one would expect some differences that research teams conducting translations could explore in more detail and/or larger studies.

## Limitations and future directions

A limitation of this mini-review is that we address only some qualities of the LUI in line with our goals, and available in existing published data. However, in the future, these seven research teams, and possibly others as published data becomes available, could be brought together to explore children's scores on the LUI in more detail (e.g., subscale performance by sex and SES) as a means of further broadening our knowledge of early pragmatics and its development cross-linguistically. Clearly seven translations are only a fraction of possible translations and thus more extensive or different adaptations might be required by other languages. Additionally, the LUI translations have, so far, excluded children whose parents report over 20% exposure to another language. Given the prevalence of bi/multilingualism, it will be important in future work to assess how scores might differ at greater levels of bi/multilingualism, and in such cases whether the methodology of a single-percent parent estimate might need to be adjusted in a way that remains clinically practical.

Further study of performance on the subscales and/or items of the LUI and its translations of children in clinical groups (i.e., developmental language disorder; SPCD, autism spectrum disorder, deaf and hard-of-hearing) could also reveal distinct profiles of strengths and weaknesses in pragmatics amongst these groups or conversely, commonalities, cross-linguistically. Moreover, greater study of pragmatics alongside other aspects of language development (e.g., structural aspects) may aid in differentiating the impairments in SPCD and language disorders (Norbury, 2014). It is also possible that performance on a certain subset of scales may be more sensitive to, or more suited to, investigating various outcomes of interest in a particular language (Rints et al., 2015; Dockrell et al., 2022). Finally, as the LUI is translated and used in more diverse contexts (e.g., low- and middle-income countries), it could be of interest to explore further methodological and/or technological adaptations that could benefit parents with low-literacy levels.

## Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work, and approved it for publication.

## Conflict of interest

DO'N is the author of the Language Use Inventory (LUI) and the founder and president of Knowledge in Development (KID), Inc. KID, Inc. holds the copyright to and publishes the original English version of the Language Use Inventory (LUI) commercially (languageuseinventory.com). KID, Inc. receives all proceeds from the LUI and thus DO'N is a beneficiary of proceeds from the LUI. This is a continuing relationship.

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

## Editor's note

Maria-José Ezeizabarrena edited the article in collaboration with Melita Kovacevic, University of Zagreb, Zagreb, Croatia.

## Publisher's note

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

## References


Abbot-Smith, K., Nurmsoo, E., Croll, R., Ferguson, H. J., and Forrester, M. A. (2015). How children aged 2;6 tailor verbal expressions to interlocutor informational needs. *J. Child Lang.* 43, 1277–1291. doi: 10.1017/S0305000915000616

AlKadhi, A. (2015). *Assessing early sociocognitive and language skills in young Saudi children*. Doctoral dissertation. London: City, University London.

American Psychiatric Association [APA] (2013). *Diagnostic and statistical manual of mental disorders*, 5th Edn. Virginia, VA: American Psychiatric Association. doi: 10.1176/appi.books.9780890425596

Bialecka-Pikul, M., Filip, A., Stepien-Nycz, M., Kus, K., and O'Neill, D. (2019). Rantunku! or just Tunku! Evidence for the reliability and concurrent validity of the Language Use Inventory-Polish. *J. Speech Lang. Hear. Res.* 62, 2317–2331. doi: 10.1044/2019_JSLHR-L-18-0230

Caynes, K., Rose, T. A., Burmester, D., Ware, R. S., and Johnston, L. M. (2021). Reproducibility and validity of the functional communication classification system for young children with cerebral palsy. *Dev. Med. Child Neurol.* 63, 866–873. doi: 10.1111/dmcn.14844

Conti, E., Chericoni, N., Costanzo, V., Lasala, R., Mancini, A., Prosperi, M., et al. (2020). Moving toward telehealth surveillance services for toddlers at risk for autism during the COVID-19 pandemic. *Front. Psychiatry* 11:565999. doi: 10.3389/fpsyt.2020.565999

Di Sante, M., Sylvestre, A., Bouchard, C., and Leblond, J. (2019). The pragmatic language skills of severely neglected 42-month-old children: Results of the ELLAN study. *Child Maltreat.* 24, 244–253. doi: 10.1177/1077559519828838

Dillon, E., Holingue, C., Herman, D., and Landa, R. J. (2021). Psychometrics of the Pragmatic Rating Scale for school-age children with a range of linguistic and social communication skills. *J. Speech Lang. Hear. Res.* 64, 3477–3488. doi: 10.1044/2021_JSLHR-20-00753

Dockrell, J. E., Forrest, C. L., Law, J., Mathers, S., and Charlton, J. (2022). Screening for language difficulties in disadvantaged populations on entry to early years education: Challenges and opportunities. *Front. Pediatr.* 10:833603. doi: 10.3389/fped.2022.833603

Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, J. S., and Bates, E. (2007). *MacArthur-Bates Communicative Development Inventories: User's guide and technical manual*, 2nd Edn. Baltimore, MD: Paul H. Brookes.

Foster-Cohen, S., and van Bysterveldt, A. K. (2016). Assessing the communication development of children with language delay through parent multi-questionnaire reporting. *Speech Lang. Hear.* 19, 79–86. doi: 10.1080/2050571X.2015.1108067

Guimarães, C., and Cruz-Santos, A. (2020). *Manual do LUI-Português (Portugal): Aplicação e interpretação das pontuações [LUI Portuguese Manual (Portugal): Administration and interpretation of scores]*. Waterloo: Knowledge in Development, Inc.

Guimarães, C., Cruz-Santos, A., and Almeida, L. (2013). Adaptation of the parent report Language Use Inventory for 18- to 47-months-old children to European Portuguese: A pilot study. *Audiol. Commun. Res.* 18, 332–338. https://www.redalyc.org/pdf/3915/391544056015.pdf

Helland, W. A., and Møllerhaug, L. H. (2020). Assessing pragmatic competence in 18-to 47-months-old Norwegian children. A pilot study with the Language Use Inventory (LUI). *RASK Int. J. Lang. Commun.* 52, 21–37. https://www.sdu.dk/-/media/files/epage/om_sdu/institutter/isk/rask_tidsskrift/rask_52/helland_artikel.pdf

Longobardi, E., Lonigro, A., Laghi, F., and O'Neill, D. (2017). Pragmatic language development in 18- to 47-month-old Italian children: A study with the Language Use Inventory. *First Lang.* 37, 252–266. doi: 10.1177/0142723716689273

Longobardi, E., Lonigro, A., Laghi, F., and O'Neill, D. K. (2021). The assessment of early pragmatic development: A study of the reliability and validity of the Language Use Inventory-Italian. *J. Speech Lang. Hear. Res.* 64, 3186–3194. doi: 10.1044/2021_JSLHR-20-00516

Miller, M., Young, G. S., Hutman, T., Johnson, S., Schwichtenberg, A., and Ozonoff, S. (2015). Early pragmatic language difficulties in siblings of children with autism: Implications for DSM-5 social communication disorder? *J. Child Psychol. Psychiatry* 56, 774–781. doi: 10.1111/jcpp.12342

Norbury, C. F. (2014). Practitioner review: Social (pragmatic) communication disorder conceptualization, evidence and clinical implications. *J. Child Psychol. Psychiatry* 55, 204–216. doi: 10.1111/jcpp.12154

O'Neill, D. (2007). The Language Use Inventory: A parent-report measure of pragmatic language development for 18- to 47-month-old children. *J. Speech Lang. Hear. Res.* 50, 214–228. doi: 10.1044/1092-4388(2007/017)

O'Neill, D. (2009). *Language Use Inventory (LUI): Manual*. Waterloo: Knowledge in Development, Inc.

O'Neill, D. K. (2014). "Assessing pragmatic language functioning in children: Its importance and challenges," in *Pragmatic development in first language acquisition*, ed. D. Matthews (Amsterdam: John Benjamins), 363–386. doi: 10.1075/tilar.10.20nei

Pesco, D., and O'Neill, D. (2012). Predicting later language outcomes from the Language Use Inventory. *J. Speech Lang. Hear. Res.* 55, 421–434. doi: 10.1044/1092-4388(2011/10-0273

Pesco, D., and O'Neill, D. (2016). Assessing early language use by French-speaking Canadian children: Introducing the LUI-French. *Can. J. Speech Lang. Pathol. Audiol.* 40, 198–217. https://www.cjslpa.ca/detail.php?ID=1203&lang=en

Qian, L., Shao, H., Fang, H., Xiao, T., Ding, N., Sun, B., et al. (2022). Reliability, validity and developmental sensitivity of the Language Use Inventory (LUI) in the Chinese context. *Int. J. Lang. Commun. Disord.* 57, 497–511. doi: 10.1111/1460-6984.12693

Rints, A., McAuley, T., and Nilsen, E. S. (2015). Social communication is predicted by inhibitory ability and ADHD traits in preschool-aged children: A mediation model. *J. Atten. Disord.* 19, 901–911. doi: 10.1177/1087054714558873

Sansavini, A., Favilla, M. E., Guasti, M. T., Marini, A., Millepiedi, S., Di Martino, M. V., et al. (2021). Developmental language disorder: Early predictors, age for the diagnosis, and diagnostic tools. A scoping review. *Brain Sci.* 11:654. doi: 10.3390/brainsci11050654

So, K. K. H., and To, C. K. S. (2022). Systematic review and meta-analysis of screening tools for language disorder. *Front. Pediatr.* 10:801220. doi: 10.3389/fped.2022.801220

Tager-Flusberg, H., Rogers, S., Cooper, J., Landa, R., Lord, C., Paul, R., et al. (2009). Defining spoken language benchmarks and selecting measures of expressive language development for young children with autism spectrum disorders. *J. Speech Lang. Hear. Res.* 52, 643–652. doi: 10.1044/1092-4388(2009/08-0136)

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 28 June 2023
DOI 10.3389/fpsyg.2023.1163252

Check for updates

OPEN ACCESS

EDITED BY
Maria-José Ezeizabarrena,
University of the Basque Country UPV/EHU, Spain

REVIEWED BY
Suvi Stolt,
University of Helsinki, Finland
Pasquale Rinaldi,
National Research Council (CNR), Italy

*CORRESPONDENCE
Virginia Varela-Moraga
✉ vvarela@uchile.cl

RECEIVED 10 February 2023
ACCEPTED 05 June 2023
PUBLISHED 28 June 2023

CITATION
Varela-Moraga V, Diethelm-Varela B and Pérez-Pereira M (2023) Effect of biomedical complications on very and extremely preterm children's language.
*Front. Psychol.* 14:1163252.
doi: 10.3389/fpsyg.2023.1163252



# Effect of biomedical complications on very and extremely preterm children's language

Virginia Varela-Moraga[1]*, Benjamín Diethelm-Varela[2] and Miguel Pérez-Pereira[3]

[1]Departamento de Fonoaudiología, Facultad de Medicina, Universidad de Chile, Santiago, Chile, [2]Department of Molecular Genetics and Microbiology, School of Biological Sciences, Pontificia Universidad Católica de Chile, Santiago, Chile, [3]Departamento de Psicoloxía Evolutiva e da Educación, Universidade de Santiago de Compostela, Santiago de Compostela, Spain

**Introduction:** Very and extremely preterm children have been found to show delays in the development of language in early years. In some investigations, however, a rigorous control of biomedical complications, such as Periventricular Leukomalacia (PVL), Intraventricular Hemorrhage (IVH) or Bronchopulmonary Dysplasia (BPD), does not always exist. For that reason, a confounding effect of low gestational age and biomedical complications may lead to erroneous conclusions about the effect of gestational age.

**Methods:** In this investigation we compare language development [use of words, sentence complexity and mean length of the three longest utterances (MLU3)] of three groups of Chilean children at 24 months of age (corrected age for preterm children). The first group was composed of 42 healthy full-term children (Full term group: FT), the second group of 60 preterm children born below 32 gestational weeks without medical complications (low risk preterm group: LRPT), and the third group was composed of 64 children below 32 gestational weeks who had medical complications (High risk preterm group: HRPT). The three groups were similar in terms of gender distribution, maternal education, and socio-economic environment. The instrument used to assess language was the Communicative Development Inventories (CDI). In addition, the Ages and Stages Questionnaire-3 (ASQ-3) was also used to assess other developmental dimensions.

**Results:** The results indicate that HRPT and LRPT children obtained significantly lower results than the FT group in the three language measures obtained through the CDI. No significant differences were observed between the HRPT and the LRPT groups, although the HRPT obtained the lowest results in the three CDI measures. The results obtained through the administration of the ASQ-3 confirm the delay of both preterm groups in communicative development when compared to the FT group. No significant differences between the FT and the PT groups were observed in gross motor, fine motor and problem solving dimensions of the ASQ-3. The LRPT group obtained results that were significantly higher than those of the FT group and the HRPT group in gross motor development.

**Discussion:** These results seem to indicate that the area of language development is particularly influenced by very or extremely low gestational age.

KEYWORDS
**biomedical complication, preterm children, language acquisition, parental questionnaires, psychological development**

## Introduction

Premature children are considered a vulnerable population due to their immaturity as a result of their early birth. Very premature and extremely premature children, who were born with <32 and 28 weeks of gestation, respectively (Goldenberg et al., 2008; Blencowe et al., 2019), present high morbidity and are exposed to greater biological risk. The lower the gestational age and the lower the birth weight, the more likely the presence of associated complications, chronic pathologies, and developmental delays (Bhutta et al., 2002). These biomedical complications could eventually have an impact on cognitive, linguistic, and behavioral performance during childhood. The probability of suffering cerebral palsy and neurosensory disorders increases as the gestational age decreases (Synnes et al., 1994). Children with gestational age (GA) between 31 and 23 weeks are those who have a higher probability of suffering neurodevelopmental disorders (Kilbride et al., 2004; Thorngren-Jerneck and Herbst, 2006; GuangXi Cooperative Research Group for Extremely Preterm Infants et al., 2019). GuangXi Cooperative Research Group for Extremely Preterm Infants et al. (2019) evaluated a sample of children born before 28 weeks of gestation and found that GA is a predictive factor for neurodevelopmental disorders, which means that as GA decreases, neurosensory disorders increase. Similarly, Baron et al. (2014) found that GA was the most important factor that determined differences in neuropsychological, intellectual, and behavioral functioning, and Anderson and Doyle (2008) found that 40% of the children under 26 weeks of gestation had cognitive delays.

Among the most frequent early biomedical complications found in preterm infants, bronchopulmonary dysplasia (BPD) stands out for its association with the risk of delay in cognitive and language development (Singer et al., 2001; Rvachew et al., 2005; Anderson and Doyle, 2006; Short et al., 2007; Sansavini et al., 2011; Gallini et al., 2021; Katz et al., 2022). The complications and sequelae caused by intraventricular hemorrhage (IVH) on cognitive, linguistic, and educational achievements are related to its severity, in such a way that the higher the grade of IVH, the worse its effects, with IVH of grades III and IV being particularly serious (Sherlock et al., 2005; Luu et al., 2009; Srinivasakumar et al., 2013; Vohr, 2022). A similar situation occurs with the presence of periventricular leukomalacia (PVL), which has a negative effect on cognitive and language development (Ohgi et al., 2005; Resic et al., 2008). Low Apgar scores in premature infants who manage to survive are related to the appearance of HIV, PVL, necrotizing enterocolitis, and retinopathy of prematurity, in addition to a long hospital stay (Phalen et al., 2012). Soares et al. (2017) found that the risk for language development in preterm children under 32 weeks of gestation was highly associated with the presence of intraventricular hemorrhage, bronchopulmonary dysplasia, maternal age of <18 years, birth weight of <1,000 g, and prolonged hospitalization (15–30 days minimum).

In addition to the aforementioned biomedical complications, there are environmental factors that provide additional influence (Reidy et al., 2013). Although neonatal medical risk consistently displays a negative impact on early childhood outcomes, socioeconomic and demographic risks (such as mothers who are single, of young age, or with less than a high school education) may affect cognitive, language, and motor delays, and the opposite, good socioeconomic status positively impacts on development (Mangin et al., 2016; Nyman et al., 2017; Kilbride et al., 2022).

When these biomedical complications occur in preterm newborns, the presence of negative consequences later in development is highly likely. Preterm children who have presented evident and demonstrable damage through medical examinations and procedures are considered at high risk for possible delays or disorders in their development. On the contrary, those who present few or no associated alterations are considered a low-risk group (da Ribeiro et al., 2017). In any case, there is so much heterogeneity among preterm children that even low-risk preterm infants show mixed outcomes in their language development (Casiro et al., 1990; Menyuk et al., 1995; Stolt et al., 2007; Cattani et al., 2010; Pérez-Pereira et al., 2014; Pérez-Pereira and Cruz, 2018; Suttora et al., 2020; Pérez-Pereira, 2021).

What appears repeatedly in research on very preterm and extremely preterm children with low birth weight and the presence of biomedical complications are low scores in language assessments. Most assessments of language development of children approximately 24 months of age have been carried out using the MacArthur-Bates scales or Communicative Development Inventories (CDIs) (Fenson et al., 2006). The spread of adaptations of this parental report instrument to many different languages has permitted the assessment of a large number of very young children. It would have otherwise been very difficult to assess these children through individual administration of conventional tests.

Most studies using the CDI have shown that very and extremely preterm children present lower scores and an evident delay in early lexical and morphosyntactic development or in vocabulary composition when compared to their term-born peers (Jansson-Verkasalo et al., 2004; Sansavini et al., 2006, 2011; Foster-Cohen et al., 2007; Gayraud and Kern, 2007; Stolt et al., 2007, 2009, 2012, 2017; Schults et al., 2013; Sentenac et al., 2020; Tulviste et al., 2020). The performance of preterm children is worse the lower the weight and gestational age, especially with <32 weeks of gestational age (Foster-Cohen et al., 2007; Gayraud and Kern, 2007).

The Ages and Stages Questionnaire-3rd edition (ASQ-3) (Squires et al., 2009), another parental questionnaire, has also been used as a screening test of infant and child development in 5 areas: communication, gross motor, fine motor, problem-solving, and personal social. This questionnaire has been proposed as a useful screening instrument for the following up of preterm children as a population at risk of developmental delay (Skellern et al., 2001; Marks et al., 2009; Simard et al., 2012; Agarwal et al., 2016; Ballantyne et al., 2016; Al-Hindi et al., 2021). It has been found that the scores obtained by preterm children approximately 4 years of age in the five areas explored by the ASQ-3 decrease as the gestational age of the children decreases and, coherently, that the risk of developmental delay of preterm children between 25 and 40 weeks of gestation ascends as their gestational age descends. The percentage of 4-year-old children with rates of abnormal total problems scale in the ASQ ranges from 4.2% among term-born children to 37.5% among children born at 24–25 weeks' gestation ($p < 0.001$) in a large Dutch study with 1,439 preterm children and 544 FT children (Kerstjens et al., 2012). A similar pattern has been observed in all underlying ASQ domains. Coincident results were found by Hua et al. (2021) in a study carried out with a huge sample of 137,530 Chinese preschoolers between 3 and 5 years of age who ranged from very and moderately preterm (<34 gestational weeks) to post-term (<41 gestational weeks). The authors found that the mean scores obtained in the five domains assessed with the ASQ-3 increased as the GA of the children grew. There was, however, a decrease in the scores obtained by the children born post-term when compared to the term-born children (Hua et al., 2021). Coherently, the percentage of children at risk of developmental delay in the five ASQ-3 domains increased as gestational age decreased, with, again, the exception of post-term children who presented an increase in relation to the FT children. The adjusted risks of GAs (very and moderately preterm, late preterm, early term, and post-term groups) on suspected developmental delays were observed in communication (odds ratios (ORs) were 1.83, 1.28, 1.13, and 1.21, respectively, each $p < 0.05$), gross motor skill (ORs were 1.67, 1.38, 1.10, and 1.05, respectively, each $p < 0.05$), and personal-social behavior (ORs were 1.01, 1.36, 1.12, and 1.18, respectively, each $p < 0.05$) (Hua et al., 2021).

In a study with 52 infants, 12-month-old late preterm (GA of 34–36 weeks) and 156 full-term infants matched for sex, Ballantyne et al. (2016) observed a trend for late preterm infants to be at risk of communication and gross motor delays as measured through the ASQ-3 at 12 months of chronological age. Neonatal intensive care unit (NICU) admission has been found to increase the risk of developmental delay. In another comparative study with 44 late preterm and 44 full-term children, Gutiérrez Cruz et al. (2019) also observed that the late preterm infants had significantly lower scores ($p < 0.005$) in the dimension of communication of the ASQ-3.

Al-Hindi et al. (2021) administered the Saudi Arabian version of the ASQ-3 to a sample of 61 very preterm children (below 32 weeks) between 18 and 24 months of age. Twenty-six infants (42.6%) had at least one neurodevelopmental delay. The percentages found of children with developmental delays in the different dimensions were the following: communication skills (11.5%), gross motor (11.5%), fine motor (19.7%), problem-solving skills (19.7%), and personal-social skills (23%). Therefore, the domains of personal-social, problem-solving, and fine motor skills seem to be the most affected in Saudi Arabian infants.

Early age evaluations in preterm children are of great importance to identify children at risk of developmental delays and to implement intervention programs (Schults et al., 2013; Vohr, 2016). Parental report instruments such as the CDI or the ASQ-3 may be of great help in this regard.

The primary objective of this study was to compare the language development of three groups of Chilean children at 24 months of age. Three different measures of language development were obtained through the administration of the MacArthur-Bates scales: (1) use of words, (2) sentence complexity, and (3) mean length of the three longest utterances (MLU3). The three groups of children differed in terms of gestational age and in terms of the presence or not of additional biomedical complications: (1) healthy full-term children, (2) very and extremely preterm children with biomedical complications, and (3) very and extremely preterm children without serious biomedical complications. The research questions are as follows: (1) Do full-term children show better language development than the two groups of very and extremely preterm children? (2) Do the extremely and very preterm children with biomedical complications have worse language results than the very and extremely preterm children without biomedical complications?

A secondary aim was to compare other dimensions of psychological development (communication, social interaction, fine motor development, gross motor development, and problem-solving) among the three groups of participants. The research questions are as follows: (1) Do the full-term children show better performance in the five dimensions of psychological development than the two groups of extremely and very preterm children? (2) Do the preterm children with biomedical complications have worse results in the five dimensions of psychological development than the preterm children without biomedical complications?

The hypotheses of the study are as follows:

- Full-term (FT) children will outperform the two groups of very and extremely preterm children in the three language measures.
- High-risk very preterm and extremely preterm infants (HRPT) will perform worse and have a higher incidence of delay in the three measures of language development than the low-risk very and extremely preterm (LRPT) children as well as the full-term children.
- Full-term children will outperform the two groups of preterm children in the other developmental measures obtained through the ASQ-3 (communication, gross motor, fine motor, problem-solving, and personal-social development).
- High-risk very preterm and extremely preterm infants (HRPT) will perform worse and have a higher incidence of delay in the ASQ-3 five developmental measures than the other two groups (low-risk very and extremely preterm (LRPT) and full-term children).

# Method

## Participants

Three groups of Chilean children were studied at 24 months of age (corrected age for preterm children).

The first group was composed of 42 healthy full-term children (Full-term group: FT), born between 37 and 40 weeks of gestation, and without medical problems. The children were recruited in different preschool centers located in Santiago de Chile, and they were of similar gender distribution and socioeconomic and maternal education levels as the children from the two preterm groups. The FT children obtained an Apgar score at 1 min of 7 points or higher.

The second group consisted of 60 preterm children born below 32 gestational weeks without serious medical complications (low-risk preterm group: LRPT). The children attended a follow-up program for preterm children in the Hospital Roberto del Río and the health center Cordillera Oriente, both of them located in Santiago de Chile. The criteria to include the children in this group were that they had no serious neurological impairment (IVH of

TABLE 1 Descriptive data of the three groups: Mean and (SD)/frequency (%).

| | HRPT | LRPT | FT |
|---|---|---|---|
| N | 64 | 60 | 42 |
| Gestational age | 28.27 (2.03) | 30.42 (0.81) | 38.98 (0.95) |
| Apgar score | 5.55 (2.42) | 7.71 (0.73) | 8.52 (0.67) |
| Days in NICU | 32.42 (31.48) | 6.98 (6.65) | 0.0 (0.0) |
| Birth weight | 1,139 (363) | 1,503 (327) | 3,317 (466) |
| Biomedical complications* | 2.45 (1.11) | 0 (0) | 0 (0) |
| Gender (girl) | 30 (47%) | 33 (55%) | 23 (55%) |

*Number of biomedical complications.

TABLE 2 Mothers' educational level per group (frequency).

| | HRPT | LRPT | FT |
|---|---|---|---|
| Level | 64 | 60 | 42 |
| 0 | 1 (1,6%) | 0 (0,0%) | 0 (0,0%) |
| 1 | 0 (0,0%) | 1 (1,7%) | 0 (0,0%) |
| 2 | 4 (6,3%) | 1 (1,7%) | 1 (2,4%) |
| 3 | 2 (3,1%) | 6 (10,0%) | 2 (4,8%) |
| 4 | 28 (43,8%) | 14 (23,3%) | 10 (23,8%) |
| 5 | 8 (12,5%) | 7 (11,7%) | 7 (16,7%) |
| 6 | 8 (12,5%) | 19 (31,7%) | 11 (26,2%) |
| 7 | 5 (7,8%) | 5 (8,3%) | 3 (7,1%) |
| 8 | 8 (12,5%) | 7 (11,7%) | 8 (19,0%) |

grades III or IV, PVL) or lung disease (BPD), that they had an Apgar score in the 1st min of 7 or above, and that they stayed in the neonatal intensive care unit (NICU) of the hospital for <30 days after being born.

The third group was composed of 64 children below 32 gestational weeks who had medical complications (high-risk preterm group: HRPT). The children attended a follow-up program for preterm children in the Hospital Roberto del Río and the Health Center Cordillera Oriente, both located in Santiago de Chile. The children of this group were included if they had suffered any of the following biomedical conditions: serious neurological impairment (IVH of grades III or IV, PVL), bronchopulmonary dysplasia, Apgar score in the 1st min below 7 points, or a stay in the neonatal intensive care unit (NICU) of the hospital of 30 days or longer after being born.

All the children were assessed at 24 months of age (corrected age for the preterm children).

Table 1 displays descriptive data of the three groups.

Table 2 shows the educational level of the mothers per group. The following categorization was used:

Level 0: no formal education; Level 1: basic or incomplete primary education; Level 2: complete basic or primary education; Level 3: middle or incomplete secondary education; Level 4: middle or complete secondary education; Level 5: incomplete technical studies; Level 6: complete technical studies; Level 7: incomplete university studies; and Level 8: complete university studies.

The $X^2$ test revealed that there were no significant differences between groups regarding gender distribution ($X^2 = 1.015$; df = 2.2; $p = 0.602$).

The $X^2$ test indicates that there were no significant differences between groups regarding the educational level of the mothers ($X^2 = 20.012$, df = 16.1, $p = 0.220$).

Therefore, the three groups could be considered comparable in relation to gender distribution and mother's educational level.

No children with cerebral palsy, metabolic or genetic syndromes, serious motor or sensorial problems, or pervasive developmental delay were admitted to the study.

## Instruments

The Mexican Spanish version of the Communicative Development Inventories (CDIs) (*Inventario para el Desarrollo de Habilidades Comunicativas:* IDHC) (Jackson-Maldonado et al., 2003) was used to assess the linguistic development of the participants. Given the age of the children, the form *Palabras y Enunciados* (Words and Sentences) for children between 16 and 30 months of age was given to the mothers. The CDI is a parental report instrument that has two parts. The first part, *Uso de las Palabras*, has two sections: (A) *List of words*, with a list of 680 words, organized into 23 categories, from which the parents must mark those that are produced by their child, with one point given for each marked word (out of the 680 words that make up the Mexican Spanish version, 666 (98%) have exactly the same form in Chilean Spanish, and 14 words have variations; the person in charge of the application explained their meanings to the parents when necessary); and (B) *Cómo usa y comprende el lenguaje el niño* (How children use and understand language), which contains five items that assess whether the child talks about situations to refer to the past, present, and future or with respect to absent objects or people and their search. The maximum score for this section is 5 points.

The second part of the Mexican Spanish CDI, *Oraciones y Gramática* (grammar and sentences), has three sections. The first one, A. *Formas de verbos* (Verb forms), explores the capacity of the child to produce forms of verbs in present (12 items), past (6 items), and imperative (6 items) for the three verb conjugations existing in Spanish. The maximum possible score is 24. The second section, B. *Combinación de palabras* (Word combination), asks the parents whether their child already combines words or not. If the answer is yes, the parents must give three examples of the longest utterances their child has recently produced. The mean length of these three longest utterances (MLU3) in words is obtained. If the child does not yet produce word combinations, the parents should stop filling out the inventory. The third part, C. *Complejidad de frases* (Sentence complexity), consists of 37 pairs of phrases. Both phrases express the same idea, although the second is always more complex (and evolved) from a morphosyntactic point of view. The parents are asked to mark the form that is more similar to that used by their child. One point is given if the parents mark the second option. The maximum possible score is 37 points.

 10.3389/fpsyg.2023.1163252

For the purposes of the present research, only the scores obtained in the list of words produced (named *use of words*), MLU3, and sentence complexity were considered.

The Mexican Spanish CDI has good validity and reliability values (Thal et al., 2000; Jackson-Maldonado et al., 2003).

The Spanish version of the Ages and Stages Questionnaire-3 (ASQ-3) (Squires et al., 2009) was used to assess the psychological development of the participants. The form for age 24 months was used. The ASQ-3 is a parental questionnaire that assesses five areas of development: communication, gross motor, fine motor, problem-solving, and personal social. Each area contains six items assessing different abilities in each domain, which can be scored as 0 (not yet), 5 (sometimes), or 10 (yes). The maximum score for each dimension is 60 points. The score is interpreted according to a normative chart that is included in the questionnaire. According to this chart and the User's Guide, each child can be classified in one of three areas: (1) above the cutoff, when the child's development appears to be on schedule; (2) close to the cutoff, when it is necessary to provide learning activities and monitor; or (3) below the cutoff, when further assessment with a professional may be needed. The validity and reliability of the ASQ-3 reach good values (validity is 0.82 to 0.88, test–retest reliability is 0.91, and inter-rater reliability is 0.92) (Squires et al., 2015). The results obtained through the ASQ-3 are coincident with those obtained through the Bayley III scales (Agarwal et al., 2016; Mackin et al., 2017).

Structured interview. An interview adapted from that used by Pérez-Pereira et al. (2014) explored aspects such as socioeconomic indicators, parental educational level, family composition, health issues of the child and the parents, daily routines, and family history of language problems.

In addition, information on biomedical problems, gestational history, and delivery was also obtained from the records at the hospitals.

## Procedure

Approval of the *Comité de Ética para Investigaciones en Seres Humanos* (Ethics Committee for Research with Human Beings) of the Faculty of Medicine of the University of Chile was obtained to carry out this research (resolution No. 1026). Prior informed consent was also given by the participants' parents.

The preterm children were selected from the two previously reported medical centers: Hospital Roberto del Río and the Health Center Cordillera Oriente in Santiago de Chile, where the preterm children were included in a follow-up program. Information about the eligible children was obtained from these centers, and those who fulfilled the age requirements and the rest of the inclusion criteria previously specified were included.

The full-term participants that presented the age requirements and inclusion criteria were chosen from preschool centers in the north area of Santiago de Chile.

During the second semester of 2019, the instruments were administered in person. From the year 2020, because of the COVID-19 pandemic, the instruments were administered remotely (through email, SMS, WhatsApp, or video call). Previously, the mother (in most cases) or the father of each child was called by phone to arrange the modality of contact. In this first call, information on the study and instructions on how to proceed were given to the parents.

The first author carried out the assessment. The assessment took place when the children were 24 months old (±15 days), using the chronological age for the full-term children and the corrected age for the preterm children.

The parents filled out the two questionnaires (Mexican Spanish CDI and ASQ-3) and the structured interview.

## Analysis performed

Analyses of variance (ANOVA) were performed to compare the scores obtained by the three groups of children (FT, LRPT, and HRPT) in the three measures taken from the CDI—that is, use of words or word production, MLU3, and sentence complexity—and the five measures obtained through the ASQ-3. The SPSS-28 was used for the analyses.

In addition, and solely for discussion purposes, Pearson's correlations between the three language measures (IDHC) and the five measures of psychological development (ASQ-3) were obtained.

# Results

Table 3 shows the results obtained in the three measures of the IDHC taken under consideration.

As can be observed, the ANOVA results indicate that there are significant differences among the groups in the use of words, MLU3, or sentence complexity. A *post hoc* Bonferroni test ($p < 0.05$) reveals that those differences are due to the significantly higher results obtained by the FT children in relation to those obtained by the HRPT and the LRPT groups; there are no significant differences, however, between the HRPT and the LRPT groups in the use of words, MLU3, or sentence complexity. The relatively higher results obtained by the LRPT children in comparison with the HRPT children do not reach significance.

The results obtained in the ASQ-3 are shown in Table 4, together with the results of the ANOVA.

The area of *communication* is where the differences between preterm children and full-term children are the highest. The difference between the FT group, on the one hand, and the HRPT and the LRPT groups, on the other hand, is highly significant, and the effect size ($\eta^2$) reaches nearly 19%. The Bonferroni *post hoc* test confirms that the difference found in the ANOVA is due to the significantly higher result obtained by the FT children in relation to both the HRPT and the LRFT children ($p < 0.001$).

In the areas of *fine motor* development and *problem-solving* capacity, there are no significant differences among the three groups, although, in this case, the group with the best results is the LRPT and not the FT group.

In the area of *gross motor* development, the LRPT group obtains the highest results, but now the ANOVA value reaches significance. The Bonferroni *post hoc* test indicates that there are significant differences between the LRPT and the HRPT groups and between the LRPT and FT groups ($p < 0.001$).

TABLE 3 ANOVA comparisons among groups in the scores of the CDI [mean and (SD)].

| | HRPT | LRPT | FT | F | df | $p$ | $\eta^2$ |
|---|---|---|---|---|---|---|---|
| N | 64 | 60 | 42 | | | | |
| Use of words | 57.27 (63.44) | 71.93 (63.16) | 233.14 (167.71) | 44.688 | 2.163 | <0.001*** | 0.354 |
| MLU3 | 1.51 (0.77) | 1.69 (0.69) | 2.62 (1.21) | 21.884 | 2.163 | <0.001*** | 0.212 |
| Sentence complexity | 0.81 (1.76) | 0.92 (1.09) | 7.50 (9.46) | 28.618 | 2.163 | <0.001*** | 0.260 |

***$p < 0.001$.

TABLE 4 ANOVA comparisons among groups in the scores of the ASQ-3 [mean and (SD)].

| | HRPT | LRPT | FT | F | df | $p$ | $\eta^2$ |
|---|---|---|---|---|---|---|---|
| Communication | 28.52 (17.13) | 33.00 (17.42) | 48.69 (13.88) | 18.785 | 2.163 | <0.001*** | 0.187 |
| Gross motor | 47.50 (13.09) | 55.42 (5.92) | 50.95 (8.78) | 5.935 | 2.163 | 0.003** | 0.068 |
| Fine motor | 45.78 (11.52) | 48.17 (7.42) | 47.38 (9.64) | 0.573 | 2.163 | 0.565 | 0.007 |
| Problem-solving | 43.28 (13.66) | 47.67 (12.12) | 43.45 (10.03) | 1.100 | 2.163 | 0.335 | 0.013 |
| Personal social | 42.42 (14.36) | 47.00 (9.62) | 49.64 (31.71) | 5.403 | 2.163 | 0.005** | 0.062 |

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

Finally, the ANOVA results indicate that there are significant differences between the groups in the area of *personal-social* development. The *post hoc* Bonferroni test confirms that the FT group obtains significantly higher results than the HRPT ($p < 0.001$) and that the LRPT group obtains significantly higher results than the HRPT group. The difference between the FT and the LRPT groups does not reach significance.

The two groups of PT children also differed in birth weight and gestational age. In order to check whether these differences would have affected the results, we have introduced these variables as covariates in univariate general linear models. Dependent variables use of words, MLU3, and sentence complexity (obtained through the CDI) and communication, gross motor, fine motor, problem-solving, and personal social (obtained through the ASQ-3) were successively introduced. Belonging to the LRPT or HRPT group has been used as an independent variable (fixed factor). Birth weight and gestational age were introduced as covariates. The results show that gestational age does not affect any of the results, while birth weight only affects gross motor skills and personal-social development. The pattern of results regarding the independent variable (LRPT vs. HRPT group) does not change in relation to the results of the ANOVA test reported in Tables 3, 4, with the sole exception of personal-social development, which when introducing the covariates does not show a significant difference between the LRPT and HRPT groups ($F = 1.532$, $p = 0.218$).

Considering the interpretation of the mean scores of each group in the different ASQ-3 areas, the area of *communication* is clear, that is, in which the two groups of preterm children obtain the worse results. In relation to the limit score in this area (which is 25.17), the FT children are clearly above the cutoff, while the HRPT and the LRPT children are close to the cutoff and would need monitoring and additional learning activities to help their development. In the four remaining areas, the scores obtained by the three groups are above the cutoff, with the children showing age-appropriate performance.

Spearman's correlations are shown in Table 5. The results clearly show that the correlations between the ASQ-3 communication scores and the IDHC scores (use of words $r = 0.735$, MLU3 $r = 0.800$, and sentence complexity $= 0.511$) are the highest ($p < 0.001$).

## Discussion

In relation to the first objective of the study, the FT group obtained significantly higher scores in the use of words, MLU3, and sentence complexity than the other two PT groups: HRPT and LRPT, as Table 3 shows. These results agree with those found in other studies comparing FT and very and extremely preterm children through the CDI (Jansson-Verkasalo et al., 2004; Sansavini et al., 2006, 2011; Foster-Cohen et al., 2007; Gayraud and Kern, 2007; Stolt et al., 2007, 2009, 2012, 2017; Schults et al., 2013; Sentenac et al., 2020; Tulviste et al., 2020). Therefore, we can say that hypothesis 1 is confirmed. The effect size ($\eta^2$) found in the use of words (0.354) was higher than the effect size obtained in the two morphosyntactic measures: MLU3 (0.212) and sentence complexity (0.260). This is probably because morphosyntactic development is barely developed at the age of 24 months, and, logically, the differences found are lower than in lexical development.

The second hypothesis is only partially confirmed; however, because even though the differences between the HRPT group and the FT group are clearly significant in the three language measures, the differences between the HRPT and the LRPT groups do not reach significance. These results do not match those found in other studies, which point to a negative effect of biomedical complications on language development (Singer et al., 2001; Ohgi et al., 2005; Rvachew et al., 2005; Sherlock et al., 2005; Anderson and Doyle, 2006; Short et al., 2007; Resic et al., 2008; Luu et al., 2009; Sansavini et al., 2011; Phalen et al., 2012; Srinivasakumar et al., 2013; Soares et al., 2017; Gallini et al., 2021; Katz et al., 2022; Vohr,

TABLE 5 Pearson's correlations (bilateral) between IDHC and ASQ-3 measures.

| | Communication | Gross motor | Fine motor | Problem-solving | Personal social |
|---|---|---|---|---|---|
| Use of words | 0.735*** | 0.112 | 0.166* | 0.119 | 0.371*** |
| MLU3 | 0.800*** | 0.126 | 0.186* | 0.178* | 0.373*** |
| Sentence complexity | 0.511*** | 0.029 | 0.138 | 0.078 | 0.250** |

*$p < 0.05$. **$p < 0.01$. ***$p < 0.001$.

2022). Few exceptions to this widespread pattern are available. Although the children from the LRPT group obtained higher mean scores than the children from the HRPT group in the use of words (71.93 vs. 57.27), MLU3 (1.69 vs. 1.51), and sentence complexity (0.92 vs. 0.81, respectively), these differences are not significant. Probably, the reduced number of participants in the sample makes it difficult to find significant differences. Furthermore, we have used the criteria of counting the number of risk circumstances (from 1 to 5), and the combinations of these biomedical risks could vary. We have not analyzed the effect of singular risk conditions (such as having BPD or IVH of grades III and IV) because the number of children who suffered from them was rather limited. In addition, many children from the HRPT group presented comorbidities, and two or more biomedical risks were present.

The comparison of the results obtained by the three groups in the five areas of psychological development assessed by the ASQ-3 indicates that hypothesis 3 is only partially confirmed. The area of *communication* is the only one in which the FT group obtains significantly higher results than the two groups of very and extremely PT children (see Table 4). This result agrees with those found by other studies (Ballantyne et al., 2016; Gutiérrez Cruz et al., 2019; Hua et al., 2021), which point to a higher risk of delay of PT children in communication. In this area, the effect size ($\eta^2$) is higher (0.187).

Along the same line, the results obtained in *personal-social* development seem to support the hypothesis that FT children should have higher results than the LRPT and the HRPT groups. The FT group obtained the highest results, followed by the LRPT group, and finally the HRPT group. In this case, significant differences were found between the FT and the HRPT groups but not between the FT and the LRPT groups. This result is, for the most part, coincident with those found by other studies (Al-Hindi et al., 2021; Hua et al., 2021), although the mean scores obtained by all the groups are above the cutoff.

In the area of *fine motor development*, however, there were no significant differences among the groups, and hypothesis 3 has not been supported. The group of LRPT children even obtained the highest scores in this dimension, although no significant differences were found. This result does not agree with that found by Al-Hindi et al. (2021), who found a relatively high percentage (19.7%) of very preterm children who showed developmental delays in fine motor development. In any case, the mean scores of all the groups are appropriate to their age.

Similar results were found for *problem-solving*. Although the group of LRPT children obtained the highest results, no significant differences were found among the groups. This result is not in agreement with what has been found in other studies (Kerstjens et al., 2012; Al-Hindi et al., 2021). Again, the mean scores of the three groups were above the cutoff, and no developmental risk has been found.

Finally, the results observed in *gross motor* development are in contradiction with hypothesis 3 because the LRPT group obtained results that were significantly higher than those of the FT group and, even more so, than those of the HRPT group. In any case, no group seemed to be at risk of developmental delay in this area, since the mean scores of all the groups were in the normal range, according to ASQ-3 norms. The results obtained in gross motor development in our study are dissonant with those studies that found that PT children obtained significantly lower results than FT children (Ballantyne et al., 2016; Hua et al., 2021).

In relation to hypothesis 4, as expected, the HRPT group obtained lower results than the FT group in the five areas of development explored by the ASQ-3, although these differences reach significance in only communicative and personal-social development. The HRPT group also obtained lower results than the LRPT group in all the areas explored, although those differences reach significance only in gross motor development.

Clearly, communicative development seems to be the area most affected by prematurity. The development of both the HRPT and LRPT children in the remaining areas seems to unfold according to expectations. Curiously, LRPT children have even higher results than the FT children in gross motor, fine motor, and problem-solving development, although only in gross motor development do differences reach significance.

Therefore, the results obtained with the ASQ-3 do not support the idea that very and extremely preterm children show delays in all areas of development at 24 months of age, as suggested by other studies carried out with 4-year-old children (Kerstjens et al., 2012; Hua et al., 2021). Their development seems to follow normal patterns in gross motor, fine motor, problems solving, and personal-social development, particularly if these PT children do not have biomedical complications. The improvement in the care provided to PT children in the hospitals in the last years might be also related to the results found.

The high correlations found between the results of communication in the ASQ-3 and the use of words MLU3 and sentence complexity in the CDI reinforce the idea that language and communication are particularly affected in very and extremely preterm children regardless of whether they present medical complications or not.

## Conclusion

The extremely and very preterm children studied obtained significantly lower results than the full-term children in all language and communicative measures taken at 24 months of age. Contrary to expectations, the biomedical risk factors have not shown any significant effect on the development of language and communication, although HRPT children obtained the lowest results in all measures.

The development of other areas is not so much affected by preterm birth and LRPT, and HRPT children seem to develop according to expectations. LRPT children obtain higher results than the HRPT and even the FT children in certain areas of development. This indicates that biomedical complications seem to have a detrimental effect on development, particularly gross motor, fine motor, problem-solving, and personal-social development.

One limitation of the present research is the relatively reduced number of participants in the study. It is possible that differences between the HRPT and the LRPT groups could appear, particularly in language and communication, if the number of participants was larger. In the same vein, the singular effect of different biomedical complications (including necrotizing enterocolitis) could be studied with a larger number of participants.

One strength of the present research is that two groups of very and extremely preterm children, with and without biomedical complications, were studied.

## Data availability statement

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

## Ethics statement

The studies involving human participants were reviewed and approved by Comité de Ética de Investigación en Seres Humanos (CEISH) Facultad de Medicina Universidad de Chile. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

## Author contributions

VV-M contributed to conception and design of the study and performed assessments of all study participants. BD-V organized the database and performed the statistical analysis. MP-P contributed to the conception and design of the study and the analysis and interpretation of the results by critically reviewing it and wrote the draft of the manuscript. All authors contributed to manuscript review, read, and approved the submitted version.

## Conflict of interest

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

## Editor's note

Maria-José Ezeizabarrena edited the article in collaboration with Melita Kovacevic, University of Zagreb, Zagreb, Croatia.

## Publisher's note



## References

Agarwal, P. K., Shi, L., Daniel, L. M., Yang, P. H., Khoo, P. C., Quek, B. H., et al. (2016). Prospective evaluation of the Ages and Stages Questionnaire 3rd Edition in very-low-birthweight infants. *Develop. Med. Child Neurol.* 59, 484–489. doi: 10.1111/dmcn.13307

Al-Hindi, M. Y., Almahdi, B. H., Alasmari, D. A., Alwagdani, R. K., Hunjur, W. M., Khalel, A. F., et al. (2021). Screening for neurodevelopmental delay for preterm very low birth weight infants at tertiary care center in Saudi Arabia. *Cureus* 13, e20092. doi: 10.7759/cureus.20092

Anderson, P., and Doyle, L. (2008). Cognitive and educational deficits in children born extremely preterm. *Semin. Perinatol.* 32, 51–58. doi: 10.1053/j.semperi.2007.12.009

Anderson, P. J., and Doyle, L. W. (2006). Neurodevelopmental outcome of bronchopulmonary dysplasia. *Semin. Perinatol.* 30, 227–232. doi: 10.1053/j.semperi.2006.05.010

Ballantyne, M., Benzies, K. M., McDonald, S., Magill-Evans, J., and Tough, S. (2016). Risk of developmental delay: Comparison of late preterm and full-term Canadian infants at age 12 months. *Early Human Develop.* 101, 27–32. doi: 10.1016/j.earlhumdev.2016.04.004

Baron, I. S., Weiss, B. A., Litman, F. R., Ahronovich, M. D., and Baker, R. (2014). Latent mean differences in executive function in at-risk preterm children: The delay-deficit dilemma. *Neuropsychology* 28, 541–551. doi: 10.1037/neu0000076

Bhutta, A. T., Cleves, M. A., Casey, P. H., Cradock, M. M., and Anand, K. J. S. (2002). Cognitive and behavioral outcomes of school-aged children who were born preterm: a meta-analysis. *J. Am. Med. Assoc.* 288, 728–737. doi: 10.1001/jama.288.6.728

Blencowe, H., Krasevec, J., de Onis, M., Black, R. E., An, X., Stevens, G. A., et al. (2019). National, regional, and worldwide estimates of low birthweight in 2015, with trends from 2000: a systematic analysis. *Lancet Global Health* 7, e849–e860. doi: 10.1016/S2214-109X(18)30565-5

Casiro, O. G., Moddemann, D. M., Stanwick, R. S., Panikkar-Thiessen, V. K., Cowan, H., and Cheang, M. S. (1990). Language development of very low birth weight infants and fullterm controls at 12 months of age. *Early Hum. Dev.* 24, 65–77. doi: 10.1016/0378-3782(90)90007-6

Cattani, A., Bonifacio, S., Fertz, M., Iverson, J. M., Zocconi, E., and Caselli, M. C. (2010). Communicative and linguistic development in preterm children: a longitudinal study from 12 to 24 months. *Int. J. Lang. Commun. Disor.* 45, 162–173. doi: 10.3109/13682820902818870

da Ribeiro, C., de Pachelli, M. R. O., de Amaral, N. C. O., and Lamônica, D. A. C. (2017). "Habilidades do desenvolvimento de crianças prematuras de baixo peso e muito baixo peso," in *CoDAS* (Sociedade Brasileira de Fonoaudiologia) 29. doi: 10.1590/2317-1782/20162016058

Fenson, L., Marchman, V. A., Thal, D., Dale, P. S., Reznick, J. S., and Bates, E. (2006). *The MacArthur-Bates Communicative Development Inventories:*

*User's Guide and Technical Manual.* Second Edition. Baltimore: Paul H. Brookes. doi: 10.1037/t11538-000

Foster-Cohen, S., Edgin, J. O., Champion, P. R., and Woodward, L. J. (2007). Early delayed language development in very preterm infants: Evidence from the MacArthur-Bates CDI. *J. Child Lang.* 34, 655–675. doi: 10.1017/S0305000907008070

Gallini, F., Coppola, M., De Rose, D. U., Maggio, L., Arena, R., Romano, V., et al. (2021). Neurodevelopmental outcomes in very preterm infants: The role of severity of Bronchopulmonary Dysplasia. *Early Human Develop.* 152, 105275. doi: 10.1016/j.earlhumdev.2020.105275

Gayraud, F., and Kern, S. (2007). Influence of preterm birth on early lexical and grammatical acquisition. *First Langu.* 27, 159–173. doi: 10.1177/0142723706075790

Goldenberg, R. L., Culhane, J. F., Iams, J. D., and Romero, R. (2008). Epidemiology and causes of preterm birth. *The Lancet* 371, 75–84. doi: 10.1016/S0140-6736(08)60074-4

GuangXi Cooperative Research Group for Extremely Preterm Infants, Li, Y., Meng, D. H., Wei, Q. F., Pan, X. N., Liang, W. H., et al. (2019). Neurodevelopmental outcomes of extremely preterm infants in southern China: A multicenter study. *Early Hum. Dev.* 133, 5–10. doi: 10.1016/j.earlhumdev.2019.04.002

Gutiérrez Cruz, N., Torres Mohedas, J., Carrasco Marina, M. L., Olabarrieta Arnal, I., Martín del Valle, F., and García García, M. L. (2019). Desarrollo psicomotor en prematuros tardíos a los dos años de edad: comparación con recién nacidos a término mediante dos herramientas diferentes. *Rev. de Neurol.* 68, 503–509. doi: 10.33588/rn.6812.2018360

Hua, J., Barnett, A. L., Williams, G. J., Dai, X., Sun, Y., Li, H., et al. (2021). Association of gestational age at birth with subsequent suspected developmental coordination disorder in early childhood in China. *JAMA Netw. Open* 4, e2137581. doi: 10.1001/jamanetworkopen.2021.37581

Jackson-Maldonado, D., Thal, D. J., Fenson, L., Marchman, V. A., Newton, T., and Conboy, B. (2003). *MacArthur Inventarios Del Desarrollo de Habilidades Comunicativas.* User's guide and technical manual. Baltimore, EEUU: Paul H. Brookes.

Jansson-Verkasalo, E., Valkama, M., Vainionp,ää, L., Pääkk,ö, E., Ilkko, E., and Lehtihalmes, M. (2004). Language development in very low birth weight preterm children: a follow-up study. *Folia Phoniatrica Logopaedica* 56, 108–119. doi: 10.1159/000076062

Katz, T. A., Vliegenthart, R. J. S., Aarnoudse-Moens, C. S. H., Leemhuis, A. G., Beuger, S., Blok, G. J., et al. (2022). Severity of bronchopulmonary dysplasia and neurodevelopmental outcome at 2 and 5 years corrected age. *J. Pediatrics* 243, 40–46. doi: 10.1016/j.jpeds.2021.12.018

Kerstjens, J. M., De Winter, A. F., Bocca-Tjerrrtes, I. F., Bos, A. F., and Reijneveld, S. A. (2012). Risk of developmental delay increases exponentially as gestational age of preterm infants decreases: a cohort study at age 4 years. *Dev. Med. Child Neurol.* 54, 1096–1101. doi: 10.1111/j.1469-8749.2012.04423.x

Kilbride, H. W., Thorstad, K., and Daily, D. K. (2004). Preschool outcome of less than 801-gram preterm infants compared with full-term siblings. *Pediatrics* 113, 742–747. doi: 10.1542/peds.113.4.742

Kilbride, H. W., Vohr, B. R., McGowan, E. M., Peralta-Carcelen, M., Stringer, K., Das, A., et al. (2022). Early neurodevelopmental follow-up in the NICHD neonatal research network: Advancing neonatal care and outcomes, opportunities for the future. *Semin. Perinatol.* 46, 151642. doi: 10.1016/j.semperi.2022.151642

Luu, T., Vohr, B., Schneider, K., Katz, K., Tucker, R., Allan, W., et al. (2009). Trajectories of receptive language development from 3 to 12 years of age for very preterm children. *Pediatrics*, 124, 333–341. doi: 10.1542/peds.2008-2587

Mackin, R., Ben Fadel, N., Feberova, J., Murray, L., Nair, A., Kuehn, S., et al. (2017). ASQ3 and/or the Bayley-III to support clinicians' decision making. *PLoS ONE* 12, e0170171. doi: 10.1371/journal.pone.0170171

Mangin, K. S., Horwood, L. J., and Woodward, L. J. (2016). Cognitive development trajectories of very preterm and typically developing children. *Child Dev.* 88, 282–298. doi: 10.1111/cdev.12585

Marks, K., Hix-Small, H., Clark, K., and Newman, J. (2009). Lowering developmental screening thresholds and raising quality improvement for preterm children. *Pediatrics* 123, 1516–1523. doi: 10.1542/peds.2008-2051

Menyuk, P., Liebergott, J. W., and Schultz, M. C. (1995). *Early Language Development in Full-Term and Premature Infants* (1era ed.). New York: Taylor and Francis Group.c

Nyman, A., Korhonen, T., Munck, P., Parkkola, R., Lehtonen, L., and Haataja, L. (2017). Factors affecting the cognitive profile of 11-year-old children born very preterm. *Pediatr. Res.* 82, 324–332. doi: 10.1038/pr.2017.64

Ohgi, S., Akiyama, T., and Fukuda, M. (2005). Neurobehavioural profile of low-birthweight infants with cystic periventricular leukomalacia. *Dev. Med. Child Neurol.* 47, 221–228. doi: 10.1017/S0012162205000447

Pérez-Pereira, M. (2021). Prevalence of language delay among healthy preterm children, language outcomes and predictive factors. *Children* 8, 282. doi: 10.3390/children8040282

Pérez-Pereira, M., and Cruz, R. (2018). A longitudinal study of vocabulary size and composition in low-risk preterm children. *First Langu.* 38, 72–94. doi: 10.1177/0142723717730484

Pérez-Pereira, M., Fernandez, P., Luisa Gomez-Taibo, M., and Resches, M. (2014). Language development of low-risk preterm infants up to the age of 30 months. *Early Hum. Dev.* 90, 649–656. doi: 10.1016/j.earlhumdev.2014.08.004

Phalen, A. G., Kirkby, S., and Dysart, K. (2012). The 5-minute Apgar score: survival and short-term outcomes in extremely low-birth-weight infants. *J. Perinatal Neonatal Nurs.* 26, 166–171. doi: 10.1097/JPN.0b013e31825277e9

Reidy, N., Morgan, A., Thompson, D. K., Inder, T. E., Doyle, L. W., and Anderson, P. J. (2013). Impaired language abilities and white matter abnormalities in children born very preterm and/or very low birth weight. *J. Pediatr.* 162, 719–724. doi: 10.1016/j.jpeds.2012.10.017

Resic, B., Tomasovic, M., Kuzmanic, R., Lozic, M., Resic, J., and Solak, M. (2008). Neurodevelopmental outcome in children with periventricular leukomalacia. *Collegium Antropol.* 32, 143–147.

Rvachew, S., Creighton, D., Feldman, N., and Sauve, R. (2005). Vocal development of infants with very low birth weight. *Clin. Lingu. Phonet.* 19, 275–294. doi: 10.1080/02699200410001703457

Sansavini, A., Guarini, A., Alessandroni, R., Faldella, G., Giovanelli, G., and Salvioli, G. (2006). Early relations between lexical and grammatical development in very immature Italian preterms. *J. Child Lang.* 33, 199–216. doi: 10.1017/S0305000905007208

Sansavini, A., Guarini, A., Savini, S., Broccoli, S., Justice, L., Alessandroni, R., et al. (2011). Longitudinal trajectories of gestural and linguistic abilities in very preterm infants in the second year of life. *Neuropsychologia* 49, 3677–3688. doi: 10.1016/j.neuropsychologia.2011.09.023

Schults, A., Tulviste, T., and Haan, E. (2013). Early vocabulary in full term and preterm Estonian children. *Early Hum. Dev.* 89, 721–726. doi: 10.1016/j.earlhumdev.2013.05.004

Sentenac, M., Johnson, S., Charkaluk, M. L., Seppanen, A. V., Aden, U., Cuttini, M., et al. (2020). 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. Commun. Health* 74, 346–353. doi: 10.1136/jech-2019-213564

Sherlock, R. L., Anderson, P. J., and Doyle, L. W. (2005). Neurodevelopmental sequelae of intraventricular haemorrhage at 8 years of age in a regional cohort of ELBW/very preterm infants. *Early Hum. Dev.* 81, 909–916. doi: 10.1016/j.earlhumdev.2005.07.007

Short, E. J., Kirchner, H. L., Asaad, G. R., Fulton, S. E., Lewis, B. A., Klein, N., et al. (2007). Developmental sequelae in preterm infants having a diagnosis of bronchopulmonary dysplasia. *Arch. Pediatr. Adoles. Med.* 161, 1082. doi: 10.1001/archpedi.161.11.1082

Simard, M.-N., Luu, T. M., and Gosselin, J. (2012). Concurrent validity of ages and stages questionnaires in preterm infants. *Pediatrics* 130, e108–e114. doi: 10.1542/peds.2011-3532

Singer, L. T., Siegel, A. C., Lewis, B., Hawkins, S., Yamashita, T., and Baley, J. (2001). Preschool language outcomes of children with history of bronchopulmonary dysplasia and very low birth weight. *J. Dev. Behav. Pediatr.* 22, 19–26. doi: 10.1097/00004703-200102000-00003

Skellern, C., Rogers, Y., and O'Callaghan, M. (2001). A parent-completed developmental questionnaire: Follow up of ex-premature infants. *J. Paediatr. Child Health* 37, 125–129. doi: 10.1046/j.1440-1754.2001.00604.x

Soares, A. C. C., Silva, K., da, and Zuanetti, P. A. (2017). Variáveis de risco para o desenvolvimento da linguagem associadas à prematuridade. *Audiol. Commun. Res.* 22, 1745. doi: 10.1590/2317-6431-2016-1745

Squires, J., Bricker, D., and Twombly, E. (2015). *ASQ: SE-2 Ages and Stages Questionnaires® Socio-Emotional. [Internet].* Brookes Publishing Co. Available online at: https://agesandstages.com (acccessed March 18, 2021).

Squires, J., Bricker, D. D., and Twombly, E. (2009). *Ages and Stages Questionnaires® Third Edition (ASQ3).* Baltimore, MD: Paul H. Brookes. doi: 10.1037/t11523-000

Srinivasakumar, P., Limbrick, D., Munro, R., Mercer, D., Rao, R., Inder, T., et al. (2013). Posthemorrhagic ventricular dilatation-impact on early neurodevelopmental outcome. *Am. J. Perinatol.* 30, 207–214. doi: 10.1055/s-0032-1323581

Stolt, S., Haataja, L., Lapinleimu, H., and Lehtonen, L. (2009). The early lexical development and its predictive value to language skills at 2 years in very-low-birth-weight children. *J. Commun. Disor.* 42, 107–123. doi: 10.1016/j.jcomdis.2008.10.002

Stolt, S., Klippi, A., Launonen, K., Munck, P., Lehtonen, L., Lapinleimu, H., et al. (2007). Size and composition of the lexicon in prematurely born very-low-birthweight and full-term Finnish children at two years of age. *J Child Lang.* 34, 283–310. doi: 10.1017/S0305000906007902

Stolt, S., Matomaki, J., Haataja, L., Lapinleimu, H., Lehtonen, L., and Pipari Study, G. (2012). The emergence of grammar in very-low-birth-weight Finnish children at two years of age. *J Child Lang.* 40, 336–357. doi: 10.1017/S0305000911000456

Stolt, S., Savini, S., Guarini, A., Caselli, M. C., Matomäki, J., Lapinleimu, H., et al. (2017). Does the native language influence lexical composition in very preterm children at the age of two years? A cross-linguistic comparison study of Italian and Finnish children. *First Lang.* 37, 368–390. doi: 10.1177/0142723717698006

Suttora, C., Guarini, A., Zuccarini, M., Aceti, A., Corvaglia, L., and Sansavini, A. (2020). 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* 17, 7684. doi: 10.3390/ijerph17207684

Synnes, A. R., Ling, E. W., Whitfield, M. F., Mackinnon, M., Lopes, L., Wong, G., et al. (1994). Perinatal outcomes of a large cohort of extremely low gestational age infants (twenty-three to twenty-eight completed weeks of gestation). *J. Pediatr.* 125, 952–960. doi: 10.1016/S0022-3476(05)82015-3

Thal, D., Jackson-Maldonado, D., and Acosta, D. (2000). Validity of a parent-report measure of vocabulary and grammar for Spanish-speaking toddlers. *J. Speech, Lang. Hear. Res.* 43, 1087–1100. doi: 10.1044/jslhr.4305.1087

Thorngren-Jerneck, K., and Herbst, A. (2006). Perinatal Factors Associated with Cerebral Palsy in Children Born in Sweden. *Obstetr. Gynecol.* 108, 1499–1505. doi: 10.1097/01.AOG.0000247174.27979.6b

Tulviste, T., Toome, L., Männamaa, M., and Varendi, H. (2020). Language skills at corrected age 2;0 is poorer in extremely and very preterm boys but not girls compared with their full-term peers. *Early Hum. Dev.* 151, 105164. doi: 10.1016/j.earlhumdev.2020.105164

Vohr, B. (2016). Language and hearing outcomes of preterm infants. *Semin. Perinatol.* 40, 510–519. doi: 10.1053/j.semperi.2016.09.003

Vohr, B. R. (2022). Neurodevelopmental outcomes of premature infants with intraventricular hemorrhage across a lifespan. *Semin. Perinatol.* 46, 151594. doi: 10.1016/j.semperi.2022.151594

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 14 July 2023
DOI 10.3389/fpsyg.2023.1176028



OPEN ACCESS

EDITED BY
Maria-José Ezeizabarrena,
University of the Basque Country UPV/EHU, Spain

REVIEWED BY
Irene Cadime,
University of Minho, Portugal
Dorthe Bleses,
Aarhus University, Denmark

*CORRESPONDENCE
Mårten Eriksson
✉ marten.eriksson@hig.se

RECEIVED 28 February 2023
ACCEPTED 14 June 2023
PUBLISHED 14 July 2023

CITATION
Eriksson M and Myrberg K (2023) How the communicative development inventories can contribute to clinical assessments of children with speech and language disorders.
*Front. Psychol.* 14:1176028.
doi: 10.3389/fpsyg.2023.1176028



# How the communicative development inventories can contribute to clinical assessments of children with speech and language disorders

Mårten Eriksson[1]* and Karin Myrberg[2,3]

[1]Faculty of Health and Occupational Studies, Department of Occupational Health Sciences and Psychology, University of Gävle, Gävle, Sweden, [2]Centre for Research and Development, Region Gävleborg/Uppsala University, Uppsala, Sweden, [3]Faculty of Health and Occupational Studies, Department of Caring Sciences, University of Gävle, Gävle, Sweden

**Introduction:** The purpose of the present study was to investigate whether information from the Swedish version of the Communicative Development Inventories III (SCDI-III) is informative to the Speech and Language Pathologist (SLP) when examining children with suspected speech and language disorders at a SLP unit.

**Method:** Parents to 50 children (25 girls, 25 boys, age 30–80 months) that had been referred to the local SLP unit completed the SCDI-III. Nine children came from multilingual families and 41 children came from monolingual, Swedish speaking homes. The children were diagnosed as having developmental speech disorders (12) or developmental language disorders (33). Five children were not diagnosed with any disorder.

**Results:** The results showed that the referred children performed significantly lower on scales for word production, grammar, and metalinguistic awareness, compared to a subset from the norms with a similar age and gender composition. Most children fell below the 10th percentile on word production and grammatical constructions. The intercorrelation between the three scales were in general substantial. Comparisons of children's performance on the vocabulary and grammar scales of SCDI-III, and the medical records revealed 18 cases of discordance that would have motivated further examination. The parents rated sometimes their child's vocabulary and grammar skills as higher and sometime as lower to the medical records.

**Discussion:** Limitations due to attrition and sample size were discussed. It was concluded that the SCDI-III can provide valuable information to the examination at the SLP clinic in addition to parent interviews, observations of children, and various tests, and that the potential for adapted versions would be particularly high for examinations of multilingual children.

KEYWORDS

**child health, language disorder, MB-CDI, multilingual children, speech and language therapy, language assessment**

## Introduction

The MacArthur-Bates Communicative Development Inventories (CDI, Dale et al., 1989; Fenson et al., 1993, 2007) and the Language Development Survey (LDS, Rescorla, 1989) revolutionized the study of child language three decades ago by showing that parents could give valid and reliable information on infant's and toddler's concurrent communication skills. New versions have developed from the original two designed for English American speaking infants and toddlers, including short versions and extensions for children up to 3–4 years (Fenson et al., 2007; Garcia et al., 2014; Eriksson, 2017; Tulviste et al., 2020; Cadime et al., 2021; Kas et al., 2022; Mokhtari et al., 2022; Stolt, 2023). In addition, different versions of the instruments have been adapted to over 100 different languages and dialects[1] probably making the CDI to the most widespread instrument around the world assessing child language.

Because of its easy administration, parent reports have allowed for studies of communication skills of large groups of children, revealing that the variation in young children's language skills is much larger than previously thought (Fenson et al., 1994). Ease of administration also facilitated the collection of data from atypical children spread over large areas, for example children with autism spectrum disorders (Veness et al., 2012; Miniscalco et al., 2014), children with Down syndrome (Caselli et al., 1998; Berglund et al., 2001) and preterm children (Pérez-Pereira et al., 2014; Schults and Tulviste, 2016; Pérez-Pereira and Cruz, 2018; Tulviste et al., 2020). As the internet has become an integrated part of most families' daily lives, digital versions of the CDI have been popular, and samples now often include data from several thousand children (e.g., Simonsen et al., 2014 included data from over 6,500 Norwegian children). The CDI has also been shown to have a remarkable predictive capacity. Assessments of Danish children at the age of 2 years with CDI predict academic achievement 10 years later (Bleses et al., 2016).

Despite the confirmations of earlier results and the many new findings made in child language based on parental reports, there also seems to be some limitations to their use. For example, one application is to use CDI as index tests in early screening for language difficulties. Yet, most reviews of screening studies including CDI or LDS as index test indicate that no study present a diagnostic accuracy sufficiently high for general screening (Law and Roy, 2008; Siu and US Preventive Services Task Force, 2015; Eriksson, 2022). However, there are other clinical applications for which parental reports on children's communication skills might be useful which are hitherto not fully researched.

Considering the increasing migration in the world, often due to wars or natural disasters, and the many adaptations of the CDI into different languages and dialects, the CDIs have a potential to provide a valuable source of information to Speech and Language Pathologists (SLPs) in addition to standard procedures when meeting a multilingual child at the SLP unit. The particular languages migrants speak tend to vary rather quickly which makes the task of assessing language status in children from migrant families even more challenging because there is little time to develop new instruments for the new languages. The great number of CDIs from different language communities is therefore a potential resource to be exploited.

The present study concerns how the CDI could be useful to a SLP in assessing whether a child being referred to a SLP unit has a language disorder or not. This should not be conflated with screening as the children visiting the clinic are already screened and found positive.

## The present study

The study was situated in Sweden where around one-third of all pre-school children are multilingual (Statistics Sweden, 2022). Most children, including children from migrant families, have regular contact with their local Child Health Clinic (CHC) during their first 5 years (Wallby and Hjern, 2011). At the CHCs, children's development and health are checked following a predetermined schedule at ages 2:5–3:0, 4:0 and 5:0 years. The first and second check-ups includes screening for language difficulties.

The children were referred to the SLP unit because they had failed the first screening at the CHC. This screening test consisted of five questions assessing language comprehension and an observation whether the child combines words. It was introduced and validated for 3-year old children by Westerlund and Sundelin (2000) and is a standard routine in half of Sweden. It was recently modified and validated for children 2:5 years (Nayeb et al., 2019) and for bilingual children (Nayeb et al., 2021). Bilingual children are first screened in one language (Swedish or the mother tongue, the order may vary). If they pass the first time, no further action is taken. If they fail the first time, the children are screened at a second occasion in their other language. Most children in the present study were screened at 2:5 years.

At the SLP unit, the child's expressive and receptive language were thoroughly examined using a combination of parent interview, informal assessments and formal assessments including various standardized tests in agreement with the description from the Catalise consortium (Bishop et al., 2016). The results from these examinations, which may be extended over several meetings, are documented by the SLP in the child's medical record. Multilingual children are to demonstrate difficulties in all their languages to be diagnosed with a speech or language disorder (*cf.* Bishop et al., 2016).

Children from Syria and Somalia were included as examples of migrant children. Because no standardized language tests in Syrian Arabic or Somali are available at Swedish SLP units, examinations of multilingual children in these languages rely heavily on language samples and dynamic assessments using an interpreter. Based on the SLPs examination, children with disorders are identified and diagnosed according to the International Statistical Classification of Diseases and Related Health Problems, 10th edition (WHO, ICD-10, https://www.who.int/standards/classifications/classification-of-diseases). All Services from the CHCs and SLPs in Sweden are free of charge.

We have used the Swedish version of CDI-III, SCDI-III, which is normed for Swedish speaking children 30–48 months old (Eriksson, 2017) in the present study. Younger children than 30 months old are rarely assessed at the SLP units in Sweden, and there is no CDI for older children. The SCDI-III differs in some respects from the version first introduced by Fenson et al. (2007) and early adaptations to Spanish (Guiberson and Rodrígez, 2010; Guiberson et al., 2011) and Basque (Garcia et al., 2014). The differences include a vocabulary part

1 https://mb-cdi.stanford.edu

restricted to four semantic categories with a focus on verbs and abstract nouns, see Eriksson, 2017 for details). This change was introduced to reduce ceiling effects, a problem that had afflicted the original version (Fenson et al., 2007). It draws upon earlier work on the composition of children's early lexicon which have shown that an increase in verbs comes as a second wave after an initial increase in number of nouns in many languages (Bates et al., 1994; Caselli et al., 1995; Bornstein et al., 2004; Schults and Tulviste, 2016) including Swedish (Berglund and Eriksson, 1994). The focus on verbs increased the difficulty of the scale and presumably facilitated the reporting task for the parent as only words from four semantic categories had to be searched in memory. Indeed, no ceiling effects were found after this change in the vocabulary part of SCDI-III for 4-year-old Swedish speaking children (Eriksson, 2017) or for Estonian (T. Tulviste, personal communication December 2, 2022) nor in Norwegian 4-year-olds (E. Holm, personal communication December 2, 2022), two additional languages with the same modification in the vocabulary section of CDI-III as SCDI-III. Another novel feature of all SCDI-III scales (vocabulary, grammar and the child's vocabulary metalinguistic awareness) was that data fitted best to a linear function in contrast to scales of vocabulary and grammar developed for younger children in both English and other languages including Swedish for which an exponential function gave the best fit.

Norms are central as they disclose a child's communication skills compared to those of other same-aged children. However, norms are sometimes used as a proxy for identifying children with language difficulties, for example children below the 10th percentile (Fenson et al., 2007; Eriksson, 2022). Yet, the validity of such proxy's is in general unsubstantiated, that is, there is a lack of studies showing that children performing below the 10th percentile are diagnosed with a language disorder by SLPs, and that children scoring above the 10th percentile have no language disorder. In the present study, we plot the vocabulary, grammar, and metalinguistic scores of SCDI-III from children referred to a SLP unit in relation to the 10th percentile to characterize the children's language. However, we make no claim of the 10th percentile bearing a particular significance apart from being a convenient reference in these figures. Neither should results from the present study be taken as evidence for or against the validity of a demarcation between children with and without a disorder at the 10th percentile because it only includes children with suspected language disorders (and the sample is way to small). Additionally, norms are based on group values and clinical judgments concern individual children. Moreover, judgments on children's language status should be based on studies of more than one aspect of children's communicative skills because a language disorder may distort the order in which language skills develops in a particular child (Eriksson, 2022), and performance of isolated aspects of language is not necessarily associated with functional language ability in everyday life (Spaulding et al., 2013).

The greatest potential for clinical use of parental reports is probably in relation to assessment of children from multilingual families (Freeman and Scroeder, 2022) as standardized tests in non-Swedish languages are not available at Swedish SLP units and communication with the parent often go by an interpreter. Yet, using the instruments with multilingual families might encounter new challenges, including, but not exclusively associated to the adaptation or translation of the instruments to other languages. There might also be advantages to investigate the use of SCDI-III in relation to assessment of children from monolingual (Swedish speaking) families, for example, some children perform below their actual competence on standardized tests due to shyness or lack of concentration. Therefore, the present study includes both children from monolingual Swedish-speaking using SCDI-III and multilingual families using translated or adapted versions of the SCDI-III.

To conclude, the overall aim of the present study was to investigate whether information from SCDI-III is informative to the SLP when examining children with suspected speech and language disorders at a SLP unit. The study includes both monolingual Swedish speaking children and multilingual children (Syrian Arabic and Somali). The following four research questions were asked:

1. Have children with a verified developmental speech disorder (DSD) or a verified developmental language disorder (DLD) lower scores on the three scales of SCDI-III (vocabulary, grammar and metalinguistic awareness) in comparison to normative data from a sample of same aged typical developing children?
2. Are the correlations between vocabulary, grammar, and metalinguistic skills the same for children with DSD, children with DLD and typical developing children from the norming group?
3. Do children with a DSD/DLD score below the 10th percentile on the three SCDI-III scales?
4. Is the information from the SCDI-III important to the SLP when deciding on a diagnosis, and is there any difference in this respect between children from monolingual Swedish speaking families compared to children from multilingual families?

# Method

## Participants

All preschool children being referred to a local SLP unit were eligible for inclusion. A total of 123 instruments were distributed by the SLPs to visiting parents: 13 in Syrian Arabic, 8 in Somali and 102 in Swedish. Of these, 7 in Syrian Arabic, 2 in Somali and 41 in Swedish were returned, corresponding to a response rate of 41%. The completed forms were from 25 girls and 25 boys with a median age of 54 months (range 32–80 months). The distribution of age group and gender over language is shown in Table 1.

## Instrument

The Swedish Communicative Development Inventory III (SCDI-III, Eriksson, 2017) designed for children 2 years 6 months to 4 years was used in its original Swedish form and in two preliminary versions, Syrian Arabic and Somali, respectively. The SCDI-III starts with a general question about the child's general level of communication, with six alternatives from "no words" to "long and complicated sentences" including examples. If a parent marked "no words," no more questions were applicable, and the parent was asked to hand in the form. Next follows a vocabulary list of 100 words with

TABLE 1 Age group, language, gender, and presence of language disorders in children referred to a SLP unit.

| Age group (months) | Swedish | | Arabic | | Somali | | Total |
|---|---|---|---|---|---|---|---|
| | Girls | Boys | Girls | Boys | Girls | Boys | |
| *Children with speech disorders n = 12* | | | | | | | |
| 30–36 | | | | | | | |
| 37–42 | | | | | | | |
| 43–48 | | | | | | | |
| 49–54 | | 3 | | | | | 3 |
| 55–60 | 4 | | | | | | 4 |
| 61–66 | 2 | | | | | | 2 |
| 67–80 | 2 | 1 | | | | | 3 |
| *Children with language disorders n = 33* | | | | | | | |
| 30–36 | 1 | 3 | | 1 | | 1 | 6 |
| 37–42 | 2 | 4 | | | | | 6 |
| 43–48 | 1 | | | 2 | | | 3 |
| 49–54 | 2 | 1 | 1 | 1 | | | 5 |
| 55–60 | 1 | | | | | | 1 |
| 61–66 | 3 | 2 | 1 | | 1 | | 7 |
| 67–80 | 1 | 3 | 1 | | | | 5 |
| *Children with no speech/language disorders n = 5* | | | | | | | |
| 30–36 | | | | | | | |
| 37–42 | | | | | | | |
| 43–48 | 1 | 1 | | | | | 2 |
| 49–54 | | 1 | | | | | 1 |
| 55–60 | | 1 | | | | | 1 |
| 61–66 | 1 | | | | | | 1 |
| 67–80 | | | | | | | |
| Total | 21 | 20 | 3 | 4 | 1 | 1 | 50 |

*N* = 50.

a focus on verbs and adjectives divided into four semantic categories; food related words, body related words, mental words, and emotion words. In a third section, 18 examples of grammar and sentence complexity was assessed. A fourth section contained seven questions on metalinguistic awareness. A final question concerned pronunciation.

The Syrian Arabic form of SCDI-III was developed with assistance from K. Floccia and A.G.S. Abdelwahab at the university of Plymouth, UK, who recently has published adaptation of the Words Only (short form) based on CDI and designed for children 8 to 30 months in 17 Arabic dialects (Abdelwahab et al., 2021). The development of the Somali version of SCDI-III started with a translation from the Swedish version by a professional interpreter. This translation was then back translated to Swedish by two Somali speaking SLP students in cooperation with I. Lundeborg Hammarström at Linköping University, Sweden. The grammar section of SCDI-III was reduced and included only three items illustrating use of elaborated phrases complexity in Syrian Arabic and Somali (Question 17, item 5, 8, and 9) because of difficulties comparing grammar development across languages.

## Procedure

Data was collected between Jan 2020 and Dec 2021. The parents were asked to complete the instrument at home and put it in a pre-stamped envelope addressed to the first author at the University of Gävle. All forms were returned within a 30-day period after the assessment. Hence, the ordinary SLP made their evaluation of children's language as usual, blind to the information from SCDI-III.

# Analyses

## Group level

Because SCDI-III was designed for children 30–48 months old, no child older than 48 months was included in the group comparisons. First, all children with a verified disorder were compared to the norming group. A second comparison was then carried out including only monolingual Swedish speaking children. The normative group included 1,134 children, but this group was reduced in each comparison to reflect the exact age and gender composition of the clinical groups (Table 1).

Differences between children with verified speech and language disorders and normative values on vocabulary, grammar and metalinguistic awareness were determined by one sample *t*-tests.

Associations between expressive vocabulary, grammar, and metalinguistic skills were investigated by Pearson correlations. Only Swedish speaking children were included in the correlational analysis because the long grammar scale was not included in the non-Swedish versions of SCDI-III. Significance of the difference between two correlations were determined by a Z-test (Soper, 2023). To control for substantial disassociations between expressive vocabulary and grammar in a few children that might distort the correlation on the group level, we looked for the number of children for which a parent had reported vocabulary skills in the lowest quartile together with grammar skills in the top quartile, or vice versa.

Performance in relation to the 10th percentile of the normative sample (Eriksson, 2017) were illustrated in figures depicting this reference as a solid line.

## Individual level

Medical records from the SLPs examination adjacent with the date of completion of the SCDI-III were obtained for each child and scrutinized by an experienced SLP (KM), not clinically involved with any of the participants. The notes filed under the record keywords "vocabulary" and "grammar" were examined with particular rigour alongside with the keyword summarizing the child's overall communicative ability. This information was compared with the information contained in the completed SCDI-IIIs and were qualitatively categorized as concordant or discordant. Discordant cases were categorized as higher as or lower on the vocabulary scale and the grammar scale than would be expected from the information in the medical records. Conflicting information was required in order for a case to be categorized as discordant. The metalinguistic category was excluded since the standard SLP assessment does not yield adjacent information.

## Research ethics

All participating parents consented to have data from their children included in the project and contributed actively by completing the SCDI-III. All data was treated confidentially, and the project was approved by the local ethical committee (dnr 2019–02780).

# Results

A total of 33 children were diagnosed with a developmental language disorder, DLD (F80.1, F80.2, F80.8 W, R470D) including the seven children with Arabic and two children with Somali background. Another 12 children were diagnosed with a developmental speech disorder, DSD (F80.0; R48.2B). All children with DSD were over 4 years. The disorders were quite evenly distributed over girls and boys. Five children had no speech or language disorder see Table 1.

## The clinical groups performance compared to that of a norming group

To investigate if children with DLD scored lower than children from the norming sample on the three scales from SCDI-III assessing vocabulary, grammar and metalinguistic scales, the mean value from the 15 children with DLD were compared to those of a subset of typical developing children described in Eriksson (2017) reflecting the same age and gender composition. The same analyses were also carried out including only the 11 monolingual Swedish speaking children as the validity of the CDI-III versions in Somali and Syrian Arabic has not been properly established and might therefore yield somewhat unreliable data.

## Vocabulary

The mean number of words on SCDI-III for the 15 children diagnosed with DLD and 48 months or younger was 20.47 words ($sd = 18.38$) out of 100 words. This was significantly lower than the norming value of 61.05 words taken from the norming sample of SCDI-III (Eriksson, 2017), reflecting the same age and gender composition based on 395 children, $t$ (14) = − −8.55, $p < 0.001$, $d = -2.21$, *CI* [−3,15, −1.24] as determined by a one-sample *t*-test.

Exclusion of the 4 children under 48 months with Somali or Syrian Arabic as best language yielded a very similar result. The mean number of words was 19.73 ($sd = 19.73$) as compared to a norming value of 60.70 words based on 278 children, $t$ (10) = − −8.41, $p < 0.001$, $d = -2.54$, *CI* [−3,77, −1.28] determined by a one-sample *t*-test.

## Grammar

The original grammar scale in SCDI-III was developed for Swedish and contained 18 items with a maximum score of 36. The mean score for the 11 monolingual Swedish speaking children, diagnosed with DLD and 48 months or younger, was 4.27 ($sd = 6.20$) as compared to a norming value of 22.29 based on 278 children. This difference is significantly lower, $t$(10) = − −9.64, $p < 0.001$, $d = -2.91$, *CI* [−4,28, −1.51].

Because comparisons of grammar across languages can be extremely difficult, we tried out a short grammar scale containing only three items reflecting how elaborate utterances the child typically use, that would be easier to use across languages. An example in which the parent should indicate which of two forms was most representative for the child's current speech, is "Turn on the light" or "Turn out the light so I can see." The correlation between the short and the full item scale for all children in the norming group was $r = 0.91$, $p < 0.001$, $n = 1{,}120$ (and for a selection reflecting the age and gender composition of the current sample, $r = 0.835$, $p < 0.01$, $n = 320$). The correlation in the present group was $r = 0.954$, $p < 0.001$, $n = 38$. Thus, it seems that much of the information from the full grammar scale can be captured by this short 3-item scale.

The mean score for the children with DLD (including one Arabic and one Somali child) on this short scale was 0.83, which should be compared to a norming value of 4.10 from the norming group, $t$ (11) = −6.11, $p < 0.001$, $d = -1.764$, CI [−2.670, −0.828] as determined by a one-sample *t*-test.

## Metalinguistic awareness

The mean score on the metalinguistic scale (maximum of 7) for the children diagnosed with DLD and 48 months old or younger was 1.93 ($sd = 1.49$). This was significantly lower than the norming value of 3.52 taken from the norming sample of SCDI-III (Eriksson, 2017) reflecting the same age and gender composition based on 359 children, $t(13) = -3.82$, $p = 0.002$, $d = -1.02$, $CI$ [−1.66, −0.36] as determined by a one-sample t-test.

Exclusion of the children with Somali or Syrian Arabic as best language yielded a very similar result, now with a mean of 2.09 ($sd = 1.30$) as compared to a norming value of 3.44 based on 278 children, $t(10) = -3.44$, $p = 0.006$, $d = -1.04$, $CI$ [−1,76, −0.28] as determined by a one-sample $t$-test. In sum, the children with a verified language disorder scored as a group significantly lower than a norming group with a similar age and gender composition.

## Interrelations between vocabulary, grammar and metalinguistic awareness

The correlations between vocabulary, grammar and metalinguistic awareness were high in the reference group. The correlations were also high for all the Swedish speaking children in the present group, see Table 2. Breaking the already small study group into subgroups is associated with great uncertainty. Yet, separate analyses of children with DSD, DLD, and no language disorder indicated that the associations are substantial for children with DLD and for children with no language disorder. However, another pattern with no or even negative associations between the three skills are indicated for children with DSD (Table 2). The difference in correlation between vocabulary and grammar for children with DSD and children with DLD was indeed significantly different, $z = 2.508$, $p = 0.001$, as was the difference in correlation between vocabulary and metalinguistic skills, $z = 3.271$, $p < 0.001$. To further investigate the associations between vocabulary and grammar on the individual level, we looked for the number of children for which a parent had reported vocabulary skills in the lowest quartile together on grammar skills in the top quartile, or vice versa. However, no such children were found. Hence, the low correlation between vocabulary and grammar was not caused by a few odd reports with large negative correlations.

## Do children with a DSD/DLD score below the 10th percentile on the three SCDI-III scales?

The 10th percentile from the norming sample is marked in Figures 1–3 by a solid line. The age of 48 months is marked with a vertical dotted line, and the 10th percentile to the right of this is thus an extrapolation. The 10th percentile distinguished perfectly between children with and without a language disorder 48 months or younger on the vocabulary scale. For the children 48 months or older with no speech or language disorder, two children scored below the 10th

TABLE 2 Correlations between scales measuring expressive vocabulary, grammar, and metalinguistic awareness in the Swedish speaking children referred to the SLP unit.

| | M | SD | Vocabulary | Grammar |
|---|---|---|---|---|
| *….Children from the reference group, n = 1,035–1,104* | | | | |
| Vocabulary | 65.05 | 18.06 | – | |
| Grammar | 23.53 | 9.28 | 0.785** | – |
| Metalinguistic awareness | 3.67 | 1.91 | 0.548** | 0.537** |
| *All Swedish speaking children referred to the SLP unit, n = 41* | | | | |
| Vocabulary | 55.37 | 30.24 | – | |
| Grammar | 16.17 | 11.61 | 0.773** | – |
| Metalinguistic awareness | 4.02 | 2.09 | 0.739** | 0.597** |
| *Children with speech disorders, n = 12 (mean age 61 months)* | | | | |
| Vocabulary | 78.25 | 10.11 | – | |
| Grammar | 26.50 | 5.00 | −0.202 | – |
| Metalinguistic awareness | 5.50 | 1.00 | −0.121 | −0.091 |
| *Children with language disorders, n = 24 (mean age 74 months)* | | | | |
| Vocabulary | 39.54 | 29.46 | – | |
| Grammar | 9.33 | 9.75 | 0.661** | – |
| Metalinguistic awareness | 3.29 | 2.03 | 0.794** | 0.491* |
| *Children with no speech/language disorders, n = 5 (mean age 55 months)* | | | | |
| Vocabulary | 76.40 | 15.75 | – | |
| Grammar | 24.20 | 7.12 | 0.714 | – |
| Metalinguistic awareness | 4 | 2.74 | 0.359 | 0.756 |

*$p < 0.05$, **$p < 0.01$.

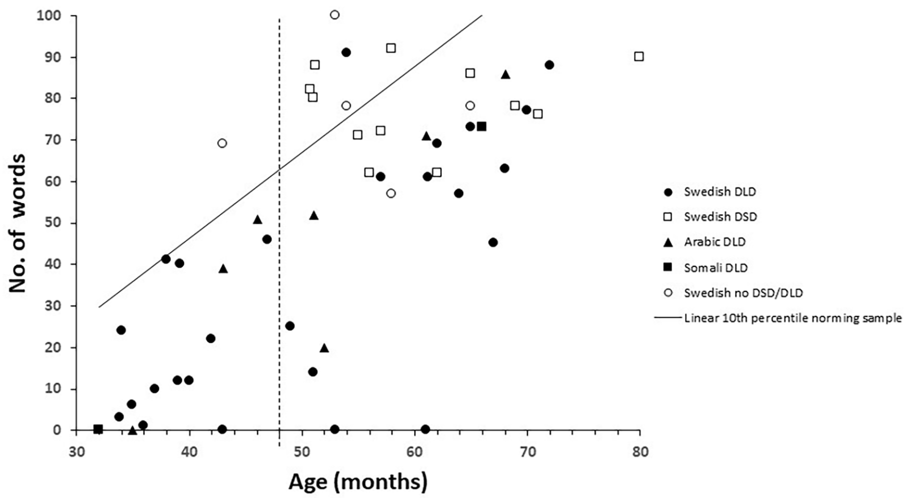


FIGURE 1
Vocabulary size plotted over age and related to the 10th percentile from the norming group in children with Developmental Language Disorders (DLD), Developmental Speech Disorders (DSD), and no disorder for children speaking Arabic, Somali, and Swedish.

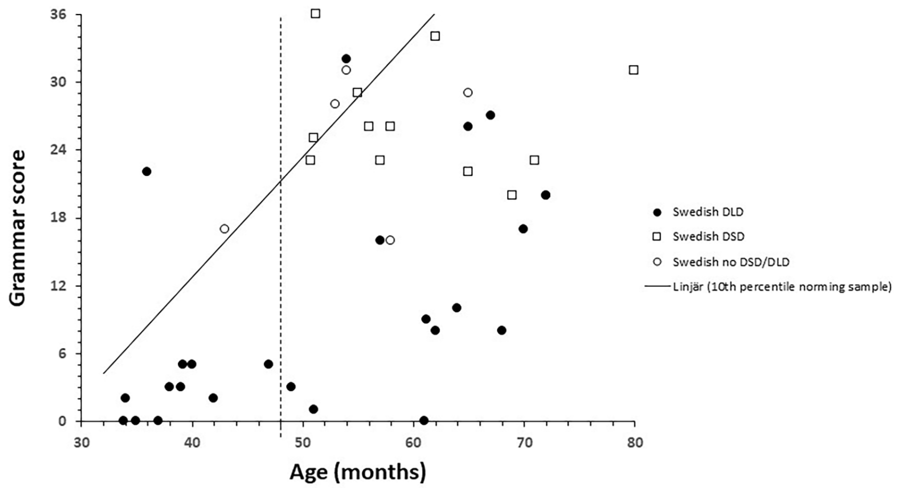


FIGURE 2
Grammar score plotted over age and related to the 10th percentile from the norming group in Swedish speaking children with Developmental Language Disorders (DLD), Developmental Speech Disorders (DSD), and no disorder.

percentile on the vocabulary scale and among the children with DSD (all 48 months or older), eight out of 12 children scored below the 10th percentile, see Figure 1. All children, except one with DLD 48 months or younger, scored below the 10th percentile on the grammar scale. One child with DLD that were older than 48 months scored above the 10th percentile while two older children without DLD scored below the 10th percentile. Among the children with DSD, nine children out of 12 scored below on the grammar scale see Figure 2. On the metalinguistic scale, slightly more than half of all children performed above the 10th percentile and three of the five children without a speech and language disorder scored below the 10th percentile see Figure 3.

In sum, the 10th percentile discriminated between children with and without a language disorder quite well on the vocabulary and the grammar scales while the 10th percentile on the metalinguistic scale was of little help in this respect. The discriminations were also best for children 48 months or younger. DSD differ from DLD in that the condition involves problems creating or forming speech sounds, not problems with expressive and/or receptive language. It was therefore surprising that a large proportion of children diagnosed with DSD were positioned below the 10th percentile on the scales.

## Is the information from the SCDI-III important to the slp when deciding on a diagnosis?

The issue of interest was whether the SCDI-III added new information to the assessment that motivated further investigation, or

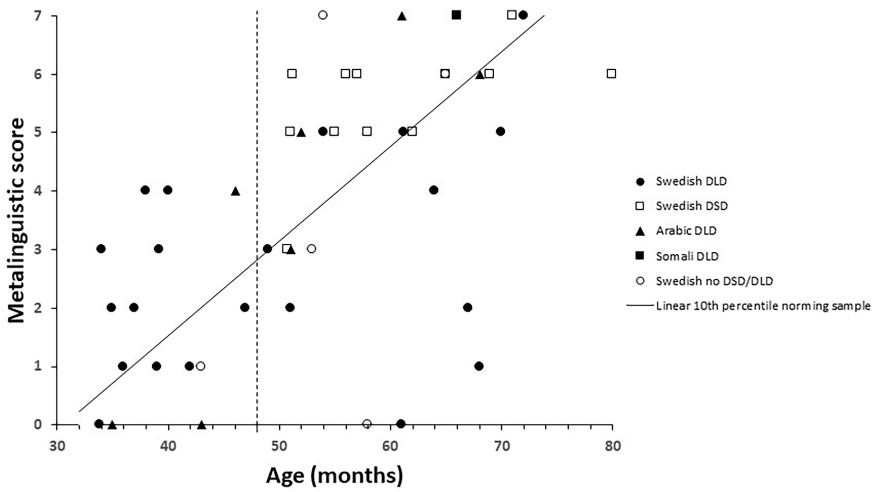


FIGURE 3
Metalinguistic score plotted over age and related to the 10th percentile from the norming group in children with Developmental Language Disorders (DLD), Developmental Speech Disorders (DSD), and no disorder for children speaking Arabic, Somali, and Swedish.

even presumptively could imply a revision of the child's diagnosis. Comparisons of the children's medical record information with the parental reports were made by qualitative analyses. We have listed the cases for which the SCDI-III provides information discordant with the medical record in Table 3. In 18 out of the 50 cases (36%), discordances were found. These cases were rather evenly distributed among gender, diagnostic group, and age span. Among the multilingual children, however, discordant results between the parent report and the medical record were observed in 6 out of 9 children (67%). An example of a discordance is child 13, diagnosed with DLD, for which the SLP reported an extremely limited expressive vocabulary and the parents report that she uses a rather large variety of the words. Here, the discordance was of the "higher" type indicating that the parents reported higher verbal skills than what was described by the SLP in the medical record. An example of the "lower" type of discordance is child 5, diagnosed with DSD, for which the parent reported much larger problems with vocabulary and grammar than was indicated in the medical records. In fact, all observed discordances including children with DSD were of this type. The review of the medical records indicates that vocabulary and grammar have been rather superficially examined in the children with DSD, with assessment primarily focusing on the production of speech sounds. The communicative skills in all six multilingual children were reported higher than assessed by the SLP (considering the child's abilities across all languages). The Arabic and Somali SCDI-IIIs are based on the child's abilities on their mother tongue. Although not formally assessed by a professional interpreter in all these cases, the medical records read that these children have none or very limited abilities on their mother tongue.

## Discussion

In this study, parent reports in the form of SCDI-III were collected from children undergoing traditional examination at a SLP unit. Although the use of CDIs at SLP units are quite widespread, there is a scarcity of studies evaluating this practice. The present study yielded four findings; (1) the referred children scored significantly lower than previously established norms for scales measuring expressive vocabulary, grammatical constructions, and metalinguistic awareness, (2) most children performed below the 10th percentile on the two former scales, (3) the interrelations between scales were high with a possible exception of children with DSD, (4) discordances between the SCDI-III and information from the medical records were found in 36% of the children.

That the referred children scored considerably lower than norms reflecting typically developing children on SCDI-III was expected and suggest that SCDI-III is a highly relevant instrument for clinical use. The term "clinical validity" is sometimes used casually for such results (eg. Lopez et al., 2023), but the term was originally used to denote the classification accuracy of a screening test (Holtzman, 1999). Evidence for the latter meaning is not demonstrated in the present study because this is not a study on screening.

The intercorrelations between the scales measuring expressive vocabulary, grammar and metalinguistic awareness were generally high. This was also expected from previous research (eg. Fenson et al., 2007; Eriksson, 2017). The correlation between vocabulary and grammar is the best documented relation. Bates and Goodman (1997) argued that a high correlation between vocabulary and grammar indicate that they operate together, in contrast to linguistic theories placing them in different modules. Furthermore, they give evidence of a high correlation between vocabulary and grammar in quite diverse populations including typically developing children, children with focal brain injury, children with Williams syndrome, children with Down syndrome, aphasic patients, and studies of on-line language processing in healthy adults. In this light, the low correlation between vocabulary and grammar in the group of children with DSD is intriguing. Although the group with DSD was small (n = 12), the difference in correlation between vocabulary and grammar for this group and children and those with DLD was significant. The main differences between children with DSD and children with DLD is that children with DSD have problems in creating or forming speech

TABLE 3 Discordances between SLP assessment and SCDI-III.

| Child | Age | Gender | Language | Vocabulary | | Grammar | |
|---|---|---|---|---|---|---|---|
| | Months | | | Higher parent rating | Lower parent rating | Higher parent rating | Lower parent rating |
| *Children with speech disorders n = 4* | | | | | | | |
| 5 | 71 | Girl | Swedish | | 1 | | 1 |
| 6 | 57 | Girl | Swedish | | 1 | | 1 |
| 31 | 62 | Girl | Swedish | | 1 | | |
| 39 | 70 | Girl | Swedish | | 1 | | 1 |
| *Children with language disorders n = 13* | | | | | | | |
| 13 | 57 | Girl | Swedish | 1 | | 1 | |
| 15 | 35 | Boy | Swedish | | 1 | | 1 |
| 16 | 66 | Girl | Swedish | 1 | | 1 | |
| 20 | 72 | Boy | Swedish | | | 1 | |
| 21 | 37 | Boy | Swedish | | 1 | | 1 |
| 22 | 61 | Girl | Swedish | 1 | | 1 | |
| 37 | 54 | Boy | Swedish | 1 | | 1 | |
| 41 | 50 | Girl | Arabic | 1 | | 1 | |
| 42 | 68 | Girl | Arabic | 1 | | 1 | |
| 44 | 46 | Boy | Arabic | 1 | | 1 | |
| 45 | 53 | Boy | Arabic | | | 1 | |
| 47 | 43 | Boy | Arabic | 1 | | 1 | |
| 49 | 67 | Girl | Somali | 1 | | 1 | |
| *Children with no speech/language disorders n = 1* | | | | | | | |
| 50 | 58 | Girl | Swedish | | 1 | | 1 |
| Total | | | | 9 | 7 | 11 | 6 |

*N* = 18.

sounds but, in contrast to children with DLD, not problems with expressive or receptive language. It can also be seen from Figures 1, 2, that the children with DSD score rather high in both vocabulary and grammar. A possible explanation to the low correlation between vocabulary and grammar might instead be related to the fact that problems in creating speech sounds is normal for toddlers, and the diagnosis DSD is therefore only given to older children. This restricts the variation in both vocabulary and grammar and the low correlation might therefore be a consequence of the low variation. More studies involving a larger group of children with DSD including a larger variation in vocabulary and grammar skills is needed to confirm this.

Complete agreement between the clinical examination and the SCDI-III would have made one of them redundant. However, discordance was fairly common, and the SCDI-III indicated sometimes that the child had more speech and language problems than recorded by the SLP, sometimes less. Thus, it was not the case that parents always rated their children's communication skills as higher than the SLP. Discordance of the opposite sort was observed for four children with DSD. The review of the medical records of these children indicated that vocabulary and grammar was rather superficially examined, probably due to referral information concerning speech-sound problems. In such cases with very specific information in the referral, SCDI-III can provide convenient information on whether to expand the examination or not.

All nine multilingual children enrolled in this study were diagnosed with DLD and discordances between the SCDI-III results and the SLP were observed in six of them. Here, all six parents reported higher linguistic competence than assessed by the SLP. These conflicting results might reflect confounding factors associated with testing multilingual children. However, conflicting information of this kind is crucial to the SLP in deciding whether a more thorough examination of a child should be undertaken or not. It might also have motivated a change in the treatment plan or in the advice given to parents.

## Strengths and limitations

Notably, this study includes both strengths and limitations. One major strength is that the study employs the SCDI-III designed for children 30–48 months old to children being referred to a SLP unit. This age range corresponds better to the age of the referrals than what CDI versions designed for younger children do. In fact, many of the referrals were older than 48 months. Another strength of the study was that the SLPs were blind to the results from SCDI-III. Hence, they followed just standard procedures when examining the children and were not influenced by the child's score on SCDI-III. This made it possible to study what SCDI-III possible could add to the standard procedure. A third strength was that the study took place in a community setting and the children had been subject to a professional SLP examination with associated ICD diagnoses.

A major limitation was that we lost control over the attrition during the pandemic. The original plan was to deliver SCDI-III consecutively to eligible children at their first visit to a SLP unit until information on 50 multilingual children (Syrian Arabic or Somali besides Swedish) were reached, and information from whatever the number of monolingual Swedish speaking children that was gathered in that time. However, priorities in the health care changed during the pandemic, and research on child language was not prioritized. Numerous appointments were redirected to telemedicine and information about the meeting was delivered by mail. Information on the current project was at first included in this information but dropped off gradually to restart again at the second half of 2021. It is also possible that attrition among the youngest children relates to a very limited expressive ability in many cases. A quick glance at the items in the questionnaire by the parent might have led to misassumptions about the target group being children with more advanced language than their child. Other reason to the attrition might reflect low parental interest in the study, little time available to complete the questionnaire, and low literacy levels in the parents. Contacts with multilingual families were also hampered by the need for an interpreter. Furthermore, a substantial attrition was associated to a small sample size. Future studies with larger study samples to validate these results, particularly for multilingual children, are warranted. To achieve this, it would probably be beneficial to ask parents to fill in the report during- or adjacent to the child's visit to the SLP unit.

## Conclusion

Overall, the results indicate that the SCDI-III would be a useful instrument in addition to parent interviews, observations of children, and standardized tests in examinations of pre-school children at the SLP clinic. In particular, further assessments are warranted when the result from SCDI-III is discordant to other results in the examination process. The potential is probably higher for multilingual children than for monolingual children, but more research including more multilingual children, and also more languages, are needed. The clinical use of SCDI-III described here should in principle be the same for adaptations to other languages. It is therefore important that CDI-IIIs are developed for more languages, and that these versions can be made available to SLP around the world from one common website.

## Data availability statement

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

## Ethics statement

The studies involving human participants were reviewed and approved by Swedish Ethical Review Authority. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

## Author contributions

ME and KM contributed to the conceptualization, methodology, and analyses of the study. ME wrote the original draft. All authors contributed to the article and approved the submitted version.


## Funding

The research was supported by the University of Gävle, Sweden.

## Acknowledgments

A preliminary version of this research, based on preliminary data and with a slightly different conclusion, was presented at the eight European Network Meeting on Communicative Development Inventories, Dubrovnik, Croatia, May, 2022. The research was supported by the University of Gävle, Sweden.


## Conflict of interest

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

## Editor's note

Maria-José Ezeizabarrena edited the article in collaboration with Melita Kovacevic, University of Zagreb, Zagreb, Croatia.

## Publisher's note



## References


Abdelwahab, A. G. S., Forbes, S., Cattani, A., Goslin, J., and Floccia, C. (2021). An adaptation of the MacArthur-Bates CDI in 17 Arabic dialects for children aged 8 to 30 months. *Lang. Learn. Dev.* 17, 425–446. doi: 10.1080/15475441.2021.1916502

Bates, E., and Goodman, J. C. (1997). On the inseparability of grammar and the lexicon: evidence from acquisition, aphasia and real-time processing. *Lang. Cog. Process* 12, 507–584. doi: 10.1080/016909697386628

Bates, E., Marchman, V., Thal, D., Fenson, L., Dale, P., and Reznick, J. S., Reilly, J., and Hartung, J. (1994). Developmental and stylistic variation in the composition of early vocabulary. *J. Child Lang.*, 21:, 85–123. doi:10.1017/S0305000900008680

Berglund, E., and Eriksson, M. (1994). "Parental reports of early language skills. Presentation of inventories and a cross-linguistic comparison" in *Proceedings from the Vth Nordic conference on child language*. ed. A. H. Wold (Oslo: Oslo University), 91–102.

Berglund, E., Eriksson, M., and Johansson, I. (2001). Parental reports of spoken language skills in children with down syndrome. *J. Speech Lang. Hear. Res.* 44, 179–191. doi: 10.1044/1092-4388(2001/016)

Bishop, D., Snowling, M., Thompson, P., Greenhalgh, T., and Consortium, C. A. T. A. L. I. S. E. (2016). CATALISE: a multinational and multidisciplinary Delphi consensus study. Identifying language impairments in children. *PLoS One* 11:e0158753. doi: 10.1371/journal.pone.0158753

Bleses, D., Makransky, G., Dale, P. S., Højen, A., and Ari, B. A. (2016). Early productive vocabulary predicts academic achievement 10 years later. *Appl. Psycholinguist.* 37, 1461–1476. doi: 10.1017/S0142716416000060

Bornstein, M. H., Cote, L. R., Maital, S., Painter, K., Park, S.-Y., Pascual, L., et al. (2004). Cross-linguistic analysis of vocabulary in young children: Spanish, Dutch, French, Hebrew, Italian, Korean, and American English. *Child Dev.* 75, 1115–1139. doi: 10.1111/j.1467-8624.2004.00729.x

Cadime, I., Santos, A. L., Ribeiro, I., and Viana, F. L. (2021). Parental reports of Preschoolers' lexical and syntactic development: validation of the CDI-III for European Portuguese. *Front. Psychol.* 12:677575. doi: 10.3389/fpsyg.2021.677575

Caselli, M. C., Bates, E., Casadio, P., Fenson, J., Fenson, L., and Sanderl, L., and, Weir, J. (1995). A cross-linguistic study of early lexical development, Cogn. Dev., 10:, 159–199. doi:10.1016/0885-2014(95)90008-X.

Caselli, M. C., Vicari, S., Longobardi, E., Lami, L., Pizzoli, C., and Stella, G. (1998). Gestures and words in early development of children with down syndrome. *J. Speech Lang. Hear. Res.* 41, 1125–1135. doi: 10.1044/jslhr.4105.1125

Dale, P. S., Bates, E., Reznick, J. S., and Morisset, C. (1989). The validity of a parent report instrument of child language at twenty months. *J. Child Lang.* 16, 239–249. doi: 10.1017/s0305000900010394

Eriksson, M. (2017). The Swedish communicative development inventory III: parent reports on language in preschool children. *Int. J. Behav. Dev.* 41, 647–654. doi: 10.1177/016502541664407

Eriksson, M. (2022). Insufficient evidence for the validity of the language development survey and the MacArthur–Bates communicative development inventories as screening tools: a critical review. *Int. J. Lang. Commun. Disord* 58, 555–575. doi: 10.1111/1460-6984.12800

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., et al. (1994). Variability in early communicative development. *Monogr. Soc. Res. Child Dev.* 59, 1–185. doi: 10.2307/1166093

Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, J. S., et alet.al. (2007). *MacArthur-Bates communicative development inventories: User's guide and technical manual*. Baltimore, MD: Brookes.

Fenson, L. D., Philip, S., Reznick, J. S., Thal, D. J., Bates, E., Hartung, J. P., et alet.al. (1993). *The MacArthur communicative development inventories. Users guide and technical manual*. San Diego: Singular Publishing Group.

Freeman, S., and Scroeder, S. R. (2022). Assessing language skills in bilingual children: current trends in research and practice. *J. Child Sci.* 12, 33–46. doi: 10.1055/s-0042-1743575

Garcia, I., Barreña, A., Ezeizabarrena, M. J., Almgren, M., Arratibel, N., and Barnes, J. (2014). Assessing the communicative development of 30 to 50-months old Basque children: the Basque version of the MacArthur-Bates CDI-III. *Psikologia* 88, 33–72. doi: 10.26876/uztaro.88.2014.3

Guiberson, M., and Rodrígez, B. L. (2010). Measurement properties and classification accuracy of two Spanish parent surveys of language development for preschool-age children. *Am. J. Speech Lang. Pathol.* 19, 225–237. doi: 10.1044/1058-0360(2010/09-0058)

Guiberson, M., Rodrígez, B. L., and Dale, P. S. (2011). Classification accuracy of brief parent report measures of language development in Spanish-speaking toddlers. *Lang. Speech Hear. Serv. Sch.* 42, 536–549. doi: 10.1044/0161-1461(2011/10-0076)

Holtzman, N. A. (1999). Promoting safe and effective genetic tests in the United States: work of the task force on genetic testing. *Clin. Chem.* 45, 732–738. doi: 10.1093/clinchem/45.5.732

Kas, B., Jakab, Z., and Lőrik, J. (2022). Development and norming of the Hungarian CDI-III: a screening tool for language delay. *Int. J. Lang. Commun. Disord.* 57, 252–273. doi: 10.1111/1460-6984.12686

Law, J., and Roy, P. (2008). Parental report of infant language skills: a review of the development and application of the communicative development inventories. *Child Adolesc. Mental Health* 13, 198–206. doi: 10.1111/j.1475-3588.2008.00503.x

Lopez, A., Caffò, A. O., Tinella, L., and Bosco, A. (2023). The four factors of mind wandering questionnaire: content, construct, and clinical validity. *Assessment* 30, 433–447. doi: 10.1177/10731911211058688

Miniscalco, C., Rudling, M., Råstam, M., Gillberg, C., and Johnels, J. Å. (2014). Imitation (rather than core language) predicts pragmatic development in young children with ASD: a preliminary longitudinal study using CDI parental reports. *Int. J. Lang. Commun. Disord.* 49, 369–375. doi: 10.1111/1460-6984.12085

Mokhtari, F., Kazemi, Y., Feizi, A., and Dale, P. (2022). Psychometric properties of the MacArthur-Bates communicative development inventories-III (CDI-III) in 30 to 37 months old Persian-speaking children. *Archives Rehabil.* 23, 372–391. doi: 10.32598/RJ.23.3.3166.3

Nayeb, L., Lagerberg, D., Sarkadi, A., Salameh, E. K., and Eriksson, M. (2021). Identifying language disorder in bilingual children aged 2.5 years requires screening in both languages. *Acta Paediatr.* 110, 265–272. doi: 10.1111/apa.15343

Nayeb, L., Lagerberg, D., Westerlund, M., Sarkadi, A., Lucas, S., and Eriksson, M. (2019). Modifying a language screening tool for three-year-old children identified severe language disorders six months earlier. *Acta Paediatr.* 108, 1642–1648. doi: 10.1111/apa.14790

Pérez-Pereira, M., and Cruz, R. (2018). A longitudinal study of vocabulary size and composition in low risk preterm children. *First Lang.* 38, 72–94. doi: 10.1177/014272371773048

Pérez-Pereira, M., Fernández, P., Gómez-Taibo, M. L., and Resches, M. (2014). Language development of low risk preterm infants up to the age of 30 months. *Early Hum. Dev.* 90, 649–656. doi: 10.1016/j.earlhumdev.2014.08.004

Rescorla, L. (1989). The Language Development Survey: a screening tool for delayed language in toddlers *J. Speech Hear. Disord.* 54, 587–599. doi: 10.1044/jshd.5404.587

Schults, A., and Tulviste, T. (2016). The very first words of Estonian children: a comparison of two parental report types. *Adv. Pediat. Res.* 2, 1–9. doi: 10.12715/apr.2015.2.15

Simonsen, H. G., Kristoffersen, K. E., Bleses, D., Wehberg, S., and Jørgensen, R. N. (2014). The Norwegian communicative development inventories: reliability, main developmental trends and gender differences. *First Lang.* 34, 3–23. doi: 10.1177/0142723713510997

Siu, A. L. and US Preventive Services Task Force (2015). Screening for speech and language delay and disorders in children aged 5 years or younger: US preventive services task force recommendation statement. *Pediatrics* 136, e474–e481. doi: 10.1542/peds.2015-1711

Soper, D. S. (2023). Significance of the difference between two correlations calculator [software]. Available at: https://www.danielsoper.com/statcalc

Spaulding, T. J., Hosmer, S., and Schechtman, C. (2013). Investigating the interchangeability and diagnostic utility of the PPVT-III and PPVT-IV for children with and without SLI. *Int. J. Speech Lang. Pathol.* 15, 153–462. doi: 10.3109/17549507.2012.762042

Stolt, S. (2023). Internal consistency and concurrent validity of the parental report instrument on language in pre-school–aged children – The Finnish Communicative Development Inventory III. *First Language.* doi: 10.1177/01427237231167301

Statistics Sweden (2022). Population in Sweden by country/region of birth, citizenship and Swedish/foreign background, [data set]. SCB statistical database. Retrieved February 1, 2023, Available at: https://www.statistikdatabasen.scb.se

Tulviste, T., Toome, L., Männamaa, M., and Varendi, H. (2020). Language skills at corrected age 2; 0 are poorer in extremely and very preterm boys but not girls compared with their full-term peers. *Early Hum. Dev.* 151:105164. doi: 10.1016/j.earlhumdev.2020.105164

Veness, C., Prior, M., Bavin, E., Eadie, P., Cini, E., and Reilly, S. (2012). Early indicators of autism spectrum disorders at 12 and 24 months of age: a prospective, longitudinal comparative study. *Autism* 16, 163–177. doi: 10.1177/1362361311399936

Wallby, T., and Hjern, A. (2011). Child health care uptake among low-income and immigrant families in a Swedish county. *Acta Paediatr.* 100, 1495–1503. doi: 10.1111/j.1651-2227.2011.02344.x

Westerlund, M., and Sundelin, C. (2000). Can severe language disability be identified in three-year-olds? Evaluation of a routine screening procedure. *Acta Paediatr.* 89, 94–100. doi: 10.1080/080352500750029149

Frontiers in Psychology

TYPE Original Research
PUBLISHED 20 July 2023
DOI 10.3389/fpsyg.2023.1188550



OPEN ACCESS

EDITED BY
Maria-José Ezeizabarrena,
University of the Basque Country UPV/EHU, Spain

REVIEWED BY
Maria Voeikova,
Institute of Linguistic Studies (RAS), Russia
Philippe Prévost,
Université de Tours, France

*CORRESPONDENCE
Jelena Kuvač Kraljević
jelena.kuvac@erf.unizg.hr

RECEIVED 17 March 2023
ACCEPTED 06 July 2023
PUBLISHED 20 July 2023

CITATION
Šmit Brleković L and Kuvač Kraljević J (2023) Parental reports on language development in toddlers and preschoolers based on the Croatian version of Communicative Development Inventories III.
*Front. Psychol.* 14:1188550.
doi: 10.3389/fpsyg.2023.1188550



# Parental reports on language development in toddlers and preschoolers based on the Croatian version of Communicative Development Inventories III

Lidija Šmit Brleković[1] and Jelena Kuvač Kraljević[2]*

[1]Postgraduate Doctoral Programme Speech, Language and Hearing Disorders, Faculty of Education and Rehabilitation Sciences University of Zagreb, Zagreb, Croatia, [2]Department of Speech and Language Pathology, Faculty of Education and Rehabilitation Sciences University of Zagreb, Zagreb, Croatia

**Introduction:** Previous studies have focused on understanding parental attempts to record language development in children, across many typologically different languages. However, many of these studies restricted their assessment to children up to the age of 3 years. The aim of this paper was to move this boundary by examining language development in typically language developed children older than 3 years.

**Methods:** Using the Croatian version of the Communicative Development Inventories III (CDI-III-HR), we investigated the contribution of parental reports of a child's lexical, grammatical, and metalinguistic awareness abilities to general language abilities assessed by clinicians. Participants included the parents of 151 children between the ages of 30 to 48 months, who completed the CDI-III-HR and reported on their child's language abilities.

**Results:** Our results show that age is significantly associated with the lexical, grammatical, and metalinguistic awareness abilities of a child's language development. These findings confirm that all three abilities increase with age and that parents can perceive changes in a child's language development. The subsections of CDI-III-HR were moderately to strongly associated with each other, with the strongest association being between lexicon and grammar, suggesting that they remain closely related after the age of 30months. Parental assessments of a child's language development are a better predictor of language production than language comprehension, with grammar making the most consistent and significant contribution.

**Discussion:** This study confirms that the development of grammatical abilities is the most prominent skill between the ages of 30 to 48 months and that parents can observe the transition in the child's language development through their usage of grammar in words to grammar in sentences. Based on the selected sample of children, we discovered different patterns of parental success in assessing the child's language ability. These findings indicate that parents can act as valuable sources of information regarding the child's language abilities, but clinical assessments of early language development should consider many other formal sources of information in addition to parental reports.

KEYWORDS

**MacArthur-Bates Communicative Development Inventories, parental reports, vocabulary, grammar, metalinguistic awareness, language development, Croatian**

## Introduction

The discussion regarding the optimal method for effectively assessing infant and toddler language development continues to be the subject of intense debate in the areas of developmental psychology and speech-language pathology for two main reasons. First, none of the methods developed so far have been able to capture the multidimensional nature of language. Second, each method has its own shortcomings (Dockrell, 2001): for example, standardised measurement instruments are very often not an appropriate method for assessing children's language in the early years of life, primarily because it is difficult to ensure the child's cooperation in a new environment and with a new person (Law and Roy, 2008).

These shortcomings can be compensated to some extent by parental reports. They have been proved to be increasingly useful as a good initial method of describing a child's language and communication abilities. Additionally, parents spend a considerable amount of time with the child and they are the most frequent interlocutors during the early years of the child's life. This allows them to observe the child's language abilities under natural conditions, as well as in a wide range of situations (Guiberson et al., 2011). Parental reports can be a valuable source of initial information, especially in cases where observational data indicate concerns about language and communication development.

Likewise, parental reports also have certain limitations. Parents may be biased and may overestimate or underestimate the child's abilities due to various clinical, educational, and social circumstances (Feldman et al., 2000; Law and Roy, 2008). For example, Jackson-Maldonado et al. (2013) showed that mothers with lower levels of education were more likely to overestimate their child's level of word understanding, especially during the early phase of development, while mothers with higher levels of education were more likely to apply a higher standard when interpreting the ability of early understanding, resulting in lower average scores in this sample of children. These limitations raise the question of the reliability of parental assessment of the child's language in the late toddler and preschool periods.

## Parental assessment of language development of children older than 30 months using MB-CDI-III

Over the past 50 years, researchers and clinicians have widely recognised and confirmed the ecological validity of standardised measures of parental reports. Parental reports of a child's language and communication abilities have become an integral part of screening and diagnostic procedures. Therefore, many standardised measurement instruments are available in order to collect specific data from parents, such as the Age and Stages Questionnaire (Bricker and Squires, 1999) or the Language Development Survey (LDS; Rescorla, 1989). However, only a few of these are designed to assess the language development of children who are more than 3 years old. For example, the Language Use Inventory (LUI; O'Neill, 2007) can be used to assess language abilities, but it is focused exclusively on specific language aspects, in this case, pragmatics.

Fenson et al. (2007) developed the MacArthur-Bates Communicative Development Inventories III (MB-CDI-III) to assess lexical and syntactic development in children aged 30 to 37 months. This version is an extension of the MacArthur-Bates Communicative Development Inventories (MB-CDI, Fenson et al., 1990, 2007). MB-CDI was originally developed in order to conduct parental assessments of broader aspects of language and communication abilities in children up to 30 months of age. This is a period during which language development is slow, thus ensuring that parents perceive the qualitative changes in the child's language. Using the MB-CDI, Dale et al. (1989) demonstrated that parental reports were accurate, valid, and reliable when parents were limited to reporting current and novel behaviors, as well as when reporting was based on a recognition format.

The MB-CDI-III consists of three subsections: vocabulary checklist (100 items), syntactic complexity (12 items), and language use (12 items). It was validated based on a sample of 356 children aged 30 to 37 months. However, ceiling effects occurred in later months covered by the instrument (34–37 months), particularly in the syntactic complexity and language use subsections (Fenson et al., 2007). Instrument showed good concurrent validity. In a study of the validity of parental reports of the language abilities of children aged 2 and 3 years, Feldman et al. (2005) found a statistically significant, moderate correlation between the scores on all three subsections of the MB-CDI-III for children aged 3 years and the scores obtained using two standardised tests: The McCarthy Scales of Children's Abilities ($r = 0.47–0.56$, $p < 0.001$) and Peabody Picture Vocabulary Test – Revised ($r = 0.41–0.49$, $p < 0.001$). In addition, the authors reported statistically significant but low correlations between the three subsections and conversational language measures in children at age three ($r = 0.26–0.42$, $p < 0.001$).

Given the many ceiling effects that the MB-CDI-III (Fenson et al., 2007) showed, especially in the upper half of the age range studied, Eriksson (2017) revised this assessment by introducing four semantic categories (food words, body words, mental words, and emotion words) into the vocabulary subsection, adding a new subsection on metalinguistic awareness, and extending the use of the instrument to 48 months of age. To our knowledge, the MB-CDI-III has been developed in nine languages (Table 1). Based on the details listed in Table 1, it can be seen that the adaptation of this instrument has been carried out in two directions – first, according to the original American inventory (as in Basque or Hungarian), and second, according to the revised Swedish inventory (as in Croatian and Estonian). Regardless of whether the languages follow the American or Swedish version, there are differences in the structure or the number of items in the newly developed versions of the CDI-III. For example, in the Portuguese inventory, there is no metalinguistic awareness subsection, or in the Estonian inventory, there are six items in pronunciation, while the Swedish inventory has only one. Third, although the instrument has broad applicability in language assessment of different populations (clinical groups, bilingual speakers, children from different social backgrounds), the authors' different original motivations for developing the instrument in one language are also evident. The Hungarian inventory, for example, was developed for screening children experiencing language delays. On the other hand, the Norwegian authors strongly emphasise the importance of the Norwegian CDI-III for multilingual speakers.

All previously published CDI-III, developed according to the American or Swedish versions, reported relatively good

TABLE 1 Overview of the CDI-III (data taken from Cadime et al., 2021 and supplemented with recent data).

| Reference | Language | Age range of target population (years) | Subsections and items | | | | |
|---|---|---|---|---|---|---|---|
| | | | Vocabulary | Syntax and morphology | Uses of language | Metalinguistic awareness | Additional items |
| Fenson et al. (2007) | English (American) Original version | 2.6–3.0 | Word checklist (100) | (a) Syntactic complexity (12) – Pairs of sentences with varying complexity (parents must flag one in each pair that most resembles what the child says) | Using language (12) – Questions on different language uses with a yes/no response | – | (a) One question on whether the child already combines words (not yet/sometimes/often)<br>(b) Mean length of utterances –Parents must list the three longest sentences that they heard from their child recently |
| Guiberson and Rodríguez (2010)* | Spanish (Pilot INV–III) | 3.0–5.2 | Word checklist (100) | (a) Syntactic complexity (12) - Pairs of sentences with varying complexity (parents must flag one in each pair that most resembles what the child says) | Using language (12) – Questions on different language uses with a yes/no response | – | (a) One question on whether the child already combines words (not yet/sometimes/often)<br>(b) Mean length of utterances –Parents must list the three longest sentences that they heard from their child recently |
| Garcia et al. (2014) | Basque | 2.6–4.2 | Word checklist (120) | (a) Syntactic complexity (29) – Pairs of sentences with varying complexity (parents must flag one in each pair that most resembles what the child says)<br>(b) Morphology – one list of suffixes (16) and one list of verbs (20) (parents should indicate the ones produced by their child) | Using language (10) – Questions on different language uses with a yes/no response | – | (a) One question on whether the child already combines words (not yet/sometimes/often)<br>(b) Mean length of utterances –Parents must list the three longest sentences that they heard from their child recently |
| Eriksson (2017) | Swedish (SCDI-III) | 2.6–4.0 | Word checklist (100) divided into four semantic categories: food words (16), body words (26), mental words (30), emotion words (28) | (a) Language complexity (10) –Pairs of sentences (parents must indicate if the child uses one of them or if both are used equally)<br>(b) Grammar (8) – questions on the use of grammar markers with a three-category frequency response scale: never/sometimes/everyday | – | Metalinguistic Awareness (7) – Questions with a yes/no response to assess phonological and orthographic awareness, as well as the awareness of the existence of other languages | (a) One question on children's general level of communication with six alternatives (e.g., not yet talking)<br>(b) Pronunciation – one question on how the child sounds compared to other children of the same age, with three response alternatives: like an age-mate, a younger, or an older child. |

*(Continued)*

TABLE 1 (Continued)

| Reference | Language | Age range of target population (years) | Subsections and items | | | | |
|---|---|---|---|---|---|---|---|
| | | | Vocabulary | Syntax and morphology | Uses of language | Metalinguistic awareness | Additional items |
| Garmann et al. (2019) | Norwegian Pilot | 2.6–4.0 | Word checklist (100) divided into four semantic categories: food words (16), body words (26), mental words (30), emotion words (28) | (a) Language complexity (10) –Pairs of sentences (parents must indicate if the child uses one of them or if both are used equally)<br>(b) Grammar (8) – questions on the use of grammar markers with a three-category frequency response scale: never/sometimes/everyday | – | Metalinguistic Awareness (7) – Questions with a yes/no response to assess phonological and orthographic awareness, as well as the awareness of the existence of other languages | (a) One question on children's general level of communication with six alternatives (e.g., not yet talking)<br>(b) Pronunciation – one question on how the child sounds compared to other children of the same age, with three response alternatives: like an age-mate, a younger, or an older child. |
| Tulviste and Schults (2020) | Estonian (ECDI-III) | 2.6–4.0** | Word checklist (101) divided into four semantic categories: food words (17), body words (26), mental words (30), emotion words (28) | (a) Language complexity (10) –Pairs of sentences (parents must indicate if the child uses one of them or if both are used alternately)<br>(b) Grammar (7) – questions on the use of grammar markers with a three-category frequency response scale: never/sometimes/everyday | – | Metalinguistic Awareness (7) – Questions with a yes/no response to assess phonological and orthographic awareness | (a) One question on children's general level of communication with six alternatives (e.g., not yet talking)<br>(b) Pronunciation – one question on how the child sounds compared to other children of the same age with three response alternatives: like an age-mate, a younger, or an older child. Five items with a yes/no response to assess specific pronunciation difficulties |
| Cadime et al. (2021) | European Portuguese (CDI-III-PT) Pilot | 2.6–4.0 | Word checklist (166) divided into four semantic categories: food words (37), body words (34), mental words (45), emotion words (50) | (a) Syntactic complexity (26) – checklist presenting examples of syntactic structures (parents must indicate yes/no if the child produces the target structure) | – | – | – |

(Continued)





10.3389/fpsyg.2023.1188550

TABLE 1 (Continued)

| Reference | Language | Age range of target population (years) | Subsections and items | | | | |
|---|---|---|---|---|---|---|---|
| | | | Vocabulary | Syntax and morphology | Uses of language | Metalinguistic awareness | Additional items |
| Kas et al. (2022) | Hungarian (HCDI-III) | 2.0–4.2 | Word checklist (124) | (a) Syntactic complexity (12) – Pairs of sentences with varying complexity (parents must flag one in each pair that most resembles what the child says)<br>(b) Morphology – one list of productive errors (12) (parents should indicate the ones produced by their child and have the possibility to add their own examples) | Using language (14) – Questions on different language uses with a yes/no response.<br>Also completed by adding two questions related to children's use of specific morphologically complex forms for asking for permission and for expressing conditional intentions. | – | (a) One question on whether the child already combines words (not yet/sometimes/often)<br>(b) Example Sentences section (Mean length of utterances) –Parents must list the three longest sentences that they heard from their child recently |
| Kuvač Kraljević et al. (n.d.) | Croatian (CDI-III-HR) | 2.6–4.0 | Word checklist (100) divided into four semantic categories: food words (16), body words (26), mental words (30), emotion words (28) | (a) Language complexity (10) –Pairs of sentences (parents must indicate if the child uses one of them or if both are used equally)<br>(b) Grammar (8) – questions on the use of grammar markers with a three-category frequency response scale: never/sometimes/everyday | – | Metalinguistic Awareness (9) – Questions with a yes/no response to assess phonological and orthographic awareness, as well as the awareness of the existence of other languages | (a) One question on children's general level of communication with six alternatives (e.g., not yet talking)<br>(b) Pronunciation – one question on how the child sounds compared to other children of the same age, with three response alternatives: like an age-mate, a younger, or an older child.<br>Five items with a yes/no response to assess specific pronunciation difficulties |

*Pilot INV–III is not an endorsed adaptation of the CDI–III, it is a translated version of the CDI–III. **So far, the study has collected data for children aged 2.10to 3.3 years ($N = 100$), but the authors aim to collect data for children in a broader age range.





10.3389/fpsyg.2023.1188550

intercorrelation between their subsections – vocabulary, grammar, and metalinguistic awareness. For example, the Hungarian CDI-III (HCD-III; Kas et al., 2022) was assessed based on a sample of 1,424 children between the ages of 24 to 50 months and showed that all variables except one – production error – were highly correlated with each other (for example, correlation between vocabulary and sentence was $r = 0.956$, $p < 0.01$). In the Swedish CDI-III (SCDI-III; Eriksson, 2017), all three subsections were significantly related to each other ($r = 0.544–0.780$, $p < 0.01$). The Estonian CDI-III (ECDI-III; Tulviste and Schults, 2020) was validated based on a sample of 100 Estonian-speaking children between the ages of 34 to 39 months. In this case, strong correlations were observed between vocabulary, grammar usage, and sentence complexity ($r = 0.71–0.88$, $p < 0.001$), but the correlation between phonological and orthographic awareness and other components was weak ($r = 0.21–0.42$, $p < 0.05$). In the European Portuguese adaptation of CDI-III (CDI-III-PT; Cadime et al., 2021), there were positive strong correlations between vocabulary and grammar ($r = 0.659$, $p < 0.001$). In Basque CDI-III (Garcia et al., 2014), assessed on a sample of 1,024 children between the ages of 30 to 50 months, all subsections were highly correlated with each other ($r = 0.59–0.81$, $p < 0.001$). In the Norwegian pilot study, a moderate correlation was found between level of communication on the one hand and vocabulary ($r = 0.483$, $p < 0.01$), grammar ($r = 0.496$, $p < 0.01$), and pronunciation ($r = 0.449$, $p < 0.01$) on the other hand. Considering these correlations with level of communication, the authors believed that it is sufficient to ask parents how their child speaks, without delving deeply into lexical and grammatical development (Garmann et al., 2019).

Among all new developed CDI-III, only four reported concurrent validity, i.e., validity that shows the extent of agreement between assessments conducted at the same time. In ECDI-III (Tulviste and Schults, 2020), concurrent validity was established with positive medium correlations between the ECDI-III components and the Estonian version of the New Reynell Developmental Language Scales subsections ($r = 0.43–0.65$, $p < 0.001$) and ECDI-II components ($r = 0.52–0.87$, $p < 0.001$). In CDI-III-PT (Cadime et al., 2021), medium to strong correlations were observed between CDI-III-PT scores and the language subsection of The Griffiths Mental Development Scales (Luiz et al., 2006). In the Basque CDI-III (Garcia et al., 2014), a medium partial correlation was found controlling for age between the components of a Basque CDI-III and the Peabody test (Dunn et al., 2006) ($r = 0.60–0.72$, $p < 0.001$). In a Norwegian pilot study (Garmann et al., 2019), 28 children were recorded participating in a 30-min spontaneous conversation and significant moderate to strong correlations were observed between the words from CDI-III and two measures related to a child's spontaneous conversation abilities: the number of different words and mean length of utterances (MLU).

## Language development in children between 2 and 4 years of age

After the age of 2 years, language abilities in children increase exponentially at all levels: not only quantitatively, but also qualitatively, in terms of complexity and the depth in the conceptual level of lexical units. For example, at 2 years, a child's lexicon contains about 250 words, while at 3 years, the lexicon contains about 1,000 words, and at 4 years about 1,600 words (Owens, 2020). Besides open-class words – nouns, verbs, and adjectives – two-year-olds, especially those with larger lexicons, have usually begun to acquire closed-class words such as articles, pronouns, prepositions, question words, and quantifiers, which are used to express grammatical meaning in sentences (Stolt, 2018). However, three-year-olds possess lexicon with more abstract words that represent different mental states. These words differ in their semantic-conceptual properties and are therefore more demanding to acquire for several reasons: their perceptual properties in relation to the referent in the child's environment are not so transparent and direct, which means that the child cannot rely on context to interpret the meaning of such words, but they have to extract it from an abstract concept. All this affects the later appearance of these words in the child's lexicon. While most concrete words such as action verbs like *walk*, *cook*, or *drink*, appear before a child's second year of life, abstract mental verbs do not appear until the third year. Some of them, especially those that are very similar in meaning (e.g., *think* and *know*), remain obscure to children until the fourth year of life (Papafragou et al., 2007).

Greater lexical knowledge is the trigger for the master of a greater number of morphological rules, and consequently, a bigger lexicon and more advanced morphology are the trigger for the production of more complex and longer syntactic structures. This relationship between expressive lexical ability and grammatical development has been reported in many monolinguistic studies in different languages (Maital et al., 2000; Stolt et al., 2009), as well as in cross-linguistic studies (Thordardottir et al., 2002; Devescovi et al., 2005; Kuvač Kraljević et al., 2021). Moreover, a cross-linguistic study involving Croatian, Estonian, and Finnish (Kuvač Kraljević et al., 2021) confirmed that two-year-old children who combined words had lexicon approximately four times larger than those who still had not yet started to combine at the same age, suggesting that lexical development can predict syntactic development. However, the trajectory of lexical and grammatical development is not always monodirectional in a way that only lexical abilities support and predict grammatical development. It is more correct to say that the relationship between grammatical and lexical development is bidirectional. As the child begins to use the grammatical system productively, it facilitates lexical growth, i.e., as the lexicon grows, it increasingly supports grammatical development (Moyle et al., 2007). The interplay between these two skills varies across the different years. Progress in the development of both abilities can sometimes occur simultaneously, but sometimes, especially in the first years of life, it can occur sequentially, which means that the development of one of these abilities is not linear (Bates and Goodman, 2001).

By age three, children begin to master many grammatical forms, but they continue to build on some that they acquired in their second year of life (for example, they continue to use negation (Tager-Flusberg, 2001), but also learn to incorporate negation into more complex forms (Reed, 2017)). Generally, after age three, children ask more and more questions and have better control on sentence-internal features such as predicate-subject-object agreement or independent and dependent relations. However, they also increase production on the discourse level through better mastering of across-sentence features. For example, although three-year-old children exhibit a lack of contextual information about the time, place, and chronological order of events when recounting a personal story (Reese et al., 2011), they can produce and tell a short story with two events (Peterson and McCabe, 1983). In short, although the period from 2 to 4 years is very variable in terms of pace and style, this period represents the stage of language development when children extend the grammar of words to sentences.

It is certainly important to observe language development from the perspective of receptive and expressive language, that is, from the connection between comprehension and production. Despite the widespread assumption that comprehension is always more advanced than production, differences in the relationship between comprehension and production in early language development are much more common than assumed. Many studies report a positive, significant, small to moderate correlation between comprehension and production, as seen in Bornstein and Hendricks's (2012) study. However, this correlation is not as simple as it first appears. For example, in a large sample of 101,250 children, ages 2;00 to 9;11, from sixteen under-researched languages, the authors found that the mean of comprehension and production varied with the child's age, reaching an asymptote at age 5;00. Thus, comprehension does not always predict production (Bauer et al., 2002). In addition, production has been found to precede comprehension for certain language phenomena, such as word order, verbal inflection or object pronouns (Hendriks, 2014). Bates (1993) also describes a dissociation between comprehension and production for typically language developing children but also in some clinical groups such as late talkers, who show enormous variability in receptive language abilities, from typical to impaired. Thus, comprehension and production can also be sometimes defined as 'dissociated' processes (Bates, 1993; Bauer et al., 2002) during language development.

After the age of three, children slowly begin to think about language not only as an object of knowledge, but also as a means of communication (Sinclair, 1986). This metalinguistic ability develops gradually and includes the knowledge that language consists of more discrete elements (e.g., phonemes, syllables, words) and can also be represented in written form. Aside from a better understanding of the inherent nature of language, metalinguistic awareness is mandatory for reading and writing at all stages of its development – from emergent and beginning to conventional literacy. Emergent literacy refers to literacy development that originates in early childhood, i.e., before the onset of formal conventional literacy instruction. Whitehurst and Lonigan (2001) emphasised that, in addition to oral language, phonological and orthographic awareness act as precursors to early literacy that correlates most strongly with later conventional reading. Phonological awareness refers to a child's ability to identify smaller language units of words such as phonemes and syllables (Anthony and Francis, 2012). Knowledge of the alphabet is one of the best predictors of reading achievement in school (Whitehurst and Lonigan, 2001). In addition to being directly related to decoding written language, letter knowledge also has an impact on phonological awareness. Another pathway of print awareness and letter knowledge is through writing, and this is called orthographic awareness. Behaviors such as pretend writing or learning to write one's own name are examples of emergent writing. Since parental linguistic input plays an important role in the development of the child's spoken language abilities, it can be assumed that it is also important in promoting metalinguistic awareness. For example, shared book reading, an activity that is part of the family literacy construct, plays a special role in the development of spoken language, especially in lexical development (Blewitt and Langan, 2016), but also in the development of emergent literacy (Justice and Kaderavek, 2002). Finally, family literacy, defined by the frequency and quality of shared reading, which can be summarised as a cultural practice and the number of books at home as an indicator of cultural capital, is known to be a good predictor of children's early and later literacy (Niklas et al., 2020).

All these determinants of language between 2 and 4 years of age – the more advanced vocabulary, complex syntax, and beginning to think about language on a metalinguistic level – raises a logical question: how do parents observe and perceive such language development, given its quantity and content? With this framework in mind, the aim of this study is to investigate the concurrent validity of CDI-HR-III by analysing and comparing parental reports on the lexical, grammatical, and metalinguistic awareness abilities of typical developing children aged 30 to 48 months with data of general language abilities of the same children based on assessments made by clinicians.

The specific objectives of the study are:

1. to examine the influence of age on lexical, grammatical, and metalinguistic abilities of children,
2. to examine the interrelationships between the different subsections of the parental report,
3. to investigate the individual contributions of lexical, grammatical, and metalinguistic awareness variables based on parental reports of the child's general language abilities, and
4. with regard to the two assessment methods – parental report and formal assessment – to examine the agreement of the language performances of those children whose language achievement were in the lower range of average performance.

# Materials and methods

## Participants and procedure

Study participants included parents of a total of 173 children, who completed the CDI-III-HR and reported on their child's lexical, grammatical, and metalinguistic abilities. Participants were recruited by speech and language pathologists (SLP) who work in the kindergartens attended by their children. After receiving consent from the parents who were willing to participate in the study, the SLPs explained how to fill out the CDI-III-HR. The participants then filled out the inventory themselves and returned the reports to the SLPs after a few days.

Once the SLPs received the completed inventories from the parents, they tested each child on the Comprehension and Production Scales of the New Reynell Developmental Language Scales (NRDLS-HR; Edwards et al., 2019) in their offices in the respective kindergartens. The analysis of the obtained data on the NRDLS-HR revealed that 22 children had below average performance on both scales – Comprehension and Production (standard scores ≤80 correspond to the 10% population of children with lowest achievement in language; see Norbury et al., 2016; Wu et al., 2023) – and were therefore excluded from further analysis. Therefore, in this study we included only those

TABLE 2 Distribution of NRDLS-HR scores obtained by the children included in the study.

| Scale | Standard score | | |
|---|---|---|---|
| | 81–90 | 91–100 | > 100 |
| NRDLS-HR Comprehension scale | 9 | 38 | 104 |
| NRDLS-HR Production scale | 22 | 20 | 109 |

children whose NRDLS-HR scores exceeded the established threshold (standard scores ≥81) i.e., those whose language performance falls within the range of typical language abilities. The distribution of the scores obtained by the children included in the study ($n = 151$) on both NRDLS-HR scales are listed in Table 2.[1]

For the final analysis, 151 children included in the study were stratified into three age groups: youngest – from 30 to 35 months ($n = 51$; 30 girls and 21 boys), middle – from 36 to 41 months ($n = 42$; 21 girls and 21 boys), and oldest – from 42 to 48 months ($n = 58$; 29 girls and 29 boys). Of the total number of children, 31% lived in the eastern part of Croatia, 29% from the Adriatic part, and 40% from the central and northern part where one/third of the entire Croatian population lives (according to the last census, Croatian Bureau of Statistics, 2021). For 96% of the children, the CDI-III-HR was completed by their mother, while for 3% of the children, it was completed by their father and for 1% of the children, by both parents together.

The study and its protocol were approved by the Ethics Committee of the Faculty of Education and Rehabilitation Sciences, University of Zagreb (approval number: 251–74/22–01/2). After guaranteeing their anonymity and dignity, informed consent was obtained from all participants. In order to collect data on language development of children from different parts of Croatia, a total of 24 SLPs participated in this study. The requirement for SLP involvement was that he or she had to have a license to use NRDLS-HR. The inclusion of SLPs in the study, the recruitment of participants at daycare centers, and the entire testing procedure were approved by the Ministry of Science and Education (MSE 533–06-21-0002).

## Adaptation of the CDI-III in Croatian

The adaptation of CDI-III-HR (Kuvač Kraljević et al., n.d.) is a continuation of the adaptation of the MB-CDI I and MB-CDI II (Kovačević et al., 2007) and the short versions of the same inventories (Kuvač Kraljević et al., 2023). The adaptation of CDI-III in Croatian began in September 2018 after receiving approval from the CDI Advisory Board and Mårten Eriksson (University of Gåvle). The Croatian version of CDI-III follows the Swedish version (Eriksson, 2017), after taking into account the peculiarities of the Croatian language, especially in the grammar part. Croatian belongs to the group of South Slavic languages, and it is highly morphologically developed with seven cases, three genders, and two numbers in the noun system, as well as seven verbal classes based on infinitive and present tense forms (Barić et al., 1997).

The CDI-III-ḪR is an instrument for parents in which they are asked to mark the words they recognise in their children's current spoken language. In addition, questions about the use of grammatical markers and the presence of metalinguistic awareness only had to be answered with *yes* or *no*. The first version of CDI-III-HR was developed in 2019 and consisted of 100 words taken from the Swedish version, with the same number of grammatical items. However, sometimes it was necessary to find suitable and equivalent substitutes for Swedish, and so completely new items were developed, reflecting the peculiarities of the Croatian language (such as verbal aspect or noun inflection). For the development of the items in the grammar and sentence complexity subsections, we performed a comprehensive review of available Croatian literature and developed the items based on empirical evidence (for example, Kovačević et al., 2009; Radić Tatar, 2013). These studies have shown that children at age of three and four mark all tenses in Croatian, include verb aspects, use more complex prepositional markers, compare adjectives, form interrogative clauses, use negation and conjunctions, as well as overgeneralise morphological rules. CDI-III-HR consisted of a list of words and pairs of sentences of different complexity.

In May 2020, we asked 45 parents of children between 30 and 48 months of age, who were also trained SLPs, to complete the CDI-III-HR and evaluate its efficacy using both parental and professional knowledge. These parents provided two types of feedback: (1) linguistic – for example, that it is necessary to give more examples for several items in the grammar subsection and (2) technical – refers to graphical solutions of subsections. For example, the structure of the sentence complexity subsection was confusing for many parents. That is why it was restructured in a way that simple and complex sentences that form a pair are listed one below the other on the left side and the scoring was on the right side (examples of this task for both Croatian and Swedish versions can be seen in Supplementary material). A pilot study of the first version of the CDI-III-HR showed a ceiling effect on a large number of words and relatively simple syntactic structures in the sentence complexity subsection.

In November 2020, we began developing a new version using the feedback from parents and data from the pilot study. Fifty additional words were selected on the basis of the Frequency Dictionary of Croatian Children's Language (Kuvač Kraljević et al., 2022) and added to the vocabulary subsection (such as *rugate se (mock), prosuti (spill), zijevati (yawn)*). The complexity of grammatical items was increased. For this reason, the standardised version contained 150 words in the vocabulary section, 16 items in the grammar-morphology section, and 14 sentence pairs in the syntax complexity section. In order to standardise the inventory, assessment reports were collected from parents of 620 children (311 girls and 309 boys) with typical language and cognitive development from all parts of Croatia in 2022, taking into account all dialectal language and regional cultural differences. Parents were encouraged to check off words from the list, even if the words did not match the child's dialect. For example, if a parent of a child from the southern part of Croatia indicated the word *frigati (fry)* for the word *pržiti* (fry), the word was accepted because both words have the same meaning. The

1 This study is part of a larger investigation examining the reliability of parental assessment as part of the Project "Standardization of CDI-III-HR and verification of parental reliability in reporting language development of children with typical language development and children with language disorder." Children with below-average language performance are not the subject of this paper but will be examined in a separate study.

norms were developed for the two-month age range to better capture changes in lexical and grammatical development.

In the end, the CDI-III-HR was developed based on the Swedish inventory, both in terms of number of items in the vocabulary and grammatical parts, as well as the distribution of word class – nouns, verbs, and adjectives. In addition to their clinical value, these comparable formats of CDI-III offered the opportunity to conduct cross-linguistic studies covering a wide range of issues analysed from the linguistic and cultural perspectives of different languages (see for example Eriksson et al., 2012; Kuvač Kraljević et al., 2021).

The final CDI-III-HR consists of the following subsections:

- Level of communication (6 items).
- Vocabulary with a total of 100 words divided into four semantic categories: food words (16 items), body words (26 items), mental words (30 items) and emotion words (28 items).
- Grammar consists of two subsections: grammar-morphology (8 items) and syntax complexity with 10 pairs of sentences.
- Metalinguistic awareness consists of 9 items related to phonological (3 items) and orthographic (6 items) awareness.
- Pronunciation consists of 6 items; one general question and five related to the child's ability to pronounce some sounds.

The Level of Communication subsection is not scored but has an exclusion criterion. Namely, if parents check one of the first two options (*He/she still does not speak.; He/she speaks, but his/her speech is unintelligible.*), they do not continue to fill out the inventory, as these choices indicate that the child is not yet using language to communicate. For each ticked item in the vocabulary and metalinguistic awareness subsection, the child receives 1 point. For the items in the pronunciation and grammar-morphology subsections, parents can choose between three options – *never*, *sometimes*, and *always* – and the point scale ranges from 0 to 2. In the sentence complexity subsection, the second sentence in each pair is more complex and receives 1 point.

Cronbach's $\alpha$ for the four vocabulary categories was: food words 0.74, body words 0.88, mental words 0.93, emotion words 0.90, and for the whole vocabulary subsection 0.97. Cronbach's $\alpha$ was 0.77 for the grammar-morphology, and 0.84 for sentence complexity. Due to the small number of items, Cronbach's $\alpha$ was the smallest for phonological awareness ($\alpha = 0.39$) and orthographic awareness ($\alpha = 0.59$) as two parts of the metalinguistic awareness subsection (Kuvač Kraljević et al., in press).

## New Reynell Developmental Language Scales (NRDLS-HR)

The SLPs who participated in this study assessed the children's language comprehension and production abilities using the Croatian version of the New Reynell Developmental Language Scales (NRDLS-HR; Edwards et al., 2019). This well-known test assesses comprehension and production of single words (nouns and verbs), morphology, and simple and complex sentences. The test has been adapted in Croatian and follows the structure of the original English version, but integrates all the peculiarities of the Croatian language, especially in the grammatical part of the test. The Comprehension Scale and the Production Scale consist of 72 items each. The test is valid for children between the ages of 2 to 7.6 years old and specific norms are available for all age groups. The norms were developed based on data collected from 791 typically developing children from different parts of Croatia and includes all dialects variations. There is a strong correlation between the Comprehension Scale and the Production Scale (Pearson correlation coefficient $r = 0.91$). The values of the reliability coefficients obtained by the split-half method (method of internal consistency) for the entire sample were 0.95 for the Comprehension Scale and 0.97 for the Production Scale. The correlation values between the two NRDLS-HR scales and the two language tests (Test for Reception of Grammar, TROG-2:HR and Peabody Picture Vocabulary Test, PPVT-III-HR) were also high, ranging from 0.74 to 0.84. The discriminant validity of the scales was verified by comparing them to a clinical sample of children with development language disorder who achieved significantly lower scores compared to children with typical language development. Measures of sensitivity and specificity were calculated. The Comprehension Scale was able to accurately identify 75% of children with a language disorder (sensitivity) and 91% of children with typical language development (specificity). The Production Scale was able to accurately identify 82% of children with a language disorder (sensitivity) and 90% of children with typical language development.

## Data analysis

A child's ability to produce a word or combine words in syntactic structures was scored with 1 point. Although the grammar-morphology subsection offers the possibility of marking the intensity of a child's use of some morphological forms on the scale – *never*, *sometimes*, *always* – here the categories *sometimes* and *always* are treated as one, which means that the entire subsection is scored with two values – 0 and 1. Standardised values, i.e., standardised scores and percentiles, are always used when analysing data from the NRDLS-HR because they ensure a clear classification of the individual's performance in relation to his or her peers.

An assessment of the distribution of all three subsections – vocabulary, grammar, and metalinguistic awareness – and their eight subsections and categories – food words, body words, mental words, emotional words, grammar-morphology, syntactic complexity, phonological and orthographic awareness – showed that most of the distributions were platykurtic (i.e.) violated one of the assumptions of normality. Only three variables – vocabulary subsection and two categories: mental words, and emotional words – met the normality assumptions in all three age groups and showed symmetric distributions ($p > 0.05$). Therefore, non-parametric analyses were performed.

First, descriptive data were calculated for all subsection of the CDI-III-HR for all three age groups individually (youngest, middle and oldest). Analysis of variance (ANOVA) was conducted to examine the effects of age on a child's language development and Spearman correlation analyses were performed to examine the associations between the different variables of the two assessment methods – parental report and formal assessment. Linear regression was used to examine whether the parental assessment of the child's lexical, grammatical, and metalinguistic awareness knowledge predicted the child's performance on formal language assessment. The predictor variables were tested *a priori* to check for the validity of the proportionality assumption and the absence of multicollinearity.

It is important to note that one participant in the youngest age group and two participants in the oldest age group had missing data for two variables – syntactic complexity and metalinguistic awareness. In addition, two participants from the middle age group and one participant from the oldest age group had missing data for pronunciation.

# Results

## Influence of age on language development

For all parts of the CDI-III-HR, we calculated the average performance of the children based on data collected from the parental reports.

### Level of communication

Of the total of 51 children in the youngest age group, there were 5 children whose parents reported that they spoke two or three words, 15 children who formed complete sentences, and 31 children who produced complex sentences. In the middle age group, the parents of 11 children indicated that they spoke in complete sentences, while 31 children were able to use complex sentences. In the oldest age group, parents reported that 5 children were able to speak in complete sentences and 53 used complex sentences.

### Pronunciation

When we considered the children in the youngest age group, four parents indicated that their children's pronunciation sounded like that of even younger children, 26 indicated that their children sounded like their peers, and 21 indicated that they sounded somewhat more advanced than their peers. Of the total of 42 children in the middle age group, 30 parents indicated that their children sounded like most of their peers and 10 indicated that their children sound somewhat more advanced than their peers. For two children, parents did not provide any information about their pronunciation. In the oldest age group, two parents indicated that their children sounded like a younger child, 36 parents indicated that their children sounded like their peers and 19 parents indicated that their children sound somewhat more advanced than their peers. For one child, parents did not provide information about his pronunciation level.

### Language subsection

Table 3 lists the average values for all subsections and categories of the three language variables corresponding to each age group. The mean values for all three variables increase with age, and this increase is most pronounced in relation to vocabulary and grammar. It is also evident that all three age groups show the same performance pattern – vocabulary and grammar showed better performance than metalinguistic awareness.

To investigate the effects of age on the language abilities of children as assessed by parents, a two-way ANOVA 3 × 3 was performed to understand the effect of age (youngest, middle, and oldest age groups) on lexical, grammatical, and metalinguistic abilities in children. The results show that age was statistically significant for all three variables. On lexical ability [$F$ (2, 148) = 16.143, $p < 0.000$], differences were observed between all three age groups – the youngest and middle age groups ($p = 0.048$), the youngest and oldest age groups ($p < 0.001$), as well as the middle and oldest age groups ($p = 0.021$). On grammatical abilities [$F$ (2, 148) = 8.159, $p < 0.000$] differences were observed between the youngest and middle age groups ($p = 0.020$), as well as the youngest and oldest age groups ($p < 0.001$) but there was no difference between middle and oldest groups ($p = 0.740$). On metalinguistic awareness [$F$ (2, 148) = 8.713, $p < 0.000$] differences were observed between the youngest and oldest age groups ($p < 0.001$), as well as the middle and oldest age groups ($p = 0.042$) but not between the youngest and middle groups ($p = 0.434$).

## Interrelationships in language variables of the CDI-III-HR

Table 4 lists the correlations among the four categories of the vocabulary subsection – food word, body words, mental words and emotional words. Medium-to-large significant correlations between all categories were obtained.

Table 5 shows the correlation between the three language subsections of the CDI-III-HR – vocabulary, grammar and metalinguistic awareness. In all three age groups, all three variables are significantly correlated, with the highest correlation coefficients observed between vocabulary and grammar.

## Individual contribution of language variables derived from parental reports to the prediction of the child's general language abilities

Before conducting the linear regression analysis, we examined the correlation between the three language variables from the CDI-III-HR – vocabulary, grammar and metalinguistic awareness – and the standardised scores on both scales of the NRDLS-HR, which were considered as a measure of general language ability.

Table 6 shows that vocabulary, grammar, and metalinguistic awareness based on the CDI-III-HR were related to language production measured using the Production Scale of NRDLS-HR, with the exception of vocabulary in the middle age group. The number of variables from CDI-III-HR, which correlated with the language comprehension measure used in the NRDLS-HR Comprehension Scale, decreased significantly with age. Comprehension was associated with vocabulary and grammar in the youngest age group, and only grammar in the middle age group. There was no association between comprehension and the three language variables in the oldest age group. Metalinguistic awareness was not related to language comprehension performance in any age group.

Linear regression analysis was performed to test whether the children's language performance assessed through parental reports significantly predicted their performance in formal assessment. Moreover, we wanted to examine the individual contribution of each variable of the CDI-III – lexicon, grammar, and metalinguistic awareness – to the prediction of the children's performance on formally assessed language comprehension and production.

TABLE 3 Descriptive statistics for language subsection of the CDI-III-HR.

| Subsection | Category | Age group | | | | | |
|---|---|---|---|---|---|---|---|
| | | youngest (30–35 months; *n* = 51) | | middle (36–41 months; *n* = 42) | | oldest (42–48 months: *n* = 58) | |
| | | Min-Max | *M* (SD) | Min-Max | *M* (SD) | Min-Max | *M* (SD) |
| Vocabulary | Food words (*n* = 16) | 5–16 | 12.20 (2.40) | 6–16 | 12.57 (2.64) | 7–16 | 13.55 (1.88) |
| | Body words (*n* = 26) | 8–26 | 17.63 (4.53) | 5–26 | 19.21 (4.68) | 14–26 | 21.60 (3.18) |
| | Mental words (*n* = 30) | 2–29 | 15.12 (7.14) | 4–30 | 18.60 (7.10) | 10–30 | 22.00 (5.69) |
| | Emotion words (*n* = 28) | 4–27 | 15.78 (5.17) | 7–28 | 18.81 (5.13) | 11–28 | 21.33 (4.66) |
| | Total (*n* = 100) | 19–95 | 60.73 (17.30) | 31–97 | 69.19 (18.11) | 47–100 | 78.48 (13.86) |
| Grammar | Morphology (*n* = 8) | 2–8 | 6.25 (1.60) | 5–8 | 7.36 (0.91) | 4–8 | 7.55 (0.80) |
| | Syntactic complexity (*n* = 10) | 0–10 | 5.02 (3.11) | 1–10 | 6.07 (2.85) | 0–10 | 6.62 (2.80) |
| | Total (*n* = 18) | 0–18 | 11.16 (4.51) | 7–18 | 13.43 (3.31) | 0–18 | 14.03 (3.56) |
| Metalinguistic awareness | Phonological awareness (*n* = 3) | 0–3 | 1.51 (0.99) | 0–3 | 1.83 (1.03) | 0–3 | 2.02 (1.00) |
| | Orthographic awareness (*n* = 6) | 0–5 | 2.55 (1.26) | 1–5 | 2.76 (1.21) | 0–6 | 3.60 (1.55) |
| | Total (*n* = 9) | 0–7 | 4.11 (1.89) | 1–8 | 4.60 (1.81) | 0–9 | 5.62 (2.17) |

M, Mean; SD, Standard deviation.

TABLE 4 Correlations between categories of the vocabulary subsection.

| Age group | Vocabulary category | Food words | Body words | Mental words | Emotions words |
|---|---|---|---|---|---|
| youngest (30–35 months; n = 51) | Food words | 1 | | | |
| | Body words | 0.599** | 1 | | |
| | Mental words | 0.676** | 0.789** | 1 | |
| | Emotions words | 0.432** | 0.758** | 0.719** | 1 |
| middle (36–41 months; n = 42) | Food words | 1 | | | |
| | Body words | 0.701** | 1 | | |
| | Mental words | 0.763** | 0.749** | 1 | |
| | Emotions words | 0.762** | 0.767** | 0.862** | 1 |
| oldest (42-48 months: n = 58) | Food words | 1 | | | |
| | Body words | 0.658** | 1 | | |
| | Mental words | 0.666** | 0.760** | 1 | |
| | Emotions words | 0.654** | 0.684** | 0.845** | 1 |

**Indicates a significant correlation at the 0.01 level (two-tailed).

The results of the linear regression analysis showed that vocabulary and grammar were statistically significant in the youngest age group (30–35 months) and only grammar was significant in the middle age group (36–41 months) (Table 7). This implies that parental reports of the child's vocabulary only up to 35 months of age and of grammar only up to 41 months of age significantly predict the child's

TABLE 5 Correlation between subsections of CDI-III-HR.

| Age group | Subsection | Vocabulary | Grammar | Metalinguistic awareness |
|---|---|---|---|---|
| youngest (30–35 months; $n = 51$) | Vocabulary | 1 | | |
| | Grammar | 0.554** | 1 | |
| | Metalinguistic awareness | 0.310* | 0.417** | 1 |
| middle (36-41 months; $n = 42$) | Vocabulary | 1 | | |
| | Grammar | 0.694** | 1 | |
| | Metalinguistic awareness | 0.457** | 0.472** | 1 |
| oldest (42-48 months: $n = 58$) | Vocabulary | 1 | | |
| | Grammar | 0.547** | 1 | |
| | Metalinguistic awareness | 0.482** | 0.442** | 1 |

**Indicates a significant correlation at the 0.01 level (two-tailed).

TABLE 6 Correlation between both scales of NRDLS-HR and all three subsections of the CDI-III-HR.

| Age group | CDI-III-HR NRDLS-HR | Vocabulary | Grammar | Metalinguistic awareness |
|---|---|---|---|---|
| Youngest (30–35 months; $n = 51$) | Comprehension scale | 0.378** | 0.463** | 0.273 |
| | Production scale | 0.472** | 0.554** | 0.435** |
| Middle (36–41 months; $n = 42$) | Comprehension scale | 0.252 | 0.390** | 0.085 |
| | Production scale | 0.276 | 0.532** | 0.356* |
| Oldest (42–48 months: $n = 58$) | Comprehension scale | 0.184 | 0.231 | 0.211 |
| | Production scale | 0.278* | 0.400** | 0.382** |

**Indicates significant correlations at the 0.01 level; *Indicates significant correlations at the 0.05 level.

comprehension abilities as determined by a formal assessment, i.e., using the NRDLS Comprehension Scale.

Although the contribution of the predictor variables to language production was very small, the regression analysis showed that all three predictors were statistically significant in the youngest age group, while grammar and metalinguistic awareness were statistically significant in the middle and oldest age groups (Table 8). This means that parental reports of the child's language, especially for grammar and metalinguistic awareness, throughout the period from 30 to 48 months, significantly predict the child's production performance as determined by formal language assessment.

## Agreement between language performances for children with scores in the lower range of average performance

For the final analysis, we selected only those children whose performance on the Comprehension Scale or Production Scale of NRDLS-HR was in the lowest 10% of the standard scores of the typical population of the range of the typical population (from 81 to 90 standard score). As presented in Table 2, we selected a total of 22 children (five children in the youngest age group, five in the middle age group, and 12 in the oldest age group). Since language development is still variable at the age of 3 and 4 years, the language performance of children whose achievement is near the 10th percentile (i.e., 80 standard score) is very sensitive and should be monitored. Therefore, we wanted to investigate how parents viewed the language development of these children.

Two parents of the children aged 30–35 months (fourth and fifth child depicted in Figure 1) overestimated their child's performance in lexical knowledge and grammar. Two parents (child no. 2 and 3) rated the child's lexical and grammatical knowledge similarly to the scores obtained on NRDLS-HR. Two parents (of child no. 2 and 3) overestimated metalinguistic awareness, which shows that this ability is difficult to assess at this age for some parents.

In the middle age group (from 36 to 41 months), two parents overestimated their child's performance (child no. 4 and 5 in Figure 2), one parent underestimated the performance (child no. 2), and the other two parents estimated their child's language performance similarly to the performance obtained on NRDLS-HR (child no. 1 and 3). In this age group, it was much easier for parents to assess metalinguistic awareness. In other words, parents did not overestimate this ability any more or less than they did with lexicon and grammar.

In the oldest age group (42 to 48 months), one parent overestimated the child's performance (child no. 1 on Figure 3) and one parent faced problems during the assessment of metalinguistic awareness (child no. 7 in Figure 3). Nearly half of the parents ($n = 5$; child no. 3 to child no. 7 in Figure 3) rated the child's language abilities significantly higher on all three variables – lexical, grammatical, and metalinguistic awareness – such that the scores between at least one of these variables exceeded one standard deviation in comparison with the child's performance on the NRDLS Production Scale. For the last five children (from child no. 8 to child no. 12 in Figure 3), the parents'

TABLE 7 Linear regression analysis to identify the factors influencing language comprehension.

| Age group | Predictors | $\beta$ | $p$ | $F$ | $p$ | $R$ | $R^2$ | $\Delta R$ |
|---|---|---|---|---|---|---|---|---|
| Youngest (30–35 months; $n=51$) | Vocabulary | 0.379 | 0.006 | 12.822 | 0.006 | 0.455 | 0.144 | 0.126 |
| | Grammar | 0.455 | <0.001 | 8.227 | <0.001 | 0.455 | 0.207 | 0.191 |
| | Metalinguistic awareness | 0.262 | 0.063 | 3.619 | 0.063 | 0.262 | 0.069 | 0.050 |
| Middle (36–41 months; $n=42$) | Vocabulary | 0.265 | 0.090 | 3.022 | 0.090 | 0.265 | 0.070 | 0.047 |
| | Grammar | 0.381 | 0.013 | 6.802 | 0.013 | 0.381 | 0.145 | 0.124 |
| | Metalinguistic awareness | 0.070 | 0.659 | 0.197 | 0.659 | 0.070 | 0.005 | −0.020 |
| Oldest (42–48 months: $n=58$) | Vocabulary | 0.202 | 0.128 | 2.393 | 0.127 | 0.202 | 0.041 | 0.024 |
| | Grammar | 0.229 | 0.084 | 3.098 | 0.084 | 0.229 | 0.052 | 0.035 |
| | Metalinguistic awareness | 0.239 | 0.070 | 3.403 | 0.070 | 0.239 | 0.057 | 0.040 |

TABLE 8 Linear regression analysis to identify the factors influencing language production.

| Age group | Predictors | $\beta$ | $p$ | $F$ | $p$ | $R$ | $R^2$ | $\Delta R$ |
|---|---|---|---|---|---|---|---|---|
| Youngest (30–35 months; $n=51$) | Vocabulary | 0.477 | <0.001 | 14.402 | <0.001 | 0.477 | 0.227 | 0.211 |
| | Grammar | 0.548 | <0.001 | 21.073 | <0.001 | 0.548 | 0.301 | 0.286 |
| | Metalinguistic awareness | 0.390 | 0.005 | 8.786 | 0.005 | 0.390 | 0.152 | 0.135 |
| Middle (36–41 months; $n=42$) | Vocabulary | 0.272 | 0.082 | 3.192 | 0.082 | 0.272 | 0.074 | 0.051 |
| | Grammar | 0.531 | <0.001 | 15.702 | 0.001 | 0.531 | 0.282 | 0.264 |
| | Metalinguistic awareness | 0.389 | 0.011 | 7.135 | 0.001 | 0.389 | 0.151 | 0.130 |
| Oldest (42–48 months: $n=58$) | Vocabulary | 0.251 | 0.057 | 3.769 | 0.057 | 0.251 | 0.063 | 0.046 |
| | Grammar | 0.315 | 0.016 | 6.184 | 0.016 | 0.315 | 0.099 | 0.083 |
| | Metalinguistic awareness | 0.356 | 0.006 | 8.123 | 0.006 | 0.356 | 0.127 | 0.111 |

assessment of the children's linguistic abilities was similar to that of the clinician.

It can be concluded that regardless of age, half of the parents were able to assess their child's language development similarly to the scores obtained by formal language assessment. Considering the three language variables, it is challenging for parents of the youngest age group to assess metalinguistic awareness.

## Discussion

Although parental reports have proven to be an effective tool to gather information about the child's language and communication development in the first three years of life, very little is known about the validity of parental assessments during the period when language becomes more lexically diverse and grammatically complex. Therefore, this study investigated the concurrent validity of parental reports of children between the ages 30 to 48 months by analysing and comparing the parental reports on language production abilities of typical developing children using the Croatian version of CDI-III with assessments of general language abilities.

First, the descriptive data from this study shows that, although it is a cross-sectional study, parents are able to recognise all three language abilities in a way that reflects the increase in children's language development after 30 months. At the same time, the highest increase was observed in the lexicon and grammar, while the smallest increase was observed in metalinguistic awareness. Despite individual variations, the age factor had a significant effect on all three abilities of children's language development, thus confirming once again that all three language abilities increase with age and that parents can perceive these developmental trajectories. In terms of the assessment method, this means that CDI-III-HR is sensitive enough to detect improvement in language development between 30 and 48 months.

When looking at lexical development based on the CDI-III-HR, it can be seen that even based on a limited and selected set of 100 words, the child's lexical development shows linear progression across all three age groups. Although we cannot talk about the lexicon size, these results are consistent with Owens (2020), who found that a child's vocabulary increases exponentially with age. The mean scores obtained in this study for the lexicon are similar to those obtained with the Swedish (Eriksson, 2017) and Estonian (Tulviste and Schults, 2020) versions of CDI-III. This suggests that the developed comparative formats of the CDI-III in different languages provide information about the similarity of lexical development over the period 30 to 48 months in languages that differ significantly typologically. Words are a building block for further grammatical development, so a lexicon that is not only quantitatively sufficient, but also qualitatively diverse is obligatory from the second year of childhood. Although it is slower than lexical development, the development of grammar also progresses with age, which is consistent with other studies confirming that children use more morphological rules and extend syntactic structures after the second and especially after the third year of life (Tager-Flusberg, 2001; Reed, 2017). The slowest increase was recorded in children's metalinguistic awareness. The same is confirmed in other languages that used CDI-III such as Swedish (Eriksson, 2017) or Estonian (Tulviste and Schults, 2020).

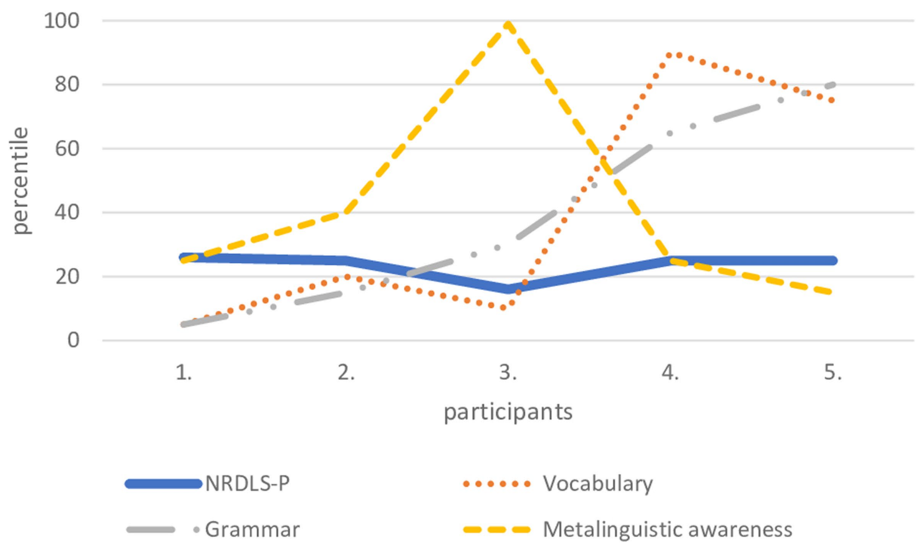


FIGURE 1
Language abilities of five children from the youngest age group obtained on the CDI-III-HR (vocabulary, grammar and metalinguistic awareness) and compared with the data obtained on the NRDLS-HR– Production scale.

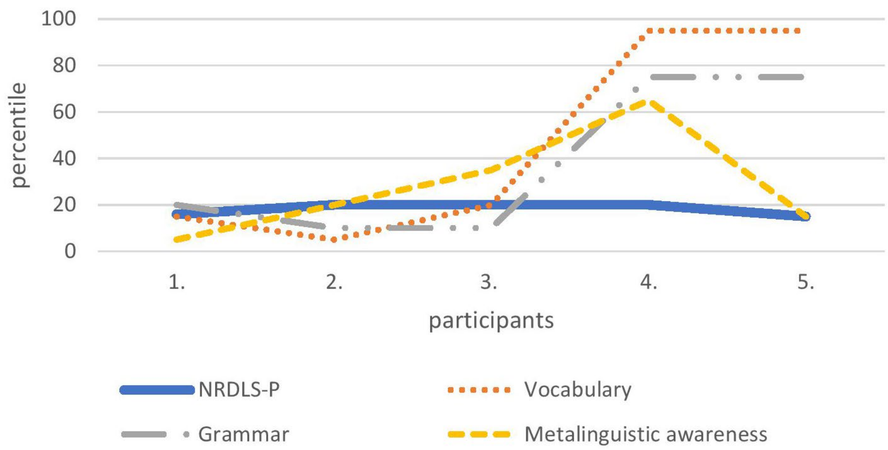


FIGURE 2
Language abilities of five children from the middle age group obtained on the CDI-III-HR (vocabulary, grammar and metalinguistic awareness) and compared with the data obtained on the NRDLS-HR– Production scale.

This finding is not surprising, since this ability is just beginning to develop at this age. Namely, for children at the age of three years, this is especially demanding because the implementation of metalinguistic awareness, especially phonological awareness, requires explicit linguistic knowledge about discrete language units. This type of knowledge cannot be extracted from the current communicative context (Sinclair, 1986; Whitehurst and Lonigan, 2001; Anthony and Francis, 2012).

Second, in all three age groups positive medium-to-large significant correlations were found among all four categories of the vocabulary subsection – food word, body words, mental words, and emotional words. The descriptive data show that children in all three age groups have the most words from the food and body words categories. Mental words were the least represented in the youngest age group but becoming more present after 36 months. Two explanations can be given for this: (1) the order of acquisition of certain semantic categories – it has been shown that words from the food category are acquired very early (in the second year of life) because they are an essential part of a child's life (Eriksson, 2017). Words from the body parts category are acquired intensively between the second and third years of life as children become more familiar with their physical features. For this reason, words that describe external body parts are acquired earlier than words that describe internal parts of the body. Words from the mental words and emotions categories are acquired after the third year of life and are an extremely important part of the child's socio-emotional development and the development of prosocial behavior (Drummond et al., 2014); (2) the concreteness of words – since the

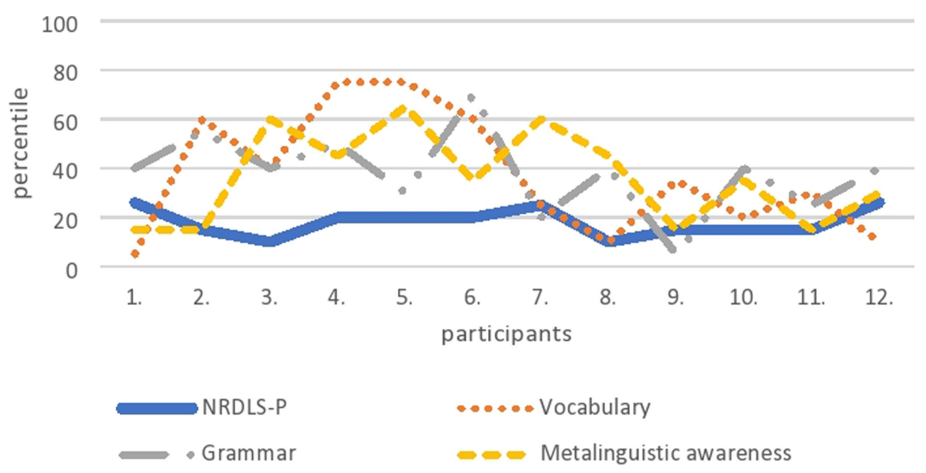


FIGURE 3
Language abilities of twelve children from the oldest age group obtained on the CDI-III-HR (vocabulary, grammar and metalinguistic awareness) and compared with the data obtained on the NRDLS-HR– Production scale.

categories of food and body consist entirely of concrete words, it is reasonable to expect words of these semantic categories to be more common in the child's early vocabulary. In the mental and emotional words categories, almost half of the words are abstract, which means they are conceptually harder for children (e.g., *believe, want, wonder*). According to Papafragou et al. (2007), a child's vocabulary before its second year of life is defined by concrete words, while abstract words appear after the third year, which is confirmed in the present study.

By examining the relationship between the different subsections of the Croatian version of CDI-III – vocabulary, grammar and metalinguistic awareness – consistent positive moderate significant correlations were observed between the vocabulary and grammar subsections through all three age groups. The same relationship in children's early language development up to the age of 30 months was confirmed in many other monolinguistic studies in different languages (Maital et al., 2000; Fenson et al., 2007; Stolt et al., 2009), as well as in cross-linguistic studies (Thordardottir et al., 2002; Devescovi et al., 2005; Kuvač Kraljević et al., 2021). This was also confirmed in studies where the CDI-III was employed to assess language development in childhood years after 30 months, for example, in Estonian (Tulviste and Schults, 2020) or Portuguese (Cadime et al., 2021). This suggests that the intertwining of expressive lexical skills and emerging grammar is a stable developmental pattern between the ages of three and four in children with typical language development.

Furthermore, the slowest pace of growth found in children's language performance in metalinguistic development as reported by the parents was also perceived in low positive significant correlations between vocabulary and metalinguistic subsections in the youngest age group. This correlation becomes more and more moderate in the middle and oldest age groups, thus confirming that metalinguistic awareness is just beginning to develop at this age and will continue to increase with age. Of course, the low values of the correlation between metalinguistic and two other language measures, which are relatively constant even in the period of 30 to 48 months, indicate a different content of knowledge that lies in the background of metalinguistic knowledge related to lexicon and grammar. It is also interesting that the metalinguistic awareness subsection has consistent positive moderate significant correlations with the grammar subsection across all three age groups, unlike its correlation with vocabulary, which is the weakest in the youngest age group. Phonological awareness, as part of metalinguistic awareness, refers to the ability to detect or manipulate the phoneme in words independent of meaning (Anthony and Francis, 2012), which means that the meaning of the word is not crucial when one thinks metaphonologically, or even metaorthographically. The Croatian language is a morphologically rich language, where morphology, for example, defines the form of words or their syntactic functions. These data lead to the conclusion that the morphological form of words begins to be closely related to the explicit knowledge of the language already at an early age. This relationship between grammar and metalinguistic awareness should be investigated more comprehensively in further studies including older preschool and school age groups.

Third, the relationship between three language domains reported by parents and two formally assessed aspects of general language ability – comprehension and production – showed that language abilities assessed by parents were more closely associated with general language productive ability than comprehension. Namely, correlations between all subsections of CDI-III-HR and NRDLS-HR Comprehension Scale rapidly decreased as age increased and they completely disappeared in the oldest age group. Significant reduction in correlation strength between comprehension and production confirms the dissociation between these two aspects of language during this period of language development. A similar result was obtained by Bornstein and Hendricks (2012) at ages 36 and 47 months, in languages very similar to Croatian (such as Serbian and Bosnian). Although asymmetries in comprehension and production development are more common in early language development (Hendriks, 2014), in this study it is shown that this can be also expected in the later toddler and preschool years. These asymmetries are highly language-related and, according to Hendriks (2014), are determined primarily by the grammar. Children sometimes produce correct sentences even though they do not know their exact meaning. For example, they can produce the correct word order, and then use that sentence structure as a basis to conclude what is the object and what is the subject. This is explained as a language-as-a-signal view (Hendriks, 2014). Since languages differ from each other in their grammatical structure, not every language will have a dissociation between comprehension and production at the same stage of language development. In our study, the regression analyses further support this finding. Only vocabulary and grammar in the youngest age group and grammar in the middle age group significantly predict comprehensive language ability. Thus, this strong correlation with formally assessed comprehension abilities indicates grammatical development, which means that there is a linear progression in the child's grammatical production observed by the parents and the child's progression in comprehension ability. Metalinguistic awareness showed no predictive values in any age group for language comprehension, confirming once again that it corresponds to different knowledge compared to language comprehension. The conducted regression analyses support these data indicating that parental reports can predict, although at a very low variance, child lexical development in the youngest age group and for a period of 1 year with respect to grammatical development. These data contain direct clinical information, so that around the age of three, parental information about the child's expressive grammar can be a reliable source of information for clinicians.

Significant positive low-to-moderate correlations between all three language domains of parental reports and general productive language ability were found in all three age groups, except with respect to vocabulary in the middle age group. Furthermore, there was a decrease in the correlations for vocabulary in the oldest age group and metalinguistic awareness in the middle and the oldest age group until they reached low significance. On the other hand, it is interesting to note that there is a consistent positive significant moderate correlation between the parental reports on their children's grammatical abilities and formal language production measures across all age groups. Grammar develops more intensively between the ages of 30 and 48 months and is therefore most noticeable to parents. However, since the number of different grammatical forms in a child's language production is not yet so great at this age, parents may notice and report all of the child's grammatical markings. The further regression analyses confirmed inconsistent predictive role of vocabulary for general productive language abilities. While the predictor for vocabulary was no longer significant in the middle and oldest age groups, the predictor for grammar and metalinguistic awareness remained significant in all age groups. Although the contribution of all significant predictor variables was small, it can still be concluded that at this age grammar has the largest contribution. The reason for this is that this is the time when the development of grammar predominates. Indeed, up to this age, children have mainly marked one– or two–word utterances morphologically. At this age, the child begins grammatical marking at the sentence level. This also means that the child begins to apply various syntactic rules of the language. The improvement of grammatical knowledge is the reason why morphology and syntax have a greater influence on general language productive abilities after the age of three.

Finally, we wanted to see how parents rated the language performance of those children whose language performance was in the lower range of average performance measured by formal language assessment. There were several reasons for choosing this target group: first, language development at ages 3 and 4 years is still variable and it is sometimes difficult to capture all the individual characteristics of each child not just for parents, but also with standardised instruments in formal language assessment; second, the language performance of children whose performance is near the 10th percentile is very sensitive and should be monitored; and third, for these two reasons, it is obvious that it is difficult to diagnose a language development disorder at age three (Bishop et al., 2017). In our selected sample, we had five children in the youngest age group, five children in the middle age group, and 12 children in the oldest age group who scored between 81 and 90 with respect to the standard score on the NRDLS-HR. In all three age groups, the same pattern was visible: half of the parents succeeded in estimating their child's language development similarly to the formal language assessment, 40% of them significantly overestimated their child's language abilities, and only about 10% underestimated them. From a clinical perspective, the 10% who underestimated their child's performance are less problematic than the 40% who overestimated their child's performance. Namely, if clinicians rely only on parental reports in language assessment, then there would certainly be some children – among those whose language abilities were overestimated by their parents – who would enter the false negative rate, i.e., those who have language difficulties, but are recognised as children of typical language development. Thus, based on this small sample, which was used only as an example to examine the success of parental assessments of language abilities, as well as the diversity of parental assessments, it is not possible to generalise parental ability to estimate child language in any direction – even if previous studies have shown that parents can do so reliably, as claimed by Dale et al. (1989) or Guiberson et al. (2011), or that they cannot make a clear assessment at all (Law and Roy, 2008). The truth about parental ability to assess the child's language lies somewhere in the middle – parents can be a valuable source of information about the child's language abilities, but these reports cannot and should not be the only source of information for clinicians. Like any other assessment, assessment of child language and communication must be comprehensive and based on a variety of assessment methods (Shipley and McAfee, 2021), so parental reports can be only one of those methods.

## Limitations and further research

This study has two limitations. The first relates to the use of the NRDLS-HR as a measure for testing external validity. The test was recently standardised in Croatian and is therefore valid and reliable. However, it provides data on general language abilities, not separately on expressive lexicon, grammar, and metalinguistic awareness skills. This indicates the importance of developing separate standardised materials in Croatian to provide more reliable data on the concurrent validity of any newly developed expressive language test. The second limitation is related to sample size for the last question, which included only 22 participants. In order to make a more meaningful statement about parental ability to assess children's language, it would be important to include more participants and expand the range of children's performances based on formal language assessment. In other words, it would be interesting to see how parents assess children who have below-average language skills according to the formal language assessment and what patterns of parental assessment can be detected in that range of distribution. However, it would be interesting to see how parents of children with developmental language disorder perceive the language abilities of their children and how the knowledge that their child has a difficulty affects the parental image of the child's language functioning.

Parental judgment is influenced by a number of socioeconomic factors (such as education, family income, inclusion in different social activities, and so on), as well as the personality characteristics of the parents themselves. Numerous studies have been conducted to define the role of these factors in different languages for different language measures in infant and toddler periods. Unfortunately, the results of these studies are contradictory, even when the studies were methodologically the same and conducted in the same age groups (Eriksson, 2017; Tulviste and Schults, 2020) or in younger age groups than those included in this study (Fenson et al., 1994; Berglund et al., 2005; Feldman et al., 2005; Rescorla et al., 2005; Nylund et al., 2021; Urm and Tulviste, 2021). Therefore, future research, using the CDI-III, should also consider these factors and examine their influence on parental reports in the phase of language development after the age of three.

## Conclusion

By conducting this study based on the Croatian version of the CDI-III, we aimed to contribute to the existing knowledge on the validity of parental reports of child language development after the age of 3 years. From the obtained data several important conclusions can be drawn.

First, these data contribute to the new evidence on parents' success in assessing their child's language in the late toddler and preschool period. In this study, parents observed the highest gains in lexicon and grammar and the lowest in metalinguistic awareness. In addition, parents observed increases in these three language skills with age, indicating that parents may perceive these developmental trajectories. Second, comparison of these data with data collected in other languages using CDI-III indicates many similarities in the timing and manner of lexical and grammatical development and development of metalinguistic awareness among languages. Third, a consistent relationship between lexical and grammatical abilities confirms that the intertwining of expressive lexical abilities and emerging grammar is a stable developmental pattern, not only in the first three years, but also between the third and fourth years of life in children with typical language development. Fourth, grammar made the largest contribution among the three predictors analysed, implying two conclusions: (a) grammar plays a prominent role in language development during this period and (b) parents may notice the child's transition in grammar development, which can be briefly described as a transition from word grammar to sentence grammar. Fifth, the slowest rate of growth found in children's language performance in metalinguistic development and the very low number of correlations between metalinguistic awareness and general language abilities indicate a different timing and nature of the development of this construct.

However, this study also has clinical significance. Parental reports can predict, albeit with very low variance, child lexical development up to 35 months and grammatical development up to 41 months. Therefore, parental information about the child's language, especially expressive grammar, can be a reliable source of information for clinicians. Nevertheless, parental reports cannot be the only source of information for SLPs or other clinicians. In fact, many other formal sources of information should be considered in addition to parent reports when clinically assessing early language development.

## Data availability statement

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

## Ethics statement

The study and its protocol were approved by the Ethics Committee of the Faculty of Education and Rehabilitation Sciences, University of Zagreb (approval number: 251–74/22–01/2). The inclusion of SLPs in the study, the recruitment of participants at daycare centres, and the entire testing procedure were approved by the Ministry of Science and Education (MSE 533–06-21-0002). Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

## Author contributions

LŠB contributed substantially to the concept and design of the study, data collection, statistical data analysis, and data interpretation. JKK contributed substantially to the development of the test, the conception of the study, data interpretation and the editing of the manuscript. All authors contributed to the article and approved the submitted version.

## Funding

This study was financially supported in part by the Faculty of Education and Rehabilitation Sciences University of Zagreb and the University of Zagreb through the University Research Grant.

## Acknowledgments

The authors would like to thank the parents and children who participated in the study and all the SLPs who helped with data collection. The authors thank the Faculty of Education and Rehabilitation Sciences, University of Zagreb for financial support of the research and Naklada Slap, and especially Valentina Ružić, PhD, for their help in data collection and analyses.

## Conflict of interest

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

## Publisher's note

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

## Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2023.1188550/full#supplementary-material

## References


Anthony, J. L., and Francis, D. (2012). Development of phonological awareness. *Curr. Dir. Psychol. Sci.* 14, 255–259. doi: 10.1111/j.0963-7214.2005.00376.x

Barić, E., Lončarić, M., Malić, D., Pavešić, S., Peti, M., Zečević, V., et al. (1997). *Hrvatska gramatika [The Croatian grammar]*. Zagreb: Školska knjiga.

Bates, E. (1993). Comprehension and Production in Early Language Development. *Monogr. Soc. Res. Child Dev.* 58, 222–242. doi: 10.1111/j.1540-5834.1993.tb00403.x

Bates, E., and Goodman, J. C. (2001). "On the inseparability of grammar and the lexicon: evidence from acquisition" in *Essential readings in developmental psychology. Language development: The essential readings*. eds. M. Tomasello and E. Bates (Malden: Blackwell Publishing), 134–162.

Bauer, D. J., Goldfield, A. B., and Reznick, J. S. (2002). Alternative approaches to analyzing individual differences in the rate of early vocabulary development. *Appl. Psycholinguist.* 23, 313–335. doi: 10.1017/.S0142716402003016

Berglund, E., Eriksson, M., and Westerlund, M. (2005). Communicative skills in relation to gender, birth order, childcare and socioeconomic status in 18-month-old children. *Scand. J. Psychol.* 46, 485–491. doi: 10.1111/j.1467-9450.2005.00480.x

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., and Greenhalgh, T.the CATALISE-2 consortium (2017). Phase 2 of CATALISE: a multinational and multidisciplinary Delphi consensus study of problems with language development: terminology. *J. Child Psychol. Psychiatry* 58, 1068–1080. doi: 10.1111/jcpp.12721

Blewitt, P., and Langan, R. (2016). Learning words during shared book reading: the role of extratextual talk designed to increase child engagement. *J. Exp. Child Psychol.* 150, 404–410. doi: 10.1016/j.jecp.2016.06.009

Bornstein, M. H., and Hendricks, C. (2012). Basic language comprehension and production in >100,000 young children from sixteen developing nations. *J. Child Lang.* 39, 899–918. doi: 10.1017/S0305000911000407

Bricker, D. D., and Squires, J. (1999). *Ages & stages questionnaires (ASQ): a parent-completed child-monitoring system*. Baltimore: Paul H. Brookes.

Cadime, I., Silva, C., Santos, A. L., Ribeiro, I., and Viana, F. L. (2021). Parental reports of preschoolers' lexical and syntactic development: validation of the CDI-III for European Portuguese. *Front. Psychol.* 12:677575. doi: 10.3389/fpsyg.2021.677575

Croatian Bureau of Statistics. *Statistical report*. (2021) Available at: https://www.dzs.hr/Hrv_Eng/publication/2016/SI-1583.pdf (Accessed January 13, 2023).

Dale, P. S., Bates, E., Reznick, J. S., and Morriset, C. (1989). The validity of a parent report instrument of child language at twenty months. *J. Child Lang.* 16, 239–249. doi: 10.1017/s0305000900010394

Devescovi, A., Caselli, M. C., Marchione, D., Pasqualetti, P., Reilly, J., and Bates, E. (2005). A crosslinguistic study of the relationship between grammar and lexical development. *J. Child Lang.* 32, 759–786. doi: 10.1017/s0305000905007105

Dockrell, J. E. (2001). Assessing language skills in preschool children. *Child Psychol. Psychiatry. Rev.* 6, 74–85. doi: 10.1111/1475-3588.00325

Drummond, J., Paul, E. F., Waugh, W. E., Hammond, S. I., and Brownell, C. A. (2014). Here, there and everywhere: emotion and mental state talk in different social contexts predicts empathic helping in toddlers. *Front. Psychol.* 5:361. doi: 10.3389/fpsyg.2014.00361

Dunn, Ll. M., Dunn, Ll. M., and Arribas, D. (2006). TVIP. *Test de vocabulario en imágenes, Peabody. Adaptación española del PPVT-III*. Madrid: TEA Ediciones.

Edwards, S., Letts, C., Sinka, I., Kuvač Kraljević, J., Kologranić Belić, L., Hržica, G., et al. (2019). *Nove Reynell razvojne jezične ljestvice – hrvatska verzija (NRDLS-HR) [New Reynell Developmental Language Scales – NRDLS]*. Jastrebarsko/Zagreb: Naklada Slap.

Eriksson, M. (2017). The Swedish communicative development inventory III: parent reports on language in preschool children. *Int. J. Behav. Dev.* 41, 647–654. doi: 10.1177/0165025416644078

Eriksson, M., Marschik, P. B., Tulviste, T., Almgren, M., Pérez Pereira, M., Wehberg, S., et al. (2012). Differences between girls and boys in emerging language skills: evidence from 10 language communities. *Br. J. Dev. Psychol.* 30, 326–343. doi: 10.1111/j.2044-835X.2011.02042.x

Feldman, H. M., Dale, P. S., Campbell, T. F., Colborn, D. K., Kurs-Lasky, M., Rockette, H. E., et al. (2005). Concurrent and predictive validity of parent reports of child language at ages 2 and 3 years. *Child Dev.* 76, 856–868. doi: 10.1111/j.1467-8624.2005.00882.x

Feldman, H. M., Dollaghan, C. A., Campbell, T. F., Kurs-Lasky, M., Janosky, J. E., and Paradise, J. L. (2000). Measurement properties of the MacArthur communicative development inventories at ages one and two years. *Child Dev.* 71, 310–322. doi: 10.1111/1467-8624.00146

Fenson, L., Dale, P. S., Reznik, S. J., Bates, E., Thal, D. J., and Pethick, S. J. (1994). Variability in early communicative development. *Monogr. Soc. Res. Child Dev.* 59, i–173. doi: 10.2307/1166093

Fenson, L., Marchman, V., Thal, D., Dale, P., Reznick, S., and Bates, E. (2007). *MacArthur-Bates communicative development inventories – Users guide and technical manual*. Baltimore, MD: Paul H. Brookes Publishing Co.

Fenson, L., Thal, D., and Bates, E. (1990) *Normed values for the "early language inventory" and three associated parent report forms for language assessment*. San Diego: San Diego State University.

Garcia, I., Barreña, A., Ezeizabarrena, M. J., Almgren, M., Arratibel, N., and Barnes, J. (2014). Haur euskaldunen komunikazio garapena neurtzen 30-50 hilabete bitartean: MacArthur-Bates CDI-III tresnaren euskal bertsioa. [assessing the communicative development of 30 to 50-month old Basque children: the Basque version of the MacArthur-Bates CDI-III]. *Uztaro* 88, 33–72. doi: 10.26876/uztaro.88.2014.3

Garmann, N. G., Romøren, A. S. H., Flygstad, N., Milder, N., Pedersen, I., Simonsen, H. G., et al. (2019). *Språkkartlegging i norske barnehager -en pilotundersøkelse av foreldres og barnehageansattes rapportering av norsktalende tre-åringers språkferdigheter ved hjelp av CDI III*. RASK: Internationalt tidsskrift for sprog og kommunikation, vol. 49, 87–103.

Guiberson, M., and Rodríguez, B. L. (2010). Measurement properties and classification accuracy of two Spanish parent surveys of language development for preschool-age children. *Am. J. Speech Lang. Pathol.* 19, 225–237. doi: 10.1044/1058-0360(2010/09-0058).6-549

Guiberson, M., Rodriguez, B. L., and Dale, P. S. (2011). Classification accuracy of brief parent report measures of language development in Spanish-speaking toddlers. *Lang. Speech Hear. Ser. Sch* 42, 536–549. doi: 10.1044/0161-1461(2011/10-0076)

Hendriks, P. (2014). *Asymmetries between language production and comprehension*. Dordrecht: Springer.

Jackson-Maldonado, D., Marchman, V. A., and Fernald, L. C. (2013). Short-form versions of the Spanish MacArthur-Bates communicative development inventories. *Appl. Psycholinguist.* 34, 837–868. doi: 10.1017/S0142716412000045

Justice, L. M., and Kaderavek, J. (2002). Using shared storybook reading to promote emergent literacy. *Teach. Except. Child.* 34, 8–13. doi: 10.1177/004005990203400401

Kas, B., Jakab, Z., and Lőrik, J. (2022). Development and norming of the Hungarian CDI-III: a screening tool for language delay. *Int. J. Lang. Commun. Disord.* 57, 252–273. doi: 10.1111/1460-6984.12686

Kovačević, M., Jelaska, Z., Kuvač Kraljević, J., and Cepanec, M. (2007). *Komunikacijske razvojne ljestvice – KORALJE [Communicative Development Inventories]*. Jastrebarsko/Zagreb: Naklada Slap.

Kovačević, M., Palmović, M., and Hržica, G. (2009). "The acquisition of case, number and gender in Croatian" in *Development of nominal inflection in first language acquisition: A cross-linguistic perspective*. eds. U. Stephany and M. D. Voeikova (Berlin: Mouton De Gruyter), 153–177.

Kuvač Kraljević, J., Blaži Ostojić, A., Schults, A., Tulviste, T., and Stolt, S. (2021). Influence of internal and external factors on early language skills: a cross-linguistic study. *Infant Behav. Dev.* 63:101552. doi: 10.1016/j.infbeh.2021.101552

Kuvač Kraljević, J., Cepanec, M., and Kovačević, M. (2023). *Komunikacijske razvojne ljestvice – kratki oblici (KORALJko) [Communicative Development Inventories - short versions]*. Jastrebarsko/Zagreb: Naklada Slap.

Kuvač Kraljević, J., Hržica, G., and Štefanec, V. (2022). *Čestotni rječnik hrvatskog dječjeg jezika: Natuknice [frequency dictionary of Croatian children's language: Lemmas]*. Jastrebarsko/Zagreb: Naklada Slap.

Kuvač Kraljević, J., Matić Škorić, A., and Eriksson, M. (n.d.) *Komunikacijske razvojne ljestvice– III [communicative development inventories – III]*. (in press). Jastrebarsko/Zagreb: Naklada Slap.

Law, J., and Roy, P. (2008). Parental report of infant language skills: a review of the development and application of the communicative development inventories. *Child Adolesc. Ment. Health* 13, 198–206. doi: 10.1111/j.1475-3588.2008.00503.x

Luiz, D., Barnard, A., Knoesen, N., Kotras, N., Horrocks, S., McAlinden, P., et al. (2006). *Griffiths Mental Development Scales—Extended Revised: Two to Eight Years: Administration Manual*. Oxford, UK: Hogrefe.

Maital, S. L., Dromi, E., Sagi, A., and Bornstein, M. H. (2000). The Hebrew communicative development inventory: language specific properties and cross-linguistic generalizations. *J. Child Lang.* 27, 43–67. doi: 10.1017/S0305000999004006

Moyle, M. J., Ellis, W. S., Evans, J. L., and Lindstrom, M. J. (2007). Longitudinal relationships between lexical and grammatical development in typical and late-talking children. *J. Speech Lang. Hear. Res.* 50, 508–528. doi: 10.1044/1092-4388(2007/035)

Niklas, F., Wirth, A., Guffler, S., Drescher, N., and Ehmig, S. C. (2020). The home literacy environment as a mediator between parental attitudes toward shared reading and children's linguistic competencies. *Front. Psychol.* 11:1628. doi: 10.3389/fpsyg.2020.01628

Norbury, C. F., Gooch, D., Wray, C., Baird, G., Charman, T., Simonoff, E., et al. (2016). The impact of nonverbal ability on prevalence and clinical presentation of language disorder: evidence from a population study. *J. Child Psychol. Psychiatry* 57, 1247–1257. doi: 10.1111/jcpp.12573

Nylund, A., Ursin, P. A., Korpilahti, P., and Rautakoski, P. A. (2021). Vocabulary growth in lexical categories between ages 13 and 24 months as a function of the Child's sex, child, and family factors. *Front. Commun.* 6:709045. doi: 10.3389/fcomm.2021.709045

O'Neill, D. K. (2007). The language use inventory for young children: a parent-report measure of pragmatic language development for 18– to 47-month-old children. *J. Speech Lang. Hear. Res.* 50, 214–228. doi: 10.1044/1092-4388(2007/017)

Owens, R. E. (2020). *Language development: an introduction* (*10th*). Cranbury, NJ: Pearson education.

Papafragou, A., Cassidy, K., and Gleitman, L. (2007). When we think about thinking: the acquisition of belief verbs. *Cognition* 105, 125–165. doi: 10.1016/j.cognition.2006.09.008

Peterson, C., and McCabe, A. (1983). *Developmental psycholinguistics: Three ways of looking at a child's narrative*. New York: Plenum.

Radić Tatar, I. (2013). Ovladanost vrstama rečenica na kraju predškolske dobi [acquisition of the sentence types at the end of preschool age]. *Lahor* 16, 165–188.

Reed, V. A. (2017). *Introduction to children with language disorders* (*5th*). Boston, MA: Pearson.

Reese, E., Haden, C. A., Baker-Ward, L., Bauer, P., Fivush, R., and Ornstein, P. A. (2011). Coherence of personal narratives across the lifespan: a multidimensional model and coding method. *J. Cogn. Dev.* 12, 424–462. doi: 10.1080/15248372.2011.587854

Rescorla, L. (1989). The language development survey: a screening tool for delayed language in toddlers. *J. Speech Hear. Disord.* 54, 587–599. doi: 10.1044/jshd.5404.587

Rescorla, L., Bernstein, R. N., Jusczyk, P., and Jusczyk, A. M. (2005). Concurrent validity of the language development survey (LDS): associations with the Macarthur–Bates communicative development inventories (CDI): words and sentences. *Am J. Speech-Lang. Pathol.* 14, 156–163. doi: 10.1044/1058-0360(2005/016)

Shipley, K. G., and McAfee, J. G. (2021). *Assessment in speech-language pathology: a resource manual* (*6th edition*). San Diego, CA: Plural Publishing.

Sinclair, A. (1986). "Metalinguistic knowledge and language development" in *Knowledge and language*. eds. I. Kurcz, G. W. Shugar and J. H. Danks (Holland: Elsevier Science Publishers), 609–627.

Stolt, S. (2018). "Early lexicon and the development that precedes it and the development that follows – a developmental view to early lexicon" in *Handbook of communication disorders: Theoretical, empirical, and applied linguistic perspectives*. eds. A. Bar-On and D. Ravid (Boston;Berlin: De Gruyter Mouton), 91–110.

Stolt, S., Haataja, L., Lapinleimu, H., and Lehtonen, L. (2009). Associations between lexicon and grammar at the end of the second year. *J. Child Lang.* 36, 779–806. doi: 10.1017/S0305000908009161

Tager-Flusberg, H. (2001). "Putting words together: morphology and syntax in the preschool years" in *The development of language*. ed. J. Berko-Gleason. *5th* ed (Boston: Allyn and Bacon), 162–212.

Thordardottir, E. T., Weismer, S. E., and Evans, J. L. (2002). Continuity in lexical and morphological development in Icelandic and English-speaking 2-year-olds. *First Lang.* 22, 3–28. doi: 10.1177/014272370202206401

Tulviste, T., and Schults, A. (2020). Parental reports of communicative development at the age of 36 months: the Estonian CDI-III. *First Lang.* 40, 64–83. doi: 10.1177/0142723719887313

Urm, A., and Tulviste, T. (2021). Toddlers' early communicative skills as assessed by the short form version of the Estonian MacArthur-Bates communicative development inventory II. *J. Speech Lang. Hear. Res.* 64, 1303–1315. doi: 10.1044/2020_JSLHR-20-00201

Whitehurst, G. J., and Lonigan, C. J. (2001). "Emergent literacy: development from prereaders to readers" in *Handbook of early literacy research*. eds. S. B. Neuman and D. K. Dickinson (New York: The Guilford Press), 11–29.

Wu, S., Zhao, J., de Villiers, J., Liu, X. L., Rolfhus, A., Sun, X., et al. (2023). Prevalence, co-occurring difficulties, and risk factors of developmental language disorder: first evidence for mandarin-speaking children in a population-based study. *Lancet Reg. Health West. Pac.* 34:100713. doi: 10.1016/j.lanwpc.2023.100713

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 20 July 2023
DOI 10.3389/fpsyg.2023.1206949



OPEN ACCESS

EDITED BY
Maria-José Ezeizabarrena,
University of the Basque Country UPV/EHU, Spain

REVIEWED BY
Katharina Korecky-Kröll,
University of Vienna, Austria
Sophie Kern,
UMR5596 Dynamique du Langage, France

*CORRESPONDENCE
Susanna Surakka
✉ susanna.surakka@helsinki.fi

RECEIVED 16 April 2023
ACCEPTED 16 June 2023
PUBLISHED 20 July 2023

CITATION
Surakka S, Vehkavuori S-M, Saaristo-Helin K and Stolt S (2023) What kind of information do early parental report instruments provide on language ability at 3;6 when used at 2;0? A longitudinal comparison study.
*Front. Psychol.* 14:1206949.
doi: 10.3389/fpsyg.2023.1206949



# What kind of information do early parental report instruments provide on language ability at 3;6 when used at 2;0? A longitudinal comparison study

Susanna Surakka*, Suvi-Maria Vehkavuori, Katri Saaristo-Helin and Suvi Stolt

Department of Psychology and Logopedics, University of Helsinki, Helsinki, Finland

**Introduction:** Various parental report instruments are available for assessing children's language skills at the end of the second year. However, comparison studies on their usability are lacking, and it is also open to question what kind of information the instruments provide when used in a parallel manner. This longitudinal study investigated which of the available three parental report instruments, when used at 2;0 (year;month), provides the most representative information on language development at 3;6. In addition, since most of the parental report instruments available focus specifically on expressive language, the role of receptive language ability was also investigated when analyzing the explanatory value of parental report instruments.

**Methods:** The participants were 68 typically developing children. At 2;0, language skills were measured using the following measures: the Infant-Toddler Checklist of the Communication and Symbolic Behavior Scales Developmental Profile (ITC), the Short Form and Long Form versions of the Finnish Communicative Development Inventories (FinCDI-SF, FinCDI-LF), and the Reynell Developmental Language Scales III (RDLS). The outcome measures were receptive/expressive/general language ability at 3;6 measured using RDLS.

**Results:** The results of parental report instruments were significantly and positively associated with language ability at 3;6. The correlation between the combined value of ITC and FinCDI-SF and later language ability was stronger than correlations for each measure separately. The regression models with the results of parental report instruments as predictors explained 18–22% ($p < 0.00$) of the variability in the total RDLS score. However, when receptive language ability at 2;0 was included in the models as a predictor, R2 increased considerably (46–48%, $p < 0.00$).

**Discussion:** The results adduce the usability of parental report measures along with the importance of measuring receptive language skills at 2 years of age. In summary, this study provides important insights into the clinical evaluation of early language ability.

KEYWORDS

**parental report instrument, language development, receptive language, early screening, language assessment, communicative development inventories, CSBS infant-toddler checklist, longitudinal study**

## 1. Introduction

Parental report instruments are useful for investigating young children's language abilities, identifying children with delays and providing valid information on early language ability (Fenson et al., 1993, 2000a, b, 2007; Wetherby and Prizant, 2002; Law and Roy, 2008; Wallace et al., 2015). Various parental report instruments, such as the Infant-Toddler Checklist (ITC) and different forms of Communicative Developmental Inventories (CDIs), are available for evaluating 2 year-old children's communication and language skills (Wetherby and Prizant, 2002; Fenson et al., 2007). Still, to our knowledge, comparison studies on their usability are lacking. Thus, it is unclear, e.g., which measures provide the most representative information on emerging language capacity, including their predictive validity. Further, since different parental report instruments focus on different types of language and communication skills, such as vocabulary comprehension, gestures, production of sounds and words, and grammar, they may provide more comprehensive information when used together than separately. However, previous studies have not focused on this issue. In addition, most parental report instruments used to measure early language ability focus on expressive language, and the role of receptive language remains unclear. In all, further comparison information is needed on the usability of parental report instruments that measure the early language development of children at the end of the second year of life. The main aim of the present study was to investigate and compare the usability of the following three parental report instruments in a longitudinal setting: the *Infant-Toddler Checklist* of the Communication and Symbolic Behavior Scales Developmental Profile (ITC) and the *Short Form* and *Long Form* versions of the Finnish Communicative Development Inventories (FinCDI-SF, FinCDI-LF).

The Infant-Toddler Checklist (ITC) is a brief parental report instrument that can be used to screen prelinguistic and early language skills from 6 to 24 months (Wetherby and Prizant, 2002). It is a part of broader early social communication assessments called the Communication and Symbolic Behavior Scales Developmental Profile (Wetherby and Prizant, 2002). The ITC can be used to measure three developmental areas: social communication, expressive speech, and symbolic. These three composites include seven language predictors: emotion and eye gaze, communication, and gestures (social); sounds and words (speech); understanding and object use (symbolic). By evaluating these early language predictors together, it is expected to get valid information for early identification of delayed language skills even before spoken language becomes the primary communication method (Wetherby et al., 2002; Wetherby and Prizant, 2002; Watt et al., 2006; Eadie et al., 2010; Laakso et al., 2011; Wallace et al., 2015; Määttä et al., 2016; Stolt and Vehkavuori, 2018; Fäldt et al., 2021). The assessment of multiple prelinguistic skills may provide a broad overview of a child's general language development both at the measurement point and over time (Wetherby et al., 2002; Wetherby and Prizant, 2002; Määttä, 2017; Borkhoff et al., 2022; Nurse et al., 2022). Results from earlier studies support the validity of ITC as a measure of prelinguistic and early language skills (Wetherby and Prizant, 2002; Wetherby et al., 2003). Later studies have also found that the ITC is a useful clinical tool for screening and predicting later language ability (Crais, 2011; Wallace et al., 2015; Määttä et al., 2016). Early language predictors measured using ITC between ages 1;0 and 2;0 have been found to be associated with receptive and expressive language outcomes at 2;0 and 3;0, in which the ITC explained 20–51% of the variance (Wetherby and Prizant, 2002; Wetherby et al., 2003). A previous study has also shown that measures of social communication between 18 to 21 months predict language outcomes at 2 and 3 years of age even better than expressive vocabulary production measures at 2;0 (Morgan et al., 2020). Regarding the Finnish language, comparable findings have been reported (Laakso et al., 2011; Määttä et al., 2016). The original American version has been translated and validated in Finnish with minor adaptations (Laakso et al., 2011). The norming study for the FinITC ($n = 508$ children) indicated significant, positive associations between the Speech and Symbolic composites when measured at 2;0 and language skills at 3;0 $r$-values ranged from 0.31 to 0.48; (Laakso et al., 2011). In addition, significant associations were found in later studies up to the age of 8, and the explanatory value was reported to range from 10.5 to 53.3% (Määttä et al., 2016).

The Communicative Development Inventories (CDIs) are parental report instruments that can be used to assess children's language development between the ages of 8 and 30 months (Fenson et al., 2000b, 2007). Different forms of CDIs can be used to assess children's early language development, including vocabulary comprehension, production, gestures, and grammar (for a review, see Law and Roy, 2008). A tool for slightly older children between 2;6 and 4 years is also available: the CDI III (Eriksson, 2017; Eriksson and Myrberg, 2023; Marchman et al., 2023; Stolt, 2023). The CDIs have been initially developed in American English. Adaptations have been made in almost 100 languages (The MacArthur-Bates Communicative Development Inventories, 2023). The original American Short Form version of the CDI (CDI-SF) includes three versions: an infant form for children between 8 and 18 months and two Toddler forms for children between 16 and 30 months (Fenson et al., 2000b). The measure has been adapted for various languages (e.g., for Spanish, Jackson-Maldonado et al., 2013; for Portuguese, Frota et al., 2016; for Swedish, Eriksson, 2017; for Finnish, Stolt and Vehkavuori, 2018; for Basque, Ezeizabarrena and Fernández, 2022), and it has been used for both clinical and research purposes (Fenson et al., 2000a; Dale et al., 2003; Pan et al., 2004; Can et al., 2013; Vehkavuori and Stolt, 2018, 2019; Lasorsa et al., 2021; Sansavini et al., 2021; Urm and Tulviste, 2021). Previous research has shown that the early expressive lexicon measured using the brief screening method, the short form version of CDI, is a valid predictor of later language skills. Early expressive lexical skills, when measured using the short-form versions of CDI, have been found to significantly explain variation in receptive language skills at 3;0 (Pan et al., 2004) and vocabulary, syntax, and semantics in children aged 5;6 to 6;8 (Can et al., 2013). The Short Form version of the CDI has also been adapted for the Finnish population (Stolt and Vehkavuori, 2018), the target population in this study. The FinCDI-SF has two different versions: an Infant Form for children aged 9–18 months and a Toddler Form for children aged 18–24 months. In previous studies, results for the FinCDI-SF Toddler version have been reported to be comparable to studies in other languages (Vehkavuori and Stolt, 2019; Vehkavuori et al., 2021). The results of the Toddler version were associated broadly with subsequent language skills, such as receptive and expressive language (Vehkavuori and Stolt, 2019) and lexical, phonological, morphological, and pre-literacy skills at 5;0 (Vehkavuori et al., 2021). Moreover, especially expressive lexicon at 1;6 and 2;0 explained 16–22% of the variation in general language ability at 5;0 (Vehkavuori et al., 2021).

The original American Long Form version of the Communicative Development Inventories (CDI-LF) can be used to collect information on different language domains, such as receptive and expressive vocabulary, gesture use, and syntactical skills (Fenson et al., 1993). The CDI-LF includes a Words and Gestures Form (WG, 8–16 months) and a Words and Sentences Form (WS, 16–30 months). The CDI-LF Words and Sentences Form used in the present study assesses early lexical ability, usage of words, emerging morphosyntactical skills, and sentence length. Different language versions of CDIs have been adapted to the language in question, and due to this, their content differs from each other. For example, the English form includes the complexity of the child's multi-word utterances (Fenson et al., 2007), and the Finnish form includes the child's usage of inflections (Lyytinen, 1999). The CDI-LF has been adapted for numerous languages (for a comparison study, see Bleses et al., 2008) and used widely (e.g., Fenson et al., 1993; Marchman and Martine-Sussmann, 2002; Stolt et al., 2009; Torppa et al., 2010; Eriksson et al., 2012; Simonsen et al., 2014; Marjanovič-Umek et al., 2017; Viana et al., 2017; Pérez-Pereira and Cruz, 2018; Cadime et al., 2019; Patrucco-Nanchen et al., 2019). As a result, there is plenty of evidence of the validity and usability of different languages in longitudinal studies (Fenson et al., 2007; Hurtado et al., 2014; Jago et al., 2023). For example, expressive vocabulary at 1;10 has been reported as a strong predictor of total vocabulary at 2;6 (Pérez-Pereira and Cruz, 2018), and early vocabulary is a good predictor of grammar acquisition, and it relates to the development of early communicative gestures (Marjanovič-Umek et al., 2017). Also, previous studies have reported associations between CDI-LF scores at 2;0 and different tests at 3;0 (Feldman et al., 2005; Korpilahti et al., 2016). Correlations between scores on vocabulary production and three longest utterances at 2;0 and scores on standardized tests at 3;0 have been reported to range from 0.32 to 0.39 (Feldman et al., 2005). The Long Form version of the CDI has been validated and normed for the Finnish population (Lyytinen, 1999) and used in longitudinal studies (e.g., Lyytinen et al., 2005; Stolt et al., 2014; Joensuu et al., 2021). One of these longitudinal studies reported that weak lexical skills at 2;6 indicate weak expressive language skills at 5;6 (Lyytinen et al., 2005). In addition, the risk for weak subsequent language skills increased if weak concurrent expressive language skills were accompanied by weak early receptive language skills (Lyytinen et al., 2005). Also, a previous longitudinal study showed that early weak language skills at 2;0 predicted later weak language skills at 5;0 in prematurely born children with very low birth weight (Stolt et al., 2014). Moreover, a recent follow-up study found significant associations between language skills at 2 years of corrected age and literacy skills at 7 years, also in preterm-born children (Joensuu et al., 2021). It was found that the small lexicon size assessed with the help of FinCDI-LF and the short mean length of the three longest utterances correlated significantly with literacy measures ($r$-values varied between 0.31 and 0.43).

Most 2 year-old children use words for communication, but variation is vast. At 2;0, children have acquired the basic lexicon of their native language, and the average lexicon size varies between 200 and 400 words (Fenson et al., 2007; Bleses et al., 2008; Stolt et al., 2008). Still, variation between individual children is extensive: some children have acquired only some words, whereas others have lexicons of over 600 words. Word combinations appear between 18 and 20 months, and grammatical development follows lexical development. Therefore, age 2;0 is a prominent age point for assessing especially lexical development. Roughly 90% of children use at least two-word combinations at 2;0, and some may use very long sentences (Fenson et al., 2000b, 2007; Stolt et al., 2009). Expressive vocabulary size is strongly associated with grammatical development (Conboy and Thal, 2006; Stolt et al., 2009). Further, weak language ability is often identified at 2;0 by assessing lexicon size with parental report instruments, such as CDIs (Feldman et al., 2005; Desmarais et al., 2008; Law and Roy, 2008). One criterion for weak language ability at this age is fewer than 50 expressive words in the lexicon or the lack of word combinations (Rescorla et al., 2005; Zubrick et al., 2007; Dollaghan, 2013; Hawa and Spanoudis, 2014; Farabolini et al., 2023). The other commonly used criterion is that the child's expressive vocabulary size remains under the 10th percentile of the population in question (Girolametto et al., 2001; Heilmann et al., 2005; Desmarais et al., 2008; Rescorla and Dale, 2013). In the present study, both cut-off values are used.

The main aim of the present study is to compare and investigate the usability of three parental report instruments. The research questions were as follows: (1) Which one of the following parental report instruments when used at 2;0 has the strongest associations with receptive/expressive/total language ability when measured using the Reynell Developmental Language Scales (RDLS) at 3;6: Communication and Symbolic Behavior Scales, Developmental Profile, Infant-Toddler Checklist (ITC), the Short Form version of the Communicative Development Inventories (FinCDI-SF), the combined value of the ITC and FinCDI-SF or the Long Form version of the Communicative Development Inventories (FinCDI-LF)? (2) Does the receptive/expressive/general language ability at 3;6 differ between those children with weak vs. typical skills measured using the ITC/FinCDI-SF/FinCDI-LF at 2;0? (3) How does early receptive language, measured using RDLS at 2;0, contribute to the explanatory value of early parental report instruments: the ITC, FinCDI-SF, and FinCDI-LF?

# 2. Materials and methods

## 2.1. Participants

The participants were 68 (30 boys) typically developing, full-term (mean gestational week 40, SD 1.5), monolingual Finnish-speaking children. The families were invited to the study during a periodic health check-up at their local healthcare center at the age of 8 months. When the families were invited to the study, the participants had no diagnosis or suspicion of neurological disorders, such as hearing impairment, autism spectrum disorder, cerebral palsy, or cognitive delay. The parents were not known to have mental health issues or alcohol or drug abuse. All the parents had completed at least 9 years of compulsory schooling (Table 1).

This study is part of the Norming and Validation Study of Finnish Short Form Versions of the Communicative Development Inventories (FinCDI-SF, Sanaseula Study; principal investigator: the last author of the present article). The FinCDI-SF study has been approved by the Ethical Committee of the University of Turku. Each family signed written consent after being informed about the study. Parents received written information on their child's language skills at both assessment points. Parents were instructed to contact their local child health clinic if a child had delayed language skills.

TABLE 1 Education level of parents.

| | Maternal, *n* (%) | Paternal, *n* (%) |
|---|---|---|
| Compulsory school | 0 (0) | 1 (1) |
| High school or vocational school | 15 (22) | 27 (40) |
| University of Applied Sciences degree | 23 (33) | 12 (18) |
| University degree | 31 (45) | 26 (39) |

Information about one father's education was missing. One of the participants had same-gender parents, whose education level information is under "maternal."

## 2.2. Measures at 2;0 and 3;6

The data were collected at two age points: 2;0 and 3;6. The following parental report instruments were used at 2;0: Communication and Symbolic Behavior Scales Developmental Profile, Infant-Toddler Checklist (ITC, Laakso et al., 2011; original version: Wetherby and Prizant, 2002), the Finnish Short Form version of Communicative Development Inventories (FinCDI-SF, Stolt and Vehkavuori, 2018; original version: Fenson et al., 2000a, b) and the Finnish Long Form version of the Communicative Development Inventories (FinCDI-LF, Words and Sentences form, Lyytinen, 1999; original version: Fenson et al., 1993, 2007). At 2;0 and 3;6, a formal test, the Reynell Developmental Language Scales III, was used (RDLS, Kortesmaa et al., 2001; original version: Edwards et al., 1997). All the measures have been normed and validated in the Finnish population.

The ITC consists of three composites: social communication (emotion and eye gaze, communication, gestures; 26 points), speech (sounds, words; 14 points), and symbolic (understanding, object use; 17 points) (Wetherby and Prizant, 2002; Laakso et al., 2011). The 24 items are rated on a 3–5-point scale. The maximum total score is 57 points. The cut-off value for the total score at 2;0 (i.e., weak skills) is ≤49 points, the lowest 10th percentile of the norming group for the measure. Typical development is defined as >49 points. The cut-off value for social communication is ≤21 points, for speech ≤12 points, and for symbolic ≤15 points (Laakso et al., 2011).

The FinCDI-SF Toddler questionnaire includes a wordlist of 100 words and one additional question (max 2 points: 0 = not yet, 1 = sometimes, 2 = often) about the child's word combination usage (Stolt and Vehkavuori, 2018). Different lexical categories are represented in it, e.g., social-pragmatic words, nouns, and verbs. The relative share of the categories is parallel to that of the Finnish Long Form version. The FinCDI-SF is a briefer method compared to the Long Form CDI and is, therefore, more suitable for screening purposes. The cut-off value for weak expressive lexical skills at 2;0 is ≤12 words, the lowest 10th percentile of the norming group for the measure (Stolt and Vehkavuori, 2018).

The FinCDI-LF Words and Sentences form was used to gather information on early lexical ability, usage of words, emerging morphosyntactic skills, and sentence length at 2;0 (Lyytinen, 1999). The vocabulary score includes 595 items from different semantic categories. Different lexical categories are represented, e.g., social terms, nouns, and verbs. The usage of words is measured with five questions (max 5 points; 0 = does not use, 1 = uses). Morphosyntactic skills are measured based on the use of 16 different inflectional forms (max 32 points: 0 = not yet, 1 = sometimes, 2 = often). The ability to use word combinations is measured with one question (scored in the present study as max 1 point: 0 = does not use, 1 = uses sometimes or often). The mean length of the three longest utterances (M3L) calculated in morphemes is counted to get information on the utterance length of the children. The cut-off values at 2;0 were as follows: ≤30 words for expressive lexical skills, 2.06 for M3L, and 1.20 for inflectional forms (Lyytinen, 1999; Eklund, personal communication, 2017).[1] These are the lowest 10th percentiles of the norming group for the measure; scores above these are defined as typical development skills.

The RDLS is a formal test (Edwards et al., 1997; Kortesmaa et al., 2001). In the present study, it was used to measure receptive language ability at 2;0 and receptive, expressive, and general language ability at 3;6. The adapted Finnish version has normative data from 1;10 to 7;0. The receptive scale measures the comprehension of lexical items (nouns, verbs, prepositions) and simple and complex sentences. The expressive scale measures the ability to use different vocabulary items and long sentences. The total score is 124 points: 62 points for receptive skills and 62 points for expressive skills. Raw points are converted into standard scores; mean 100 points, +/− 1 SD 15 points. Weak general language ability was defined as ≤85 standard scores, and typical language ability as >85 (Kortesmaa et al., 2001).

## 2.3. Data analysis

Pearson's correlation coefficient values were used to test the associations between the results of early parental report instruments at 2;0 and general language ability at 3;6. Two sum scores were created to analyze the associations of the combined values of the measures: (1) ITC total score + FinCDI-SF vocabulary score and (2) vocabulary + M3L of the FinCDI-LF. The sum scores were created based on the *z*-values of each measure. The combined value for the ITC total score and the FinCDI SF was used to test whether it is possible to derive more comprehensive information on a child's early language with the help of two different types of brief parental report instruments together than when using either instrument alone. The combined value for the vocabulary score and the M3L value of the FinCDI-LF was used to test what kind of information these two values provide together compared to when used alone. The standard scores for the RDLS were used in all statistical analyses.

The Mann–Whitney *U* test was used to compare the differences between two groups; children with weak language skills at 2;0 and children with typical language skills at 2;0. Children with weak skills were defined based on the normative values for each measure (≤10%), ≤50 first words, or word combinations not used.

Six linear regression models were created to investigate how much of the variance in general language ability at 3;6 can be explained by the results of the parental report instruments when used at 2;0. In addition, the role of early receptive language was also investigated when measured using the RDLS at 2;0 in these models. The total score of the RDLS measured at 3;6 was used as an outcome variable in all models. The first model used the ITC and FinCDI-SF total scores as predictors. In the second model, the receptive language ability measured at 2;0 was added as a predictor. The FinCDI-LF vocabulary score was used as a predictor in the third model, and the receptive

1 Eklund, K. (2017). Email to Suvi Stolt, 6 February (personal communication).

language score was included in the fourth model. In the fifth model, the M3L value of the FinCDI-LF was used as a predictor, and in the sixth model, receptive language ability was included. The following background factors were included in all models: gender and maternal education (four groups, Table 1). All analyses were carried out using IBM SPSS Statistics for Macintosh, version 28.0.

## 3. Results

### 3.1. Data description

The descriptive statistics for the ITC, FinCDI-SF, FinCDI-LF, and RDLS are presented in Table 2. The mean value of the ITC was 52.5 (SD 3.2), and most children had typical skills when measured using ITC. Eleven children (16%, 8 boys) had a total score at or below the cut-off value. Fifteen children (22%, 8 boys) had weak social communication skills, seven children (10%, 5 boys) had weak speech skills, and four children (6%, 4 boys) had weak symbolic skills. When the FinCDI-SF values were considered, a considerable variation in expressive lexical development was detected. Children used roughly half of the items included in the measure. Six children (9%, 4 boys) had weak lexical skills.

Also, based on the FinCDI-LF, the children's expressive lexical skills varied considerably (Table 2). The mean value of lexicon size for the FinCDI-LF was 271 words (SD 148.3). Four children (6%, 4 boys) had weak lexical skills when the 10th percentile value of the norming sample for the FinCDI-LF was used as a cut-off value. In addition, 7 children (10%, 4 boys) had <50 words in their lexicons at 2;0. Children used roughly nine morphological inflections. Three children (4%, 2 boys) had weak skills in using inflectional forms. The mean length of the three longest utterances was 7, and seven children (10%, 5 boys) used very short utterances based on the M3L. Most children (91%, 25 boys) had started to use word combinations. Six children (9%, 5 boys) did not use word combinations at 2;0.

In the RDLS, most children (>85 standard scores, $n = 60$, 88%, 25 boys) had typical receptive language ability at 2;0 (Table 2). However, there was considerable variation in receptive language ability at 2;0. Eight children (12%, 5 boys) had weak receptive language skills.

The descriptive statistics for language skills at 3;6 are presented in Table 3. The mean and median values were within normal variation. There was significant variation in the language abilities of the participants at 3;6. There were five children (7%, 5 boys) with weak language skills. Of these five children, three had weak receptive skills, three children had weak expressive skills, and three children had weak general language ability.

### 3.2. Association between early communication and language skills and language ability at 3;6

Most correlations between the results of early parental report instruments and later language ability were statistically significant (Table 4). The total score of the ITC, when used at 2;0, correlated positively and significantly with receptive, expressive, and general language ability at 3;6 ($r$-values varied between 0.26 and 0.34, and $p$-values varied between 0.01 and 0.04). From the separate composites of the ITC, only the social communication composite was significantly associated with later receptive language. However, speech and symbolic composites were associated with later expressive language ability (Table 4). The strongest correlations were found between the speech composite and later expressive language ability. Regarding the FinCDI-SF, expressive vocabulary at 2;0 correlated clearly and relatively evenly with later receptive and expressive language ability at 3;6 ($r$-values varied between 0.32 and 0.38, and $p$-values varied between 0.00 and 0.01).

When the variables of FinCDI-LF were considered, expressive vocabulary, word use, inflectional forms, and the mean length of utterances were associated significantly and positively with later general language ability (Table 4). Statistically significant positive correlations were found between the FinCDI-LF variables and later receptive, and particularly expressive, language ability ($r$-values varied between 0.24 and 0.46, and $p$-values varied between 0.00 and 0.05). The strongest association with general language ability at 3;6 was found between expressive vocabulary and M3L when measured using the FinCDI-LF ($r$-values varied between 0.29 and 0.47, and $p$-values varied between 0.00 and 0.02).

TABLE 2 Descriptive statistics for test results measuring language skills at 2;0 ($n$=68).

| Method | | Mean (SD) | Median | Min.–Max. |
|---|---|---|---|---|
| ITC | | | | |
| | Social communication | 22.7 (2.6) | 23.0 | 17–26 |
| | Speech | 13.2 (1.2) | 14.0 | 8–14 |
| | Symbolic | 16.6 (0.6) | 17.0 | 15–17 |
| | Total score | 52.5 (3.2) | 53.0 | 44–57 |
| FinCDI-SF | | | | |
| | Expressive vocabulary | 57.1 (26.7) | 60.5 | 4–100 |
| FinCDI-LF | | | | |
| | Expressive vocabulary | 271.3 (148.3) | 294.0 | 10–528 |
| | Word use | 8.4 (1.8) | 9.0 | 1–10 |
| | Inflectional forms | 16.3 (9.7) | 9.7 | 0–32 |
| | M3L* | 6.6 (3.3) | 7.0 | 1–17 |
| RDLS | | | | |
| | Receptive | 107.7 (16.2) | 107.0 | 73–142 |

ITC, Infant-Toddler Checklist of the Communication and Symbolic Behavior Scales Developmental Profile; FinCDI-SF, Finnish Short Form versions of the Communicative Development Inventories; FinCDI-LF, Finnish Long Form versions of the Communicative Development Inventories; RDLS, Reynell Developmental Language Scales III (standard scores). *One missing value.

TABLE 3 Descriptive statistics for test results (standard scores) measuring receptive, expressive, and general language skills at 3;6 using the RDLS ($n$=68).

| Method | Mean (SD) | Median | Min.–Max. |
|---|---|---|---|
| RDLS | | | |
| Receptive | 107.4 (12.6) | 109.0 | 50–134 |
| Expressive | 103.5 (12.8) | 104.0 | 75–131 |
| Total score | 105.4 (13.0) | 107.0 | 55–130 |

RDLS, Reynell Developmental Language Scales III (standard scores).

TABLE 4 Correlations (Pearson's *r*-values and *p*-values) between early communication and language skills at 2;0 and language skills at 3;6 (*n*=68).

| Method | RDLS standard scores at 3;6 | | | | | |
|---|---|---|---|---|---|---|
| | Receptive | | Expressive | | Total score | |
| | *r* | *p* | *r* | *p* | *r* | *p* |
| **ITC** | | | | | | |
| Social communication | **0.28** | **0.02** | 0.09 | 0.49 | 0.23 | 0.06 |
| Speech | 0.20 | 0.10 | **0.36** | **0.00** | **0.31** | **0.01** |
| Symbolic | 0.10 | 0.41 | **0.27** | **0.03** | 0.20 | 0.10 |
| Total score | **0.32** | **0.01** | **0.26** | **0.04** | **0.34** | **0.01** |
| **FinCDI-SF** | | | | | | |
| Expressive vocabulary | **0.32** | **0.01** | **0.34** | **0.00** | **0.38** | **0.00** |
| Sum score of ITC+FinCDI-SF | **0.38** | **0.00** | **0.35** | **0.00** | **0.42** | **0.00** |
| **FinCDI-LF** | | | | | | |
| Expressive vocabulary | **0.38** | **0.00** | **0.44** | **0.00** | **0.46** | **0.00** |
| Word use | 0.15 | 0.22 | **0.27** | **0.03** | 0.22 | 0.07 |
| Inflections | **0.24** | **0.05** | **0.29** | **0.02** | **0.29** | **0.02** |
| M3L | **0.32** | **0.01** | **0.46** | **0.00** | **0.42** | **0.00** |
| Sum score of LF vocabulary + LF M3L | **0.37** | **0.00** | **0.49** | **0.00** | **0.47** | **0.00** |
| **RDLS** | | | | | | |
| Receptive | **0.52** | **0.00** | **0.62** | **0.00** | **0.68** | **0.00** |

Significant correlations ($p < 0.05$) are marked in bold. ITC, Infant-Toddler Checklist of the Communication and Symbolic Behavior Scales Developmental Profile; FinCDI-SF, Finnish Short Form versions of the Communicative Development Inventories; FinCDI-LF, Finnish Long Form versions of the Communicative Development Inventories; RDLS, Reynell Developmental Language Scales III.

TABLE 5 Receptive, expressive, and general language ability at 3;6 in children with weak vs. typical skills measured using the ITC, FinCDI-SF, and FinCDI-LF at 2;0 (Mann–Whitney U test).

| | RDLS standard scores at 3;6 | | | |
|---|---|---|---|---|
| | Mean (SD) | Mean (SD) | *U* | *p* |
| | Weak skills group at 2;0 | Typical skills group at 2;0 | | |
| | ITC | ITC | | |
| Receptive | 98.9 (18.8) | 109.0 (10.5) | 197.50 | 0.053 |
| Expressive | 95.3 (13.9) | 105.1 (12.1) | 189.50 | **0.039** |
| Total score | 94.6 (17.4) | 107.5 (11.0) | 161.00 | **0.011** |
| | FinCDI-SF | FinCDI-SF | | |
| Receptive | 98.5(9.2) | 108.2 (12.6) | 81.00 | **0.021** |
| Expressive | 89.8 (13.1) | 104.9 (12.1) | 73.00 | **0.012** |
| Total score | 92.8 (9.7) | 106.6 (12.7) | 57.00 | **0.003** |
| | FinCDI-LF (vocabulary) | FinCDI-LF (vocabulary) | | |
| Receptive | 99.8 (11.6) | 107.9 (12.6) | 95.00 | 0.178 |
| Expressive | 84.3 (9.8) | 104.8 (12.1) | 22.50 | **0.002** |
| Total score | 89.8 (9.8) | 106.4 (12.6) | 29.00 | **0.006** |
| | ≤50 first words | >50 first words | | |
| Receptive | 100.3 (9.7) | 108.2 (12.7) | 116.50 | 0.050 |
| Expressive | 91.6 (12.8) | 104.9 (12.2) | 99.00 | **0.021** |
| Total score | 94.9 (10.4) | 106.7 (12.8) | 83.50 | **0.009** |
| | No word combinations | Word combinations | | |
| Receptive | 99.8 (9.4) | 108.1 (12.7) | 95.00 | **0.048** |
| Expressive | 92.7 (15.1) | 104.6 (12.2) | 105.00 | 0.079 |
| Total score | 95.2 (11.3) | 106.4 (12.8) | 82.00 | **0.024** |
| | FinCDI-LF (M3L) | FinCDI-LF (M3L) | | |
| Receptive | 97.4 (8.4) | 108.6 (12.7) | 71.50 | **0.004** |
| Expressive | 91.4 (13.8) | 105.2 (12.0) | 91.00 | **0.015** |
| Total score | 93.3 (10.3) | 107.1 (12.6) | 66.50 | **0.003** |

Weak skills were defined as the ≤10th percentile of the measure in question. Standard scores of the RDLS at 3;6 are presented. Group comparisons are also presented. Significant correlations are marked in bold. ITC, Infant-Toddler Checklist of the Communication and Symbolic Behavior Scales Developmental Profile; FinCDI-SF, Finnish Short Form versions of the Communicative Development Inventories; FinCDI-LF, Finnish Long Form versions of the Communicative Development Inventories; RDLS, Reynell Developmental Language Scales III.

The combined value of the ITC and FinCDI-SF correlated more strongly with later language scores than the result of each of the measures separately (Table 4). The strongest correlation coefficient values were found between the receptive language ability, measured using the RDLS at 2;0, and receptive, expressive, and general language ability, measured using the RDLS at 3;6 (*r*-values varied between 0.52 and 0.68, and *p*-values were 0.00).

## 3.3. Comparison of language ability at 3;6 in children with weak vs. typical skills at 2;0

Comparisons between the weak- vs. typical-skills groups, when defined using different parental report instruments at 2;0 and regarding children's language ability at 3;6, are presented in Table 5. Based on the comparisons, many significant differences were found. Most of the investigated variables showed significant differences between the two groups (weak vs. typical skills at 2;0) in expressive and general language ability when measured using the RDLS at 3;6. In other words, those children with weak skills at 2;0 still had weaker skills than the rest of the group at 3;6.

## 3.4. Explanatory value of early parental report instruments and the role of receptive language ability

The regression models used to investigate the explanatory value of the results of parental report instruments are presented in Table 6. All models were statistically significant ($p < 0.00$). Regarding the

TABLE 6 The explanatory value of early parental report instruments and receptive language ability measured at 2;0 regarding general language ability at 3;6 – the regression models.

| Model information | Predictors at 2;0 | Beta | *t* | Sig. |
|---|---|---|---|---|
| **Model 1** | | | | |
| Outcome variable: RDLS total | ITC | 0.20 | 1.56 | 0.12 |
| $F(4,63) = 4.57, p < 0.00$ | FinCDI-SF | 0.25 | 1.96 | **0.05** |
| $R^2_{adj.} = 0.18$ | Gender | −0.18 | −1.56 | 0.12 |
| | Maternal education | 0.10 | 0.89 | 0.38 |
| **Model 2** | | | | |
| Outcome variable: RDLS total | ITC | 0.20 | 2.06 | **0.04** |
| $F(5,62) = 13.49, p < 0.00$ | FinCDI-SF | −0.04 | −0.38 | 0.70 |
| $R^2_{adj.} = 0.48$ | RDLS receptive | 0.64 | 6.19 | **0.00** |
| | Gender | −0.12 | −1.33 | 0.19 |
| | Maternal education | −0.02 | −0.22 | 0.82 |
| **Model 3** | | | | |
| Outcome variable: RDLS total | FinCDI-LF (vocabulary) | 0.41 | 3.64 | **0.00** |
| $F(3,64) = 7.11, p < 0.00$ | Gender | −0.18 | −1.63 | 0.11 |
| $R^2_{adj.} = 0.22$ | Maternal education | 0.07 | 0.07 | 0.54 |
| **Model 4** | | | | |
| Outcome variable: RDLS total | FinCDI-LF (vocabulary) | 0.12 | 1.12 | 0.27 |
| $F(4,63) = 15.56, p < 0.00$ | RDLS receptive | 0.60 | 5.56 | **0.00** |
| $R^2_{adj.} = 0.47$ | Gender | −0.14 | −1.50 | 0.14 |
| | Maternal education | −0.04 | −0.39 | 0.70 |
| **Model 5** | | | | |
| Outcome variable: RDLS total | FinCDI-LF (M3L) | 0.38 | 3.21 | **0.00** |
| $F(3,63) = 5.78, p < 0.00$ | Gender | −0.21 | −1.82 | 0.07 |
| $R^2_{adj.} = 0.18$ | Maternal education | −0.01 | −0.05 | 0.96 |
| **Model 6** | | | | |
| Outcome variable: RDLS total | FinCDI-LF (M3L) | 0.10 | 0.95 | 0.35 |
| $F(4,62) = 15.11, p < 0.00$ | RDLS receptive | 0.62 | 5.83 | **0.00** |
| $R^2_{adj.} = 0.46$ | Gender | −0.14 | −1.50 | 0.14 |
| | Maternal education | −0.07 | −0.73 | 0.47 |

ITC, Infant-Toddler Checklist of the Communication and Symbolic Behavior Scales Developmental Profile; FinCDI-SF, Finnish Short Form versions of the Communicative Development Inventories; FinCDI-LF, Finnish Long Form versions of the Communicative Development Inventories; RDLS, Reynell Developmental Language Scales III.

models that did not include receptive language ability at 2;0 as a predictor, the explanatory values varied between 18 and 22%. When receptive language ability was added to the models as a predictor, the explanatory values of the models increased considerably (46–48%). Background factors were not statistically significant in any of the models. The best model for explaining the general language ability at 3;6 included three variables: the ITC, the FinCDI-SF, and the receptive score of the RDLS at 2;0. This model explained 48% of the variation in general language ability at 3;6.

## 4. Discussion

The present study investigated and compared the associations of three parental report instruments when used at 2;0 and language skills at 3;6. In addition, it was also investigated if receptive/expressive/general language ability at 3;6 of those children with weak skills at 2;0 differed from the language ability of those children with typical language skills at 2;0. Further, the role of receptive language ability when assessing the possible predictive value of early language ability using the three parental report instruments was also analyzed. Most of the correlations between the results of early measures and later language ability were positive and statistically significant. The correlation coefficient value between the combined value of the ITC and FinCDI-SF and a later language score was higher than the one for individual measures. Early receptive language skills correlated clearly and significantly with later receptive, expressive, and general language abilities. In general, the participants with weak language skills at 2;0, such as vocabulary, word combinations, or mean length of the three longest utterances, had weaker language skills at 3;6 compared with the participants with typical early language development. All regression models, which were modified to investigate the possible explanatory value of parental report instruments when used at 2;0, significantly explained the variability in language skills at 3;6. However, the best models included the receptive language score at 2;0 as an explaining factor.

The parental report instruments are widely used, and the findings of this study support their being valuable tools for assessing early language skills and predicting later language ability. The results of all three parental report instruments used in this study were significantly associated with later language ability. These findings align with various studies and strengthen previous results (Feldman et al., 2000; Wetherby et al., 2002; Pan et al., 2004; Can et al., 2013; Wallace et al., 2015; Vehkavuori and Stolt, 2019; Vehkavuori, 2021). For example, the ITC is an acceptably sensitive and specific screening instrument for parents to complete and for early identification of language delays or disorders (Wetherby et al., 2003; Wallace et al., 2015). Further, in this study, the different composites of ITC provided slightly different information on later language ability. This finding supports the results of previous studies, which showed that the social communication composite correlated only with receptive language skills at 3;6, and speech and symbolic composites were associated significantly with expressive language skills at 3;6 (Laakso et al., 2011; Määttä et al., 2016). In addition, in this study, the FinCDI-SF correlated equally well with receptive and expressive language skills at 3;6. This finding confirms that the FinCDI-SF can be used to predict later receptive and expressive skills. It is in line with a previous study which has shown associations between early lexicon and general language ability at 5;0 (Vehkavuori et al., 2021).

Regarding the FinCDI-LF results, the strong correlations between expressive vocabulary and mean length of the three longest utterances with later language skills at 3;6 indicate that these measures are reliable indicators of language development. This finding is in line with previous research that has shown the importance of early expressive lexicon in predicting later lexical skills (Pan et al., 2004; Can et al., 2013; Pérez-Pereira and Cruz, 2018). Moreover, the present study found that the strongest correlation between early language measures and later language ability was observed when the combined value of expressive vocabulary size and M3L was used. This finding suggests that early language assessments should consider both expressive vocabulary development and morphosyntactic skills. In other words, it is important to evaluate a child's ability to use morphosyntax when assessing their language ability at 2;0.

To our knowledge, longitudinal information on the combined value of the ITC and the short form version of the CDI, the FinCDI-SF in this case, has not been previously presented (see however Vehkavuori, 2021). Thus, our finding that the combined value of the ITC and the short form version of the CDI provides more comprehensive information on the language ability in young children than when these instruments are used separately, is novel. This combination was used to derive as comprehensive information on a child's early language skills as possible with the help of two different types of brief parental report instruments. Our finding suggested that more representative information on early language development could be derived when two short instruments were used together than if used separately. This finding may be explained by the fact that these two instruments have been modified differently and they assess different domains of language, communication, and symbolic skills. Therefore, more comprehensive information on early language development can be derived when two different instruments are used together than if used separately.

Significant differences in language ability at 3;6 were found based on the comparison between children with weak vs. typical skills defined using the parental report instruments at 2;0. This result is parallel with studies examining late talkers or children who exhibit delayed expressive language skills, which have shown that weak early language skills are a risk factor for later weak language skills (Rescorla, 2011). However, most late-talking children catch up with their peers by age three, and a subset of them continue to struggle with language development. More longitudinal studies are needed to evidence outcomes of weak early expressive skills and to identify precursors of persistent language difficulties (Rescorla, 2011; Rescorla and Dale, 2013). The present study's finding about differences between the two groups is consistent with a previous study which showed that children who produced word combinations at 2;0 had better expressive language skills at age 8 (Poll and Miller, 2013). However, when interpreting the findings of the present study, it is important to take into consideration that the mean values of the weak group defined at 2;0 were within the typical variation at 3;6. This could be due to the fact that the present sample included only typically developing and generally healthy children without any known specific diagnoses which could have impacted the language ability of the weak group. Suppose the sample included children with language difficulties or more children with weak skills; in that case, one may assume that even more evident differences could have been detected between the groups at 3;6. The uneven number of participants in weak vs. typically developing groups may also have influenced the finding. Still, despite these factors, our result supports the view that the parental report instruments which were used in the present study could identify those with weaker language skills from those with better skills.

All regression models that included results from parental report instruments as predictors explained later language ability statistically significantly, which is consistent with previous studies (Law and Roy, 2008; Wallace et al., 2015). However, when receptive language ability measured using the RDLS at 2;0 was included in the model as a predictor, the explanatory value of the model increased considerably. This finding emphasizes the importance of early receptive language ability when assessing the language skills of 2 year-old children. Previous research has demonstrated that early receptive language is a significant predictor of later expressive language skills (Fisher, 2017), while early expressive vocabulary alone may not be the most reliable indicator of persistent language difficulties (Dollaghan, 2013). To our knowledge, the combined explanatory value of the parental report instruments, which primarily focus on expressive language ability, and receptive language measured using formal tests has not been previously investigated. Thus, our novel result provides important information to the field.

Regarding the strengths of the present study, one may conclude that this study provides novel comparison information on parental report instruments. While the difference between using two brief parental report instruments in parallel vs. separately was modest, even further information can be derived by using them together. Both receptive and expressive language ability at two age points was taken into consideration, which is another strength of this study. The ITC and the CDIs measure different domains of language, communication, and symbolic skills, which allows clinicians and researchers to get more comprehensive information on early language skills when used together. A limitation of this study is the small group of children with weak language skills, which may have affected the comparisons of the two subgroups. A larger group of children with weak language skills at 2;0 would have allowed investigation of the differences between the groups at 3;6 in a more detailed manner.

The present longitudinal study aimed to investigate early language development in children in a longitudinal setting, with a focus on comparing different parental report instruments. The study's clinical implications are twofold. First, our findings suggest that clinicians should consider using multiple parental report instruments in parallel to obtain more representative information about early language development. Second, it is important to assess early receptive language skills at 2;0 as this study indicates. However, many existing parental report instruments primarily focus on expressive language. This study underscores the value of validated parental report instruments while also highlighting the need for new instruments that capture receptive language development more accurately.

In conclusion, the present study contributes novel insights into the comparative usability of three different parental report instruments. Our results demonstrate that utilizing two brief parental report instruments in parallel yields more comprehensive information on later language skills than using either instrument alone. Additionally, significant differences in later language skills were observed between children with weak vs. typical skills at 2;0, as defined by the three different parental report instruments. Our findings suggest that parental report instruments can provide a useful indication of later language skills, at least to some extent. Furthermore, our study highlights the significance of assessing receptive language skills in 2 year-old children.

## Data availability statement

The datasets presented in this article are not readily available because of ethical and privacy issues. Requests to access the datasets should be directed to SSt, suvi.stolt@helsinki.fi.

## Ethics statement

The studies involving human participants were reviewed and approved by the Ethics Committee of the University of Turku. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

## Author contributions

SSu, S-MV, KS-H, and SSt contributed to the conception and design of the study. SSu, S-MV, and SSt organized the database. SSu performed the statistical analyses with the help of a statistician, wrote the first draft of the manuscript, and finalized the manuscript. S-MV, KS-H, and SSt critically reviewed the manuscript for important intellectual content. All authors contributed to the article and approved the submitted version.

## Funding

This study is supported by the Finnish Concordia Fund. The Sanaseula Study has been supported by the Päivikki and Sakari Sohlberg Foundation and the Jenny and Antti Wihuri Foundation. The publication fee is supported by the University of Helsinki.

## Acknowledgments

The authors would like to thank all the families who participated in the study.

## Conflict of interest

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

## Editor's note

Maria-José Ezeizabarrena edited the article in collaboration with Melita Kovacevic, University of Zagreb, Zagreb, Croatia.

## Publisher's note



## References

Bleses, D., Vach, W., Slott, M., Wehberg, S., Thomsen, P., Madsen, T. O., et al. (2008). Early vocabulary development in Danish and other languages: a CDI-based comparison. *J. Child Lang.* 35, 619–650. doi: 10.1017/S0305000908008714

Borkhoff, C. M., Atalla, M., Bayoumi, I., Birken, C. S., Maguire, J. L., and Parkin, P. C. (2022). Predictive validity of the infant toddler checklist in primary care at the 18-month visit and developmental diagnosis at 3-5 years: a prospective cohort study. *BMJ Paediatr. Open* 6:e001524. doi: 10.1136/bmjpo-2022-001524

Cadime, I., Moreira, C. S., Santos, A. L., Silva, C., Ribeiro, I., and Viana, F. L. (2019). The development of vocabulary and grammar: a longitudinal study of European Portuguese-speaking toddlers. *J. Child Lang.* 46, 653–681. doi: 10.1017/S0305000919000060

Can, D. D., Ginsburg-Block, M., Golinkoff, R. M., and Hirsh-Pasek, K. (2013). 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.* 40, 821–835. doi: 10.1017/S030500091200030X

Conboy, B. T., and Thal, D. J. (2006). Ties between the lexicon and grammar: cross-sectional and longitudinal studies of bilingual toddlers. *Child Dev.* 77, 712–735. doi: 10.1111/j.1467-8624.2006.00899.x

Crais, E. R. (2011). Testing and beyond: strategies and tools for evaluating and assessing infants and toddlers. *Lang. Speech Hear. Serv. Sch.* 42, 341–364. doi: 10.1044/0161-1461(2010/09-0061)

Dale, P. S., Price, T. S., Bishop, D. V. M., and Plomin, R. (2003). Outcomes of early language delay: I. predicting persistent and transient language difficulties at 3 and 4 years. *J. Speech Lang. Hear. Res.* 46, 544–560. doi: 10.1044/1092-4388(2003/044)

Desmarais, C., Sylvestre, A., Meyer, F., Bairati, I., and Rouleau, N. (2008). Systematic review of the literature on characteristics of late-talking toddlers. *Int. J. Lang. Commun. Disord.* 43, 361–389. doi: 10.1080/13682820701546854

Dollaghan, C. (2013). Late talker as a clinical category: a critical evaluation, *Late talkers: language development, interventions, and outcomes*, (Eds.) L. A. Rescorla and P. S. Dale Baltimore, MD: Paul H. Brookes Publishing Co., 91–112.

Eadie, P. A., Ukoumunne, O., Skeat, J., Prior, M. R., Bavin, E., Bretherton, L., et al. (2010). Assessing early communication behaviours: structure and validity of the communication and symbolic behaviour scales-developmental profile (CSBS-DP) in 12-month-old infants. *Int. J. Lang. Commun. Disord.* 45, 572–585. doi: 10.3109/13682820903277944

Edwards, S., Fletcher, P., Garman, M., Hughes, A., and Letts, C. (1997). *Reynell developmental language scales III (RDLS III)*. Berkshire: NFER-Nelson.

Eriksson, M. (2017). The Swedish communicative development inventory III: parent reports on language in preschool children. *Int. J. Behav. Dev.* 41, 647–654. doi: 10.1177/0165025416644078

Eriksson, M., Marschik, P. B., Tulviste, T., Almgren, M., Pérez Pereira, M., Wehberg, S., et al. (2012). Differences between girls and boys in emerging language skills: evidence from 10 language communities. *Br. J. Dev. Psychol.* 30, 326–343. doi: 10.1111/j.2044-835X.2011.02042.x

Eriksson, M., and Myrberg, K. (2023). How the communicative development inventories can contribute to clinical assessments of children with speech and language disorders. *Front. Psychol.* 14. doi: 10.3389/fpsyg.2023.1176028 (Accessed Jun 14, 2023)

Ezeizabarrena, M. J., and Fernández, I. G. (2022). Tres indicadores de exposición a la lengua y su efecto en la adquisición del léxico de una lengua minoritaria. *Rev. Logop. Fon. Audiol.* 43:100289. doi: 10.1016/j.rlfa.2022.04.003

Fäldt, A., Fabian, H., Dahlberg, A., Thunberg, G., Durbeej, N., and Lucas, S. (2021). Infant-toddler checklist identifies 18-month-old children with communication difficulties in the Swedish child healthcare setting. *Acta Paediatr.* 110, 1505–1512. doi: 10.1111/apa.15696

Farabolini, G., Ceravolo, M. G., and Marini, A. (2023). Towards a characterization of late talkers: the developmental profile of children with late language emergence through a web-based communicative-language assessment. *Int. J. Environ. Res. Public Health* 20:1563. doi: 10.3390/ijerph20021563

Feldman, H. M., Dale, P. S., Campbell, T. F., Colborn, D. K., Kurs-Lasky, M., Rockette, H., et al. (2005). Concurrent and predictive validity of parent reports of child language at ages 2 and 3 years. *Child Dev.* 76, 856–868. doi: 10.1111/j.1467-8624.2005.00882.x

Feldman, H. M., Dollaghan, C. A., Campbell, T. F., Kurs-Lasky, M., Janosky, J. E., and Paradise, J. L. (2000). Measurement properties of the MacArthur communicative development inventories at ages one and two years. *Child Dev.* 71, 310–322. doi: 10.1111/1467-8624.00146

Fenson, L., Bates, E., Dale, P., Goodman, J., Reznick, J., and Thal, D. J. (2000a). Reply: measuring variability in early child language: don't shoot the messenger. *Child Dev.* 71, 323–328. doi: 10.1111/1467-8624.00147

Fenson, L., Dale, P. S., Reznick, J. S., Thal, D., Bates, E., Hartung, J. P., et al, (1993). *MacArthur communicative development inventories: user's guide and technical manual.* San Diego: Singular Pub Group.

Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, J. S., and Bates, E. (2007). *MacArthur-Bates communicative development inventories: user's guide and technical manual* (*2nd*). Baltimore: Paul H. Brookes Publishing Co.

Fenson, L., Pethick, S., Renda, C., Cox, J., Dale, P., and Reznick, J. S. (2000b). Short-form versions of the MacArthur communicative development inventories. *Appl. Psycholinguist.* 21, 95–116. doi: 10.1017/S0142716400001053

Fisher, E. L. (2017). A systematic review and meta-analysis of predictors of expressive-language outcomes among late talkers. *J. Speech Lang. Hear. Res.* 60, 2935–2948. doi: 10.1044/2017_JSLHR-L-16-0310

Frota, S., Butler, J., Correia, S., Severino, C., Vicente, S., and Vigário, M. (2016). Infant communicative development assessed with the European Portuguese MacArthur-Bates communicative development inventories short forms. *First Lang.* 36, 525–545. doi: 10.1177/0142723716648867

Girolametto, L., Wiigs, M., Smyth, R., Weitzman, E., and Pearce, P. S. (2001). Children with a history of expressive vocabulary delay. *Am. J. Speech Lang. Pathol.* 10, 358–369. doi: 10.1044/1058-0360(2001/030)

Hawa, V. V., and Spanoudis, G. (2014). Toddlers with delayed expressive language: an overview of the characteristics, risk factors and language outcomes. *Res. Dev. Disabil.* 35, 400–407. doi: 10.1016/j.ridd.2013.10.027

Heilmann, J., Ellis Weismer, S., Evans, J., and Hollar, C. (2005). Utility of the MacArthur-Bates communicative development inventory in identifying language abilities of late-talking and typically developing toddlers. *Am. J. Speech Lang. Pathol.* 14, 40–51. doi: 10.1044/1058-0360(2005/006)

Hurtado, N., Grüter, T., Marchman, V. A., and Fernald, A. (2014). Relative language exposure, processing efficiency and vocabulary in Spanish–English bilingual toddlers. *Biling. Lang. Cogn.* 17, 189–202. doi: 10.1017/s136672891300014x

Jackson-Maldonado, D., Marchman, V. A., and Fernald, L. C. H. (2013). Short-form versions of the Spanish MacArthur-Bates communicative development inventories. *Appl. Psycholinguist.* 34, 837–868. doi: 10.1017/S0142716412000045

Jago, L., Alcock, K., Meints, K., Pine, J. M., and Rowland, C. F. (2023). Language outcomes from the UK-CDI project: can risk factors, vocabulary skills and gesture scores in infancy predict later concern for language development? *Front. Psychol.* 14. doi: 10.3389/fpsyg.2023.1167810 (Accepted May 19, 2023)

Joensuu, E., Munck, P., Setänen, S., Lipsanen, J., Huhtala, M., Lapinleimu, H., et al. (2021). 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. *Child. Basel* 8:510. doi: 10.3390/children8060510

Korpilahti, P., Kaljonen, A., and Jansson-Verkasalo, E. (2016). Population-based screening for language delay: let's talk STEPS study. *Psychology* 7, 205–214. doi: 10.4236/psych.2016.72023

Kortesmaa, M., Heimonen, K., Merikoski, H., Warma, M.-L., and Varpela, V. (2001). *Reynell developmental language scales III, Finnish scales*. Helsinki: Psykologien Kustannus Oy.

Laakso, M.-L., Eklund, K., and Poikkeus, A.-M. (2011). *ESIKKO – Lapsen esikielellisen kommunikaation ja kielen ensikartoitus*. Jyväskylä: Niilo Mäki Instituutti.

Lasorsa, F. R., Caselli, M. C., Simonelli, I., Stefanini, S., Pasqualetti, P., and Rinaldi, P. (2021). Exploring agreement between MB-CDI short forms for evaluating the language skills of Italian children aged 18–24 months. *Folia Phoniatr. Logop.* 73, 552–564. doi: 10.1159/000513243

Law, J., and Roy, P. (2008). Parental report of infant language skills: a review of the development and application of the communicative development inventories. *Child Adolesc. Ment. Health* 13, 198–206. doi: 10.1111/j.1475-3588.2008.00503.x

Lyytinen, P. (1999). *Varhaisen kommunikaation ja kielen kehityksen arviointimenetelmä*. Jyväskylä: Jyväskylän yliopiston Lapsitutkimuskeskus and Niilo Mäki Instituutti.

Lyytinen, P., Eklund, K., and Lyytinen, H. (2005). Language development and literacy skills in late-talking toddlers with and without familial risk for dyslexia. *Ann. Dyslexia* 55, 166–192. doi: 10.1007/s11881-005-0010-y

Määttä, S. (2017). Developmental pathways of language development: a longitudinal predictive study from prelinguistic stage to outcome at school entry. *Dissertation*. Jyväskylä: University of Jyväskylä

Määttä, S., Laakso, M.-L., Tolvanen, T. A. A., Westerholm, J., and Aro, T. (2016). Continuity from prelinguistic communication to later language ability: a follow-up study from infancy to early school age. *J. Speech Lang. Hear. Res.* 59, 1357–1372. doi: 10.1044/2016_JSLHR-L-15-0209

Marchman, V. A., Dale, P. S., and Fenson, L. (2023). *MacArthur-Bates communicative development inventories: user's guide and technical manual* (*3rd*). Baltimore: Paul H. Brookes Publishing Co.

Marchman, V. A., and Martine-Sussmann, C. (2002). Concurrent validity of caregiver/parent report measures of language for children who are learning both English and Spanish. *J. Speech Lang. Hear. Res.* 45, 983–997. doi: 10.1044/1092-4388(2002/080)

Marjanovič-Umek, L., Fekonja-Peklaj, U., and Sočan, G. (2017). Early vocabulary, parental education, and the frequency of shared reading as predictors of toddler's vocabulary and grammar at age 2;7: a Slovenian longitudinal CDI study. *J. Child Lang.* 44, 457–479. doi: 10.1017/S0305000916000167

Morgan, L., Delehanty, A., Cleary Dillon, J., Schatschneider, C., and Wetherby, A. M. (2020). Measures of early social communication and vocabulary production to predict language outcomes at 2 and 3 years in late-talking toddlers. *Early Child Res. Q.* 51, 366–378. doi: 10.1016/j.ecresq.2019.12.005

Nurse, K. M., Janus, M., Birken, C. S., Keown-Stoneman, C. D. G., Omand, J. A., Maguire, J. L., et al. (2022). Predictive validity of the infant toddler checklist in primary care at the 18-month visit and school readiness at 4 to 6 years. *Acad. Pediatr.* 23, 322–328. doi: 10.1016/j.acap.2022.09.004

Pan, B. A., Rowe, M. L., Spier, E., and Tamis-LeMonda, C. (2004). Measuring productive vocabulary of toddlers in low-income families: concurrent and predictive validity of three sources of data. *J. Child Lang.* 31, 587–608. doi: 10.1017/s0305000904006270

Patrucco-Nanchen, T., Friend, M., Poulin-Dubois, D., and Zesiger, P. (2019). Do early lexical skills predict language outcome at 3 years? A longitudinal study of French-speaking children. *Infant Behav. Dev.* 57:101379. doi: 10.1016/j.infbeh.2019.101379

Pérez-Pereira, M., and Cruz, R. (2018). A longitudinal study of vocabulary size and composition in low risk preterm children. *First Lang.* 38, 72–94. doi: 10.1177/0142723717730484

Poll, G. H., and Miller, C. A. (2013). Late talking, typical talking, and weak language skills at middle childhood. *Learn. Individ. Differ.* 26, 177–184. doi: 10.1016/j.lindif.2013.01.008

Rescorla, L. (2011). Late talkers: do good predictors of outcome exist? *Dev. Disabil. Res. Rev.* 17, 141–150. doi: 10.1002/ddrr.1108

Rescorla, L. A., and Dale, P. (2013). Where do we stand now? Conclusions and future directions, *Late talkers: language development, interventions, and outcomes*, (Eds.) L. A. Rescorla and P. S. Dale Baltimore, MD: Paul H. Brookes Publishing Co., 377–387.

Rescorla, L., Ratner, N. B., Jusczyk, P., and Jusczyk, A. M. (2005). Concurrent validity of the language development survey: associations with the MacArthur-Bates communicative development inventories: words and sentences. *Am. J. Speech Lang. Pathol.* 14, 156–163. doi: 10.1044/1058-0360(2005/016)

Sansavini, A., Favilla, M. E., Guasti, M. T., Marini, A., Millepiedi, S., Di Martino, M. V., et al. (2021). Developmental language disorder: early predictors, age for the diagnosis, and diagnostic tools. A scoping review. *Brain Sci.* 11:654. doi: 10.3390/brainsci11050654

Simonsen, H. G., Kristoffersen, K. E., Bleses, D., Wehberg, S., and Jørgensen, R. N. (2014). The Norwegian communicative development inventories: reliability, main developmental trends and gender differences. *First Lang.* 34, 3–23. doi: 10.1177/0142723713510997

Stolt, S. (2023). Internal consistency and concurrent validity of the parental report instrument on language in pre-school-aged children – the Finnish communicative development inventory III. *First Lang.* 2023:014272372311673. doi: 10.1177/01427237231167301

Stolt, S., Haataja, L., Lapinleimu, H., and Lehtonen, L. (2008). Early lexical development of Finnish children: a longitudinal study. *First Lang.* 28, 259–279. doi: 10.1177/0142723708091051

Stolt, S., Haataja, L., Lapinleimu, H., and Lehtonen, L. (2009). Associations between lexicon and grammar at the end of the second year in Finnish children. *J. Child Lang.* 36, 779–806. doi: 10.1017/S0305000908009161

Stolt, S., Matomäki, J., Lind, A., Lapinleimu, H., Haataja, L., and Lehtonen, L. (2014). The prevalence and predictive value of weak language skills in children with very low birth weight – a longitudinal study. *Acta Paediatr.* 103, 651–658. doi: 10.1111/apa.12607

Stolt, S., and Vehkavuori, S. (2018). *Sanaseula: MacArthur-Bates Communicative Development Inventories -arviointimenetelmän lyhyt, suomalainen versio*. Jyväskylä: Niilo Mäki Instituutti.

The MacArthur-Bates Communicative Development Inventories (2023). Adaptations in other languages. Available at: https://mb-cdi.stanford.edu/adaptations.html (Accessed April 15, 2023).

Torppa, M., Lyytinen, P., Erskine, J., Eklund, K., and Lyytinen, H. (2010). Language development, literacy skills, and predictive connections to Reading in Finnish children with and without familial risk for dyslexia. *J. Learn. Disabil.* 43, 308–321. doi: 10.1177/0022219410369096

Urm, A., and Tulviste, T. (2021). Toddlers' early communicative skills as assessed by the short form version of the Estonian MacArthur-Bates communicative development inventory II. *J. Speech Lang. Hear. Res.* 64, 1303–1315. doi: 10.1044/2020_JSLHR-20-00201

Vehkavuori, S.-M. (2021). Early lexicon: associations to later language skills and screening. *Dissertation*. Oulu: University of Oulu

Vehkavuori, S.-M., Kämäräinen, M., and Stolt, S. (2021). Early receptive and expressive lexicons and language and pre-literacy skills at 5;0 years – a longitudinal study. *Early Hum. Dev.* 156:105345. doi: 10.1016/j.earlhumdev.2021.105345

Vehkavuori, S.-M., and Stolt, S. (2018). Screening language skills at 2;0. *Infant Behav. Dev.* 50, 174–179. doi: 10.1016/j.infbeh.2018.01.001

Vehkavuori, S.-M., and Stolt, S. (2019). Early lexicon and language skills at 42 months. *Clin. Linguist. Phon.* 33, 854–868. doi: 10.1080/02699206.2019.1584721

Viana, F. L., Pérez-Pereira, M., Cadime, I., Silva, C., Santos, S., and Ribeiro, I. (2017). Lexical, morphological and syntactic development in toddlers between 16 and 30 months old: a comparison across European Portuguese and Galician. *First Lang.* 37, 285–300. doi: 10.1177/0142723717691087

Wallace, I. F., Berkman, N. D., Watson, L. R., Coyne-Beasley, T., Wood, C. T., Cullen, K., et al. (2015). Screening for speech and language delay in children 5 years old and younger: a systematic review. *Pediatrics* 136, e448–e462. doi: 10.1542/peds.2014-3889

Watt, N., Wetherby, A., and Shumway, S. (2006). Prelinguistic predictors of language outcome at 3 years of age. *J. Speech Lang. Hear. Res.* 49, 1224–1237. doi: 10.1044/1092-4388(2006/088)

Wetherby, A. M., Allen, L., Cleary, J., Kublin, K., and Goldstein, H. (2002). Validity and reliability of the communication and symbolic behavior scales developmental profile with very young children. *J. Speech Lang. Hear. Res.* 45, 1202–1218. doi: 10.1044/1092-4388(2002/097)

Wetherby, A. M., Goldstein, H., Cleary, J., Allen, L., and Kublin, K. (2003). Early identification of children with communication disorders: concurrent and predictive validity of the CSBS developmental profile. *Infants Young Child.* 16, 161–174. doi: 10.1097/00001163-200304000-00008

Wetherby, A., and Prizant, B. (2002). *Communication and symbolic behavior scales – Developmental profile*. Baltimore: Paul H. Brookes Publishing Co.

Zubrick, S. R., Taylor, C. L., Rice, M. L., and Slegers, D. W. (2007). Late language emergence at 24 months: an epidemiological study of prevalence, predictors, and covariates. *J. Speech Lang. Hear. Res.* 50, 1562–1592. doi: 10.1044/1092-4388(2007/106)

Frontiers in Psychology

TYPE Original Research
PUBLISHED 24 July 2023
DOI 10.3389/fpsyg.2023.1177161

# Communicative and linguistic factors influencing language development at 30 months of age in preterm and full-term children: a longitudinal study using the CDI

Anastasiia Ogneva and Miguel Pérez-Pereira*

Department of Developmental and Educational Psychology, University of Santiago de Compostela, Santiago de Compostela, Spain

Check for updates

OPEN ACCESS

EDITED BY
Melita Kovacevic,
University of Zagreb, Zagreb, Croatia

REVIEWED BY
Nayeli Gonzalez-Gomez,
Oxford Brookes University, United Kingdom
Bahia Guellai,
Université Paris Nanterre, France

*CORRESPONDENCE
Miguel Pérez-Pereira
miguel.perez.pereira@usc.es

RECEIVED 01 March 2023
ACCEPTED 22 June 2023
PUBLISHED 24 July 2023

CITATION
Ogneva A and Pérez-Pereira M (2023) Communicative and linguistic factors influencing language development at 30 months of age in preterm and full-term children: a longitudinal study using the CDI. *Front. Psychol.* 14:1177161.
doi: 10.3389/fpsyg.2023.1177161



**Introduction:** Previous studies showed that very preterm children have a delay in communicative (gestures) and linguistic development as compared to full-term children. Earlier use of gestures, as well as of word comprehension and production, have been found to be predictive of subsequent word production and/or language delay in both very preterm and full-term children. Not many studies on communicative antecedents of language, however, have been carried out with low-risk preterm children in comparison to full-term children.

**Methods:** In the present study a sample ($N$=142) of low-risk preterm children has been followed using the Galician version of the Communicative Development Inventories (CDI) at the ages of 10, 22, and 30 months of age and their results were compared to the results from a sample ($N$=49) of full-term children at the same ages. The determinants of language measures (vocabulary and grammar) at 30 months of age have been studied through linear regression analyses.

**Results:** ANOVA results indicate that there were no significant differences between the groups in any of the measures obtained with the CDI at any time, nor were there any differences in lexical or grammatical developmental trajectories between both groups (repeated measures ANOVA). Linear regression analyses showed that the predictors of language at 30 months of age are somewhat different for the full-term than for the preterm group.

**Discussion:** While the use of first communicative gestures at 10 months is a predictor of word production at 30 months of age for the full-term group, participation in games and routines seems to play a significant predictive role for preterm children. Word production at 22 months is the factor with a major incidence on word production at the age of 30 months for both groups. Previous specific measures of grammatical development have a clear determinant role in grammar measures at 30 months of age for the full-term children, while in the case of preterm children previous lexical development seems to be more relevant.

KEYWORDS

**low-risk preterm children, gestures, lexical development, grammatical development, determinants of language development**

# 1. Introduction

## 1.1. Early precursors of language in full-term children

Children start to use their first words around 12 months of age. They show other abilities shortly before that time, however, which are considered to be precursors of language. Among these are the abilities to imitate the actions of the caregiver, to participate in social games and daily routines, and to use gestures. From a theoretical point of view, the emergence of these abilities has been linked to important advances in socio-cognitive capacities. From Vygotsky's (1978) pioneering proposal that the use of gestures arises as a result of the interiorization process that is forged in social interactions, other authors have developed this idea, and have proposed other socio-cognitive abilities to explain the emergence of imitation of adults' actions, participation in interaction routines and social games, and use of gestures, as well as the first communicative actions (protoimperatives and protodeclaratives; Bates et al., 1975; Ratner and Bruner, 1978; Bruner, 1983; Nelson, 1985; Tomasello et al., 1993; Tomasello, 2003). Those proposals share a socio-pragmatic perspective on language acquisition, which is inspired in authors such as Wittgenstein (Nelson, 2009). According to Tomasello and others (Tomasello et al., 1993; Carpenter et al., 1998; Tomasello, 2003), the attainment of three socio-cognitive abilities is the foundation not only for the development of communicative abilities but also for the development of first language: (1) joint attention, (2) intentional reading, and (3) cultural learning or the capacity of role reversal imitation. These three abilities emerge between 9 and 12 months of age in this order and are also crucial for shared intentionality (Tomasello and Carpenter, 2007; Tomasello, 2008).

Gestures and their role for the development of language have been more widely studied than imitation of actions or participation in social games and daily interactional routines.

Young children communicate using gestures before they produce their first words (Bates, 1976). Gestures are reported to reflect cognitive and socio-cognitive developmental changes in infancy (Tomasello, 2003; Kuvač et al., 2014). There are several systems of gesture classification. According to Farkas (2007), most current studies follow the classification by Capirci et al. (1996) who described deictic and symbolic gestures which are the earliest to appear in child communication. Iconic gestures, proposed by Nicoladis et al. (1999), are produced when children have already acquired some verbal language.

At around 9–10 months, children start using deictic gestures which can be considered the first signs of intentional communication and their referential meaning is given entirely by the context (Özçalişkan et al., 2014). The most commonly studied deictic gestures are pointing, reaching, showing and giving (Crais et al., 2004). Children use them to draw parent's attention to an object, for example, pointing at a bottle to indicate a bottle. Deictic gestures constitute a useful tool for children to refer to objects before they can verbally name them. Previous research suggests that pointing at a particular object increases the chances for the child to learn the word for an object (Iverson and Goldin-Meadow, 2005) and generally paves the way for verbal language development (Goodwyn et al., 2000). A second class of gestures, referred to as "symbolic" or "representational," typically appear in children's communication at approximately 12–15 months of age. Some authors distinguish between social gestures, action-related or object-related gestures (Farkas, 2007; Stefanini et al., 2009). Gestural routines are part of everyday interactional routines (e.g., waving "bye-bye" or shaking head for "no") and are also considered to be the first communicative gestures (Fenson et al., 1993). Action-related gestures and object-related gestures are used to refer to the function of a referent or the referent itself.

The early gestures produced by children are not only considered to be precursors of words, but they are also predictors of them. Previous research has shown that it is possible to predict a large portion of the words that will eventually appear in children's spoken vocabulary. Lexical items that were initially expressed with gestures appear in the verbal lexicon 3 months later (Rowe et al., 2008). Silva et al. (2017) studied communicative development of Portuguese infants aged between 8 and 15 months and concluded that although gestures are a good predictor of vocabulary development, they are more closely associated with vocabulary comprehension than with vocabulary production. Similarly, Cadime et al. (2017), who studied 48 children at 9, 12 and 15 months of age longitudinally with the Portuguese version of the Communicative Development Inventories (CDI), found that the total number of actions and gestures and the number of early gestures produced at 9 and 12 months predicted the number of words comprehended at 15 months of age. The number of words produced, however, was predicted by actions and words only at 9 and 12 months, but not later.

Rowe and Goldin-Meadow (2009) suggested that early gestures predict later language development in a selective manner. It was found that gesture use at 18 months selectively predicted lexical and syntactic skills at 42 months. Specifically, different meanings conveyed in gestures at 18 months predicted vocabulary at 42 months, but the number of gesture and speech combinations did not predict later vocabulary. Similar results were obtained by Kuvač et al. (2014) who carried out a study with 250 infants aged 8–16 months to analyze predictive roles of different types of gestures on the onset of first word categories in early expressive vocabulary. According to their results, different types of gestures predict different types of words. For example, open-class words (such as common noun and predicates) were strongly predicted by object gestures, whereas social terms were predicted by gestural routines.

Some studies have reported an association between earlier and later verbal abilities in typically developing children. Specifically, early comprehension is claimed to be associated with later receptive (Bates et al., 1988) and expressive vocabulary (Bavin et al., 2008). The association between production of words and gestures and later expressive vocabulary skills has been reported as well (Capirci et al., 1996). Some studies also have suggested that typically developing children benefit from observing referential iconic gestures in narrative comprehension (Dargue and Sweller, 2020) and that increasing exposure to gestures produced by mothers may impact 10–12 month old infants' language development through an effect on sensorimotor brain activity (Salo et al., 2023).

Gestures have also been found to be correlated with language impairment in some studies. More specifically, children at later risk of language impairment were found to present significantly less gesture use and vocabulary abilities compared to the typically developing peers. Therefore, scarce gesture use may potentially serve as a diagnostic tool to identify children at risk for language impairment (Jackson-Maldonado, 2004; Goldin-Meadow et al., 2014; Hsu and Iyer, 2016). Similarly, Thal et al. (1991) conducted a follow-up study with 10 children who were 10% below their age peers in verbal language production when first measured and it was found that those children presented a significantly lower use of gestures.

The participation of infants in social games appears for the first time even before 8 months of age, and it becomes a very frequent activity in infants' lives. Children's participation in everyday routines and first social games, such as peek a boo, is firstly scaffolded by the adult, although later the children will be able to initiate social games by themselves. In this way they can affect the behavior of their parents and convey their wishes to play (Clark, 1978; Ratner and Bruner, 1978; Bruner, 1983; Camaioni and Laicardi, 1985). Participation in conventional social games is considered to favor language development because of the characteristics of social games: their high predictability, which will allow the children to anticipate what will happen next, and their organization in participants' turns, which may be reversible (Ratner and Bruner, 1978; Bruner, 1983). When the integrated multimodal structure of the game is violated, children are less engaged in it (Fantasia et al., 2014).

Infants' first words are mainly produced in contexts of social games, in which the mothers tend to use very repetitive and predictable language (Bruner, 1983; Camaioni and Laicardi, 1985). This finding has been corroborated by Dromi and Zaidman-Zait (2011) who found that participation in social games (peek a boo) and book reading activities (but not use of pointing gestures) was significantly associated with the number of words produced by 154 children between 12 and 15 months of age. The authors conclude that the transition into conventional language takes place within a rich context of non-verbal communicative behaviors (Dromi and Zaidman-Zait, 2011).

As for role reversal imitation of adult actions, its appearance is closely linked to the emergence of cultural learning, which is based on previous abilities for shared attention and the interpretation of the other's intention (intentional reading). Role reversal imitation is of capital importance for the learning of cultural tools (spoons, glasses, keys, computers …) and the appropriation of culture by human beings (Tomasello et al., 1993). Manifestations of cultural learning capacity appear around 10–12 months of age. After this age, children learn to use many things relating to their cultural background and how to behave in different circumstances. Progress in role reversal imitation ability occurs after 12 months of age. Carpenter et al. (2005) found that imitations of other's actions are just as common in typically developing infants at 12 months of age as at 18 months; role reversal actions which involve acting on an object (triadic object related role reversals), however, are more difficult for 12 month old children than for those of 18 months. Children with autism spectrum disorder were found to have a very limited use of role reversal imitation (Carpenter et al., 2005). The authors found positive relations between role reversal imitation and measures of language development at 18 months of age. Imitative actions, language comprehension, and language production at 18 months uniquely contributed to the prediction of late development of language production at 30 months in a sample of nearly 30,000 Norwegian children, while pointing gestures did not (Zambrana et al., 2013). Action imitation, therefore, seems to be a better predictor of late language development than pointing gesture.

## 1.2. Early precursors of language in preterm children

Although there have been several studies that have examined the development of gestural communication among atypically developing children such as children with Down syndrome (Iverson et al., 2003) or children with Williams syndrome (Laing et al., 2002), research focused on preterm children is still rare. Preterm birth has been reported to be a factor that negatively affects early communication development (during the period of 8–15 months), especially among those children who were born under 32 weeks of gestation (Pérez-Pereira et al., 2014).

Suttora and Salerni's longitudinal observational study (Suttora and Salerni, 2012) explored the development of communicative gestures in 16 preterm children [mean gestational age (GA) = 30 weeks] and two groups of full-term children at different periods (12, 18, and 24 months of age). Deictic gestures were the most frequently produced by the FT and the PT children at 12, 18, and 28 months of age, followed by referential gestures. No differences in the use of gestures or gesture types were found between the FT and the PT children. Their findings suggest that for preterm children the production of communicative pointing at 12 months is positively related to the linguistic skills at 18 and 24 months of age. The presence of pointing in children's communication at 12 months predicted their vocabulary size at 18 months and the spontaneous lexical productivity and complexity at 24 months.

Sansavini et al. (2011b) reported that 104 preterm children (mean GA = 29.5 weeks), who were measured through the Italian short form of the CDI, showed a slower rate of development in gesture/action production, word comprehension, and word production than 20 FT children, with an increasing divergence between the two groups from 12 to 24 months. Nevertheless, the preterm sample used in these studies included very or extremely preterm children (<32 and <28 weeks of gestation) or children with very low birth weight (<1,500 g). Lexical competencies at 12 months, together with gestures/actions at 18 months, were predictive of word production at 24 months.

There is controversy on the long-term effect of early development of gestures and receptive and expressive language on later language skills. The results of Pérez-Pereira et al.'s (2014) study indicate that although gestures and early word comprehension (measured at 10 months) predict very early word production, this effect disappears with time showing no correlation after 24 months of age. Similar results were obtained by Stolt et al. (2014) who found a significant effect of gestures measured at 15 months on language scores at 24 months of age, but no significant predictive value of gestures measured at 9 months on language at 24 months. However, Stolt et al. (2016) reported that the development of gestures measured between the ages of 9 months and 15 months, as well as the receptive and expressive language ability measured at 24 months, correlate significantly and positively with language skills at 60 months in preterm children with very low birthweight (GA range = 23–34 weeks of gestation).

There is controversy as well in the results of research focused on language development in preterm children. On one hand, many studies have reported that preterm (mainly very and extremely preterm) children present smaller vocabulary size as well as lower grammatical skills in comparison with full-term children (Sansavini et al., 2010, 2011a; Stolt et al., 2012, 2013; Varela-Moraga et al., 2023).

On the contrary, a few studies conducted with healthy preterm children with a wider range of gestational age have not found differences between full-term and preterm children in language acquisition (Sansavini et al., 2006; Gayraud and Kern, 2007; Pérez-Pereira et al., 2014; Pérez-Pereira and Cruz, 2018; Suttora et al., 2020).

Sansavini et al. (2006) investigated early lexical and grammatical development in Italian preterms (GA < 33 weeks) and fullterms at the age of 30 months. The result of this study suggested that most of the preterm sample displayed linguistic abilities within the normal range. As for the factors influencing preterms' language development, birthweight, gestational age and gender were shown to have the major effect. Specifically, total words number and MLU scores are affected by an extremely low birthweight, a gestational age < 31 weeks and male gender.

Gayraud and Kern (2007) studied early grammatical and lexical development in 323 preterm children compared to full-term peers at 24 months using the French MacArthur-Bates parental report. Preterm children were grouped according to their GA: extremely preterm (under 28 weeks of gestation), very preterm (between 28 and 31 weeks of gestation) and moderately preterm (between 32 and 36 weeks of gestation). Results showed that preterm children understood fewer words and produced more games, routines and onomatopoeia words. Overall, no differences were found between preterm and full-term children, if the extremely preterm group was not considered. The results obtained in this study showed that pre-term children obtained scores similar to those of younger full-term children. Therefore, the authors suggest that differences observed between groups are delays rather than deviances from the typical course of language development. These authors suggest that, as preterm children mature, differences between preterm and full-term children decline (Gayraud and Kern, 2007).

In Pérez-Pereira et al.'s (2014) study no significant differences were found between 3 groups of preterm children with different GAs (extremely and very preterm, moderately preterm and late preterm) and full-term children in communicative, lexical or grammatical development.

Preterm and full-term children were also reported to have similar developmental paths in lexical development. Specifically, Pérez-Pereira and Cruz (2018) compared the vocabulary size and composition of preterm children with different gestational age (very and extremely preterm group: 26–31 weeks, moderately preterm group: 32–33 weeks, late preterm group: 34–36 weeks) and full-term children at different periods of time (10, 22, and 30 months). Growth curve analyses showed no differences in word categories or vocabulary size among the four groups of participants. The main predictors of total vocabulary and word categories at 30 months were cognitive scores and word production measured at 22 months.

To the best of our knowledge, no study has been performed with PT children to investigate the effect of action imitation or participation in social games on later language development.

The existence of parental inventories has made it possible to gather an extensive amount of information on early communicative and linguistic development. The CDI permits the assessment of: use of gestures, participation in social games and routines, action imitation ability, as well as word comprehension and production between 8 and 15 months of age. The CDI also enables us to explore the abilities of children in word production, as well as morphosyntactic development (see the instruments section below) between 16 and 30 months of age. Therefore, the CDI seems to be an adequate, reliable and easy to use instrument to explore longitudinal relationships between early communicative and linguistic abilities and later language development.

To summarize, communicative antecedents of language have not been studied to a great extent in low-risk preterm children, children without associated medical complications. Therefore, the main aims of this study are the following:

1. To compare (cross-sectional analysis) the results obtained by the PT and the FT groups in the measures taken at 10, 22, and 30 months of age (see the instruments section and the analysis performed section below).
2. To compare the developmental trajectories throughout time of preterm and full-term children in the measures taken at different occasions (see the analysis performed section below).
3. To identify the factors predicting language development (word production, use of regular morphemes, MLU3 and sentence complexity) at 30 months of age in preterm and full-term children (see the analysis performed section).

The hypotheses of the study are as follows:

1. There will not be significant differences between the preterm and full-term groups of children in the scores obtained in the different measures of the Inventario para o Desenvolvemento de Habilidades Comunicativas: the Galician CDI (IDHC) taken at 10, 22, and 30 months of age, given the low-risk condition of the PT children.
2. No significant differences between the FT and the PT groups (inter-subjects differences) will exist in the developmental trajectories throughout time of the measures taken on different occasions: word production, MLU3, sentence complexity and regular suffixes.
3. The use of first communicative gestures will have an influence on some of the linguistic measures taken at 30 months of age in the FT as well as in the PT children (see the analysis performed section).
4. There will be variations in the determinants which have an effect of later language development (30 months of age) between the full term and the preterm children.

# 2. Methods

## 2.1. Participants

This study has been carried out using part of the data gathered in a long longitudinal project carried out with an initial sample of 150 low-risk preterm children (PT) and 49 full-term (FT) children who were studied from birth until their 9th birthday. The children and their families were recruited from 4 different hospitals in Galicia (Spain).

For the purposes of the present study, data on language and communicative development gathered at 10, 22, and 30 months of age will be presented. Corrected age has been used for the PT participants.

At 10 months of age the sample comprised 142 PT children, and 49 FT children. There were 45 PT children below 32 weeks of gestation, 36 PT children with a GA of 32 or 33 weeks, and 61 PT children with a GA between 34 and 36 weeks. The next assessment occasion took place when the children were 22 months of age. At this moment, there were 137 PT children, and 43 FT children. There were 43 PT children below 32 weeks of gestation, 36 PT children with a GA of 32 or

33 weeks, and 58 PT children with a GA between 34 and 36 weeks. At 30 months of age the children were assessed again. At this time, the PT sample consisted of 117 children, and the FT sample of 37 children. There were 37 PT children below 32 weeks of gestation, 32 PT children of 32 or 33 weeks of gestation, and 48 PT children with a GA between 34 and 36 weeks.

PT children with further serious complications were excluded from the study. Among the exclusion criteria were babies suffering from periventricular leukomalacia (PVL), intraventricular hemorrhage (IVH) greater than grade II, cerebral palsy (as diagnosed up until 9 months of age), hydrocephalus, encephalopathy, genetic malformations, chromosomal syndromes, metabolic syndromes associated to mental retardation, or important motor or sensorial impairments. Neonates with Apgar scores below 6 at 5 min were also excluded.

Descriptive data of the children at different occasions are shown in Table 1.

Both groups were similar in terms of distribution by gender [$X^2(1) = 0.025$, $p = 0.874$], mothers' education [$X^2(2) = 4.008$, $p = 0.135$] and Apgar score [$t(197) = -0.909$, $p = 0.365$], at the beginning of the study, and throughout the duration of the study.

The former data (Table 1) indicate that the children who still continued in the project at 30 months of age had similar characteristics to the original sample. Thus, there was no substantial change in sample composition throughout time.

Taking into account the Apgar mean score, the inexistence of children with serious medical complications, and the characteristics of their families (mother's education), the sample of PT children may be considered as a low-risk sample.

## 2.2. Instruments

The children participating in the study were assessed at 10, 22, and 30 months of age through the *Inventario para o Desenvolvemento de Habilidades Comunicativas* (IDHC; Pérez-Pereira and García Soto, 2003; Pérez-Pereira and Resches, 2011), a well-known parental inventory which is the Galician version of the MacArthur-Bates Communicative Development Inventories (CDI; Fenson et al., 2007). The form for children between 8 and 15 months (*Palabras e Xestos* "Words and Gestures") of this parental inventory has been administered to the participants' parents when the children were 10 months of age. This form evaluates different aspects of communicative abilities and first language (see Pérez-Pereira and García Soto, 2003; Pérez-Pereira and Resches, 2011; Pérez-Pereira, 2008) for a description of the instrument. From the results obtained, the following measurements have been considered for the analysis: Phrases (understanding of phrases), vocabulary comprehension, vocabulary production, first communicative gestures, games and routines, actions (total score obtained from the sum up of actions with objects, pretending to be a parent, imitating other adult actions).

The form *Palabras e Oracións* (Words and sentences) for children between 16 and 30 months of age was administered to the parents (mainly mothers) of the participants at 22 and 30 months of age. This form assesses different aspects of lexical and grammar development of children (for a description of the instrument see Pérez-Pereira and García Soto, 2003; Pérez-Pereira and Resches, 2011; Pérez-Pereira, 2008). The following measures were used for the analyses: Word production, Use of regular suffixes (forms of words), Mean length of the three longest utterances in words produced by the child (MLU3) and Sentence complexity.

In addition, a complete interview was applied to the mothers in order to get information on the sociodemographic characteristics of the families (educational level of both parents, occupation, family composition, etc.), the health of the children and the caregivers, and other relevant characteristics of the children.

## 2.3. Procedure

Parents' consent, and approval (2008/010) by the Galician Ethics Committee of Clinical Research were obtained before the beginning of the research.

The interview to mothers was administered shortly after the birth of the children, and again at 30 months of age in order to update information.

The IDHC-words and sentences were administered to the parents of the participants when they were 10 months of age (+15 days), while the IDHC-words and sentences were applied when the children were 22 and 30 months of age (+15 days).

## 2.4. Analysis performed

ANOVA analyses have been performed to compare the results obtained by the FT and the PT groups in the different measures taken. The effects of the independent variable (PT vs. FT group) on the following dependent variables have been analysed: understanding of phrases, vocabulary comprehension, vocabulary production, first communicative gestures, games and routines and actions (obtained through the IDHC at 10 months of age); word production, use of regular suffixes, MLU3 and sentence complexity (obtained through the IDHC at 22 and 30 months of age). Previous analysis with the division of the PT children into three different GA groups (<32 weeks, 32–33 weeks, and 34–36 weeks) have not found any significant difference among them; for this reason, all PT children were integrated into a single group.

Repeated measures ANOVAs have been carried out with measures of word production taken at 10, 22, and 30 months of age and with measures of MLU3, sentence complexity and use of regular suffixes, at 22 and 30 months of age, in order to test whether developmental trajectories differed between the two groups (PT vs. FT) or not. Therefore, 2 different models were used: (1) a 2 (age) × 2 (group) repeated measures ANOVA has been used in the case of the measures of which there were two different scores obtained at 22 and 30 months of age: MLU3, sentence complexity and use of regular suffixes; (2) a 3 (age) × 2 (group) repeated measures ANOVA has been used to analyse the scores obtained in word production at 10, 22, and 30 months of age. In this way we could test if there were intra-subjects differences (age related differences in the same participants), inter subjects differences among groups (PT vs. FT), and a combined effect age × group.

Linear regression analyses have been performed to identify those determinants of language measures (dependent variables (DV)) taken at 30 months of age (word production, MLU3, sentence complexity

**TABLE 1** Descriptive data of the sample.

| | *N* | GA mean (SD) | GA range | Apgar | BW mean (SD) | Gender (male) | Maternal education |
|---|---|---|---|---|---|---|---|
| PT newborn | 150 | 32.60 (2.46) | 26–36 | 7.87 (1.43) | 1727 (0.447) | 52.10% | 25.3%[a]<br>39.3%[b]<br>35.3%[c] |
| FT newborn | 49 | 39.84 (1.44) | 37–42 | 8.08 (1.25) | 3,378 (0.414) | 51.00% | 38.8%[a]<br>26.5%[b]<br>34.7%[c] |
| PT 10 m | 142 | 32.61 (2.40) | 26–36 | 7.94 (1.33) | 1718 (0.430) | 52.10% | 23.9%[a]<br>40.1%[b]<br>35,9%[c] |
| FT 10 m | 49 | 39.84 (1.44) | 37–42 | 8.08 (1.25) | 3,378 (0.414) | 51.00% | 38.8%[a]<br>26.5%[b]<br>34.7%[c] |
| PT 22 m | 137 | 32.62 (2.41) | 26–36 | 7.94 (1.30) | 1721 (0.435) | 52.60% | 24.8%[a]<br>40.9%[b]<br>34.3%[c] |
| FT 22 m | 43 | 39.70 (1.48) | 37–42 | 8.13 (1.20) | 3,373 (0.433) | 53.50% | 39.5%[a]<br>23.3%[b]<br>37.2%[c] |
| PT 30 m | 117 | 32.56 (2.49) | 26–36 | 7.94 (1.27) | 1712 (0.428) | 56.50% | 22.6%[a]<br>45.2%[b]<br>32.2%[c] |
| FT 30 m | 37 | 39.76 (1.49) | 37–42 | 8.16 (1.25) | 3,377 (0.443) | 51.40% | 37.8%[a]<br>27.0%[b]<br>35.1%[c] |

GA, Gestational age.
BW, Birth weight.
Maternal education: [a]Basic education.
[b]High school and technical school education.
[c]University degree.

and use of regular suffixes). Forward method has been used. The following measures have been introduced as independent variables. In Block 1 a series of measures taken at the age of 10 months were introduced:

Phrase understanding at 10 months of age.
Word comprehension at 10 months of age.
Word production at 10 months.
First communicative gestures at 10 months.
Games and routines at 10 months.
Total imitation at 10 months of age.
In Block 2, measures taken at 22 months of age were added:
Word production at 22 months of age.
Regular suffixes at 22 months.
MLU3 at 22 months of age.
Sentence complexity at 22 months.

The use of these two blocks allows us to identify the effect of variables taken at a longer distance (10 months of age), the effects of which could not be detected if they were mixed with more proximal variables in the same block.

## 3. Results

Table 2 shows descriptive data and ANOVA results.

As can be observed, no significant difference between the PT and the FT groups is found in any of the measures. Only one trend is found ($p = 0.053$) in Games and routines. Size effects are very low, ranging from 0.008 (MLU3 at 30 months of age) to 0.074 (Games and routines at 10 months), which indicates that the effect of group (PT vs. FT) is minimal on the different measures of language and communicative development taken.

The results of the repeated measures ANOVAs indicate that there is a highly significant effect of age (intra-subjects differences) on Word production [$F(2) = 309.430$, $p < 0.001$, $\eta^2 = 0.805$]; no significant combined effect of age x group is found [$F(2) = 0.901$, $p = 0.408$, $\eta^2 = 0.012$]; no significant difference between groups (PT vs. FT) (inter-subjects effects) is found [$F(1) = 0.145$, $p = 0.704$, $\eta^2 = 0.001$] in word production.

In relation to MLU3 as a dependent variable, the results of the repeated measures ANOVA indicate that there is a highly significant

TABLE 2 Scores of the language measures of the two groups and ANOVA results.

| | | GA group mean (SD) | | | | | |
|---|---|---|---|---|---|---|---|
| | N PT/FT | Preterm | Full-term | *F* | Degrees of freedom | Sign. | Partial eta squared |
| Phrases 10 m | 142/49 | 13.67 (6.5) | 14.45 (6.4) | 0.523 | 190 | 0.470 | 0.036 |
| Word underst. 10 m | 142/49 | 79.17 (74.1) | 71.86 (58.8) | 0.391 | 190 | 0.533 | 0.034 |
| Word product. 10 m | 142/49 | 5.30 (7.7) | 6.39 (21.9) | 0.260 | 190 | 0.610 | 0.030 |
| First gestures 10 m | 142/49 | 7.09 (2.6) | 7.53 (2.5) | 1.035 | 190 | 0.310 | 0.044 |
| Games and rout. 10 m | 142/49 | 4.37 (1.8) | 4.96 (1.6) | 3.794 | 190 | 0.053 | 0.074 |
| Total imitation 10 m | 142/47 | 9.66 (6.4) | 10.93 (7.5) | 1.235 | 188 | 0.268 | 0.048 |
| Word product. 22 m | 137/43 | 158.65 (147.2) | 173.77 (137.1) | 0.356 | 179 | 0.552 | 0.035 |
| Regular suffixes 22 m | 137/43 | 1.53 (2.1) | 1.79 (1.9) | 0.506 | 179 | 0.478 | 0.038 |
| MLU3 22 m | 135/43 | 2.65 (2.1) | 2.69 (2.0) | 0.280 | 177 | 0.597 | 0.033 |
| Sentence compl. 22 m | 137/43 | 2.53 (4.9) | 2.35 (4.3) | 0.048 | 179 | 0.827 | 0.021 |
| Word product. 30 m | 117/37 | 416.19 (175.6) | 411.49 (171.3) | 0.020 | 153 | 0.887 | 0.019 |
| Regular suffixes 30 m | 112/35 | 5.86 (2.8) | 5.20 (3.2) | 1.343 | 146 | 0.248 | 0.062 |
| MLU3 30 m | 106/37 | 7.00 (4.4) | 7.05 (5.7) | 0.003 | 142 | 0.956 | 0.008 |
| Sentence compl. 30 m | 112/35 | 20.81 (14.3) | 20.49 (13.3) | 0.014 | 146 | 0.905 | 0.018 |

effect of age (intra-subjects differences) [$F(1) = 136.055$, $p < 0.001$, $\eta^2 = 0.496$]; no significant combined effect of age × group is found [$F(1) = 0.008$, $p = 0.928$, $\eta^2 = 0.000$]; no significant difference between groups (PT/FT) (inter-subjects effects) is found [$F(1) = 0.070$, $p = 0.791$, $\eta^2 = 0.001$].

In relation to Sentence complexity as a dependent variable, the results of the repeated measures ANOVA indicate that there is a highly significant effect of age (intra-subjects differences) [$F(1) = 208.618$, $p < 0.001$, $\eta^2 = 0.592$]; no significant combined effect of age × group is found [$F(1) = 0.000$, $p = 0.985$, $\eta^2 = 0.000$]; no significant difference between groups (PT/FT) (inter-subjects effects) is found [$F(1) = 0.010$, $p = 0.922$, $\eta^2 = 0.000$].

In relation to the Use of regular suffixes as a dependent variable, the results of the repeated measures ANOVA indicate that there is a highly significant effect of age (intra-subjects differences) [$F(1) = 234.122$, $p < 0.001$, $\eta^2 = 0.619$]; no significant combined effect of age × group is found [$F(1) = 0.2.505$, $p = 0.116$, $\eta^2 = 0.017$]; no significant difference between groups (PT/FT) (inter-subjects effects) is found [$F(1) = 0.2.578$, $p = 0.592$, $\eta^2 = 0.002$].

The results of the longitudinal regression analyses with Word production at 30 months of age as a dependent variable for the preterm and the full-term groups appear in Tables 3, 4, respectively.

For the FT group, from the variables of Block 1 (taken at 10 months of age) First communicative gestures has been selected in Model 1 as having a significant effect on Word production ($p < 0.05$) and explains 17.4% of the variance (R2). In Model 2, First communicative gestures and Word understanding explain 27.5% of the variance (change in R2 increases 10.1% and reaches significance). Model 2 reaches significance ($p < 0.01$). When the variables of Block 2 are considered, the variance of the dependent variable explained is 53.5% and change in R2 increases 26% and reaches significance ($p < 0.001$). The two variables which are significant in Model 2 lose their significance in Model 3, and Word production at 22 months of age is the only variable which has a unique significant effect on word production at 30 months.

For the PT group, Model 1 incorporates Games and routines as a predictive variable of Word production at 30 months of age. The model reaches significance ($p < 0.05$) and explains 5.7% of the variance of the dependent variable. In Model 2 two variables, Games and routines and Word production at 10 months of age, have a significant effect. Model 2 explains 9.3% of the variance of the DV. Change in R2 increments 3.6% and reaches significance ($p < 0.05$). In Model 3 a new variable is included, Word production at 22 months of age, which reaches a high level of significance (standardized β). Now the variable Games and routines loses its significance, however Word production at 10 months continues to have a significant effect as well. The variance explained by Model 3 reaches to 38.3%, and change in R2 increases 29%, and is clearly significant ($p < 0.001$). Finally in Model 4 Use of regular suffixes is added to Games and routines, Word production at 10 months, and Word production at 22 months of age (all of which have a significant effect). Model 4 explains 41% of the variance of the DV and change in R2 reaches 2.7% and is significant ($p < 0.05$).

The results of the longitudinal regression analyses for the preterm and the full-term groups with MLU3 at 30 months of age as a dependent variable appear in Tables 5, 6, respectively.

In relation to the FT group, only MLU at 22 months appears as a predictor of MLU3 at 30 months of age in Model 1. The model reaches significance ($p < 0.01$), and the variance of the DV explained reaches 26%.

As for the PT group the results are similar. The only variable which appears to have effect on the DV is MLU3 measured at 22 months of age. Model 1 explains 56.4% of the variance and its significance level reaches $p < 0.001$.

The results of the longitudinal regression analyses for the preterm and the full-term groups with Sentence complexity at 30 months of age as a dependent variable appear in Tables 7, 8, respectively.

For the FT group, three models are obtained. In Model 1 Word production at 10 months has a significant effect ($p = 0.01$) and explains 19.6% of the variance of Sentence complexity at 30 months of age. In Model 2, a new variable, Games and routines, is added to Word

TABLE 3 Linear regression analysis: predictors of word production at 30months of age: full-term group.

| Predictors | Standardized β | Sig. | $R^2$ | Change in $R^2$ | Change in $F$ | Significance change in $F$ | $F$ | df | $p$ |
|---|---|---|---|---|---|---|---|---|---|
| Model 1 | | | 0.174 | 0.174 | 6.966 | 0.013 | 6.966 | 1.33 | 0.013 |
| First comm. gestures | 0.417 | 0.013 | | | | | | | |
| Model 2 | | | 0.275 | 0.101 | 4.436 | 0.043 | 6.064 | 2.32 | 0.006 |
| First comm. gestures | 0.347 | 0.031 | | | | | | | |
| Word understand. 10 m | 0.325 | 0.043 | | | | | | | |
| Model 3 | | | 0.535 | 0.26 | 17.35 | <0.001 | 11.891 | 3.31 | <0.001 |
| First comm. gestures | 0.238 | 0.073 | | | | | | | |
| Word understand. 10 m | 0.219 | 0.097 | | | | | | | |
| Word production 22 m | 0.537 | <0.001 | | | | | | | |

TABLE 4 Linear regression analysis: predictors of word production at 30months of age: preterm group.

| Predictors | Standardized β | Sig. | $R^2$ | Change in $R^2$ | Change in $F$ | Significance change in $F$ | $F$ | df | $p$ |
|---|---|---|---|---|---|---|---|---|---|
| Model 1 | | | 0.057 | 0.057 | 6.767 | 0.011 | 6.767 | 1.112 | 0.011 |
| Games and routines | 0.239 | 0.011 | | | | | | | |
| Model 2 | | | 0.093 | 0.036 | 4.382 | 0.039 | 5.677 | 2.111 | 0.004 |
| Games and routines | 0.327 | 0.001 | | | | | | | |
| Word production 10 m | −0.209 | 0.039 | | | | | | | |
| Model 3 | | | 0.383 | 0.29 | 51.746 | <0.001 | 22.764 | 3.11 | <0.001 |
| Games and routines | 0.156 | 0.072 | | | | | | | |
| Word production 10 m | −0.229 | 0.007 | | | | | | | |
| Word production 22 m | 0.568 | <0.001 | | | | | | | |
| Model 4 | | | 0.41 | 0.027 | 4.93 | 0.028 | 18.915 | 4.109 | <0.001 |
| Games and routines | 0.168 | 0.050 | | | | | | | |
| Word production 10 m | 0.194 | 0.021 | | | | | | | |
| Word production 22 m | 0.749 | <0.001 | | | | | | | |
| Regular suffixes 22 m | −0.254 | 0.028 | | | | | | | |

production at 10 months. Model 2 explains a higher percentage of the variance of the DV (31.1%). The change in R2 reaches 11.5% and is significant ($p < 0.05$). Finally, in Model 3, Word production at 10 months loses significance ($p = 0.06$), Games and routines continues to have a significant effect and the incorporation of MLU3 at 22 months produces an increment of 26.1% in R2, a change which is clearly significant ($p < 0.001$). Model 3 explains 57.3% of the variance of sentence complexity at the age of 30 months and reaches a high level of significance ($p < 0.001$).

As for the PT group, the only variable which contributes to the explanation of the DV is word production at 22 months of age. In this case, Model 1 explains 36.6% of the variance of sentence complexity at the age of 30 months and the model reaches significance ($p < 0.001$).

The results of the longitudinal regression analyses for the preterm and the full-term groups with Use of regular suffixes as a dependent variable appear in Tables 9, 10, respectively.

For the FT group, two models are obtained. In Model 1 Use of regular suffixes at 22 months of age explains 48% of the variance of the DV Use of regular suffixes at 30 months. Model 1 reaches significance ($p < 0.001$). In Model 2, a new variable is added to the former, MLU3 at 22 months of age. Now the variance explained reaches 57.2%, with an increment in R2 respect to Model 1 of 9.2%, which reaches significance ($p < 0.05$).

In relation to the PT group, 2 models are obtained. Model 1 contains Games and routines, which explains 5.2% of the variance and reaches significance ($p < 0.05$). In Model 2 a new variable is included, Word production at 22 months, and Games and routines loses significance. Model 2 explains 30.3% of the variance and has a significant effect on the use of regular suffixes at 30 months of age ($p < 0.001$). Change in R2 reaches 25.2% and is significant.

## 4. Discussion

In relation to objective 1, the results we found support hypothesis 1, which is that there will not be significant differences between the two groups in the scores obtained in the measures taken at any time. The results of the ANOVA are quite clear, and no significant differences were found between the PT and the FT groups, although the FT children show slightly higher results in all the measures taken

TABLE 5 Linear regression analysis: predictors of MLU3 at 30months of age: full-term group.

| Predictors | Standardized β | Sig. | $R^2$ | Change in $R^2$ | Change in $F$ | Significance change in $F$ | $F$ | df | $p$ |
|---|---|---|---|---|---|---|---|---|---|
| Model 1 | | | 0.26 | 0.26 | 11.601 | 0.002 | 11.601 | 1.33 | 0.002 |
| MLU3 22 months | 0.51 | 0.002 | | | | | | | |

TABLE 6 Linear regression analysis: predictors of MLU3 at 30months of age: preterm group.

| Predictors | Standardized β | Sig. | $R^2$ | Change in $R^2$ | Change in F | Significance change in $F$ | $F$ | df | $p$ |
|---|---|---|---|---|---|---|---|---|---|
| Model 1 | | | 0.564 | 0.564 | 130.421 | <0.001 | 130.421 | 1.101 | <0.001 |
| MLU3 22 months | 0.564 | 0.002 | | | | | | | |

TABLE 7 Linear regression analysis: predictors of sentence complexity at 30months of age: full-term group.

| Predictors | Standardized β | Sig. | $R^2$ | Change in $R^2$ | Change in $F$ | Significance change in $F$ | $F$ | df | $p$ |
|---|---|---|---|---|---|---|---|---|---|
| Model 1 | | | 0.196 | 0.196 | 7.566 | 0.01 | 7.566 | 1.31 | 0.01 |
| Word production 10 m | 0.443 | 0.01 | | | | | | | |
| Model 2 | | | 0.311 | 0.115 | 5.017 | 0.033 | 6.782 | 2.3 | 0.004 |
| Word production 10 m | 0.38 | 0.02 | | | | | | | |
| Games and routines | 0.345 | 0.033 | | | | | | | |
| Model 3 | | | 0.573 | 0.261 | 17.736 | <0.001 | 12.956 | 3.29 | <0.001 |
| Word production 10 m | 0.249 | 0.06 | | | | | | | |
| Games and routines | 0.348 | 0.009 | | | | | | | |
| MLU3 22 months | 0.528 | <0.001 | | | | | | | |

except in word understanding at 10 months, sentence complexity at 22 months and in word production, regular suffixes and sentence complexity at 30 months of age. This indicates that the performance of the PT children seems to improve relatively as they grow older, supporting previous findings, because PT children obtain relatively better results when compared to FT children at 30 months of age (Gayraud and Kern, 2007; Pérez-Pereira, 2021). The fact that prematurity correction for age has been used may be behind these findings, since correction for age is less pertinent, and may have a higher effect, at 30 months of age than at 10 and 22 months of age. In addition, and coherently, size effects were minimal, and always below 0.075. The results found in terms of language development support the findings obtained in other studies carried out with low-risk preterm children (Sansavini et al., 2006; Gayraud and Kern, 2007; Pérez-Pereira et al., 2014; Suttora et al., 2020), and indicate that the results obtained in other studies with very and extremely preterm children, who were found to have smaller vocabulary size and grammatical skills than full-term children (Sansavini et al., 2010, 2011a; Stolt et al., 2012, 2013; Varela-Moraga et al., 2023), cannot be generalized to the overall group of preterm children.

The results found indicate that the precursors of language (use of gestures, participation in social games and routines, and role reversal imitation) are not delayed in the sample of low-risk PT children we studied. Suttora and Salerni (2012), using observational data, also found that there were no differences between preterm and full-term children in their use of communicative gestures at 12, 18, and 24 months of age. Our results are in contradiction with those obtained in another study (Sansavini et al., 2011b) which found a lower production of gestures/actions in the preterm children when compared to full-term children. This study, however, was carried out with very and extremely preterm children, and used the Italian CDI short form (not the complete form).

In relation to objective 2, the results we found indicate that the developmental trend that both groups follow are similar, supporting hypothesis 2. No significant differences were found in any of the repeated measures ANOVA performed on inter-subject differences (group effect) or combined effects of age by group. Therefore, the PT and the FT groups follow similar longitudinal trajectories both in lexical and in morphosyntactic development (MLU3, sentence complexity and use of regular suffixes), thus confirming hypothesis 2. These results agree with those found by Pérez-Pereira and Cruz (2018), who have found a similar pattern of lexical development in 3 groups of low-risk preterm children with different gestational ages and one group of full-term children through a growth curve analysis. On the contrary, the intra-subject effects found were important, which indicates an important change of the linguistic abilities (word production, MLU3, sentence complexity and use of regular suffixes) with age. The greatest intra-subject differences were found in word production ($\eta^2 = 0.805$).

In relation to objective 3, the results found in the regression analyses do not seem to fully support hypothesis 3. Certainly, the use of first gestures has a predictive effect on word production at 30 months of age for the FT group, explaining over 17% of the variance, but not for the PT group. The results found give support to

TABLE 8 Linear regression analysis: predictors of sentence complexity at 30months of age: preterm group.

| Predictors | Standardized β | Sig. | $R^2$ | Change in $R^2$ | Change in *F* | Significance change in *F* | *F* | df | *p* |
|---|---|---|---|---|---|---|---|---|---|
| Model 1 | | | 0.366 | 0.366 | 61.89 | <0.001 | 61.89 | 1.107 | <0.001 |
| Word production 22m | 0.605 | <0.001 | | | | | | | |

TABLE 9 Linear regression analysis: predictors of regular suffixes at 30months of age: full-term group.

| Predictors | Standardized β | Sig. | $R^2$ | Change in $R^2$ | Change in *F* | Significance change in *F* | *F* | df | *p* |
|---|---|---|---|---|---|---|---|---|---|
| Model 1 | | | 0.48 | 0.48 | 28.665 | <0.001 | 28.665 | 1.31 | <0.001 |
| Regular suffixes 22m | 0.693 | <0.001 | | | | | | | |
| Model 2 | | | 0.572 | 0.092 | 6.418 | 0.017 | 20.047 | 2.3 | <0.001 |
| Regular suffixes 22m | 0.428 | 0.011 | | | | | | | |
| MLU3 22months | 0.402 | 0.017 | | | | | | | |

those found in other studies with FT children (Rowe et al., 2008; Rowe and Goldin-Meadow, 2009; Kuvač et al., 2014; Cadime et al., 2017; Silva et al., 2017). For FT children word understanding is also included as a predictive variable in model 2 and has a moderate effect on word production at 30months. Other studies with FT children have also found an effect of word understanding on later word production (Capirci et al., 1996; Bavin et al., 2008; Zambrana et al., 2013). The effect of first communicative gestures on word production at 30months of age in FT children appears as significant when the variables of Block 1 (taken at 10months of age) are introduced as predictors, although this effect disappears when the variables of Block 2 are introduced. Then (Model 3 in Table 3), word production at 22months of age is what has the predominant effect on word production at 30months of age. This indicates that the effect of first communicative gestures fades as the age of the children increases, just as other studies have also found (Cadime et al., 2017). There is no other effect of first communicative gestures on any other grammatical measure at 30months of age (MLU3, sentence complexity and use of regular suffixes). In this sense, the effect of the use of the first communicative gestures seems to be restricted only to lexical development in the case of full-term children.

In relation to the preterm children, the use of gestures has no significant effect on word production at 30months of age, in contrast to what occurs for the low-risk FT group. This result disagrees with those of other studies carried out with very preterm or very low birth weight children which did find an effect of first communicative gestures on later word production (Sansavini et al., 2011a,b; Suttora and Salerni, 2012; Stolt et al., 2014, 2016). Other studies have also found that gestures are not significantly associated to word production in FT children (Dromi and Zaidman-Zait, 2011).

No effect of gestures on grammatical development was found either for the PT group, in contrast with the results found by other studies (Sansavini et al., 2011b; Stolt et al., 2014, 2016), which observed an effect of first communicative gestures on later grammatical development of very preterm children.

The results obtained for the low-risk PT group are considerably different and new. This time the use of first communicative gestures does not seem to have a significant effect on word production at 30months of age, as already commented on; but the participation in games and routines, which has a modest (5.7%) although significant effect on the variance of word production at 30months of age, does have a significant effect. To our knowledge this is the first time that this variable is reported as having a predictive effect on later language development of PT children and coincides with the findings of other studies carried out with FT children (Camaioni and Laicardi, 1985; Dromi and Zaidman-Zait, 2011). The characteristics of social games and routines, with their high predictability which will allow the children to anticipate what will happen next because of their regularity, the use of repetitive language by the mothers and their organization in participants' turns (Bruner, 1983), all together seem to help children to understand and use first language, and surely constitute a very supportive environment. Therefore, the more children participate in social games and routines, the more first words they will use. Apparently, preterm children are particularly benefitted by the supportive context that social games and routines constitute.

In Model 2, word production at 10months of age is also included as a predictor, this time with a negative relationship with word production at 30months of age. In any case its effect seems to be very moderate. Finally, when variables measured at 22months of age are included in Block 2, word production appears as the factor with the greatest influence on later vocabulary. The use of regular suffixes at 22months is also included as having a modest (and negative) significant effect. Therefore, word production measured at 22months has the largest influence on word production at 30months of age for the PT group (similarly to the FT group). The effect of prelinguistic factors, however, still reaches a 0.50 significance level at 30months of age (in the case of PT children), when variables taken at 22months are included (see Model 4 in Table 4). This seems to point to a longer-lasting effect of variables taken at 10months (participation in games and routines and word production at 10months) on vocabulary development at 30months of age in PT children (Rowe and Goldin-Meadow, 2009; Suttora and Salerni, 2012).

The reported effect of imitation on later word production (Carpendale et al., 2005; Zambrana et al., 2013) could not be confirmed in our study.

TABLE 10 Linear regression analysis: predictors of regular suffixes at 30 months of age: preterm group.

| Predictors | Standardized β | Sig. | $R^2$ | Change in $R^2$ | Change in $F$ | Significance change in $F$ | $F$ | df | $p$ |
|---|---|---|---|---|---|---|---|---|---|
| Model 1 | | | 0.052 | 0.052 | 5.825 | 0.017 | 5.825 | 1.107 | 0.017 |
| Games and routines | 0.693 | <0.001 | | | | | | | |
| Model 2 | | | 0.303 | 0.252 | 38.321 | <0.001 | 23.089 | 2.106 | <0.001 |
| Games and routines | 0.047 | 0.583 | | | | | | | |
| Word production 22 m | 0.533 | <0.001 | | | | | | | |

Therefore, hypothesis 3 is only partially confirmed, since first communicative gestures influenced later word production only in FT children, but not in PT children. In addition, this effect is circumscribed to word production, and no other grammatical measures have been affected by the use of first communicative gestures.

The regression analysis for the mean length of the three longest utterances at 22 months of age as DV shows similar results for the FT and the PT groups. In both cases the former measure of this same variable at 22 months is the only predictor which shows a significant effect. The amount of variance explained is 26 and 56.4% for the FT and the PT children, respectively. In this case a specific determinant effect of the same previous measure on a later measure of the same variable is observed. As far as we know, this relationship has not been previously noticed.

The other two grammatical measures obtained at 30 months of age taken as DVs seem to have different predictors for the FT and the PT groups, as hypothesized (hypothesis 4).

In relation to sentence complexity, word production and games and routines are the variables taken at 10 months of age which have a significant effect in the case of FT children. Their effect is noticeable ($R^2 = 0.311$). When the variables measured at 22 months of age are included (Block 2), the variance explained increments of 0.261, and the effect of the three variables reaches 57.3%. Now (Model 3 in Table 7) the variables with a significant effect are MLU3 at 22 months of age and games and routines (word production at 10 months shows a trend). Again, the effect of games and routines lasts up to 30 months of age. In the case of the PT infants, no variable measured at 10 months of age seems to have an effect on sentence complexity, and the only variable which has a significant effect ($R^2 = 0.366$) is word production at 22 months of age.

In relation to the predictors of the use of regular suffixes at 30 months of age, for the FT children no variable of Block 1 seems to have a significant effect, and the only two variables which reach significance are the use of regular suffixes at 22 months of age (Model 1), with an important effect ($R^2 = 0.480$), and MLU3 at 22 months of age, with a modest effect (change in $R^2 = 0.092$). For the PT children, games and routines has a modest, although significant, effect ($R^2 = 0.052$) on regular suffixes at 30 months of age. When the variables of Block 2 are incorporated into the analysis, there is one variable, word production at 22 months of age, which shows a clear significant effect (change in $R^2 = 0.252$), and games and routines loses significance (Model 2 in Table 10). The total variance explained reaches 30.3%.

Therefore, hypothesis 4 is confirmed, since the type of predictive variables which have a significant effect on word production, sentence complexity and use of regular suffixes at 30 months of age vary between the FT and the PT groups. The only exception occurs with MLU3, for which MLU3 at 22 months is the only explanatory variable found for the PT group as well as for the FT group. Its effect, however, is greater for the PT children.

One surprising and original finding is the role played by the participation in games and routines at 10 months of age as a predictor of later grammatical development for the FT and the PT children. In the case of PT children, this predicts the use of regular suffixes while in the case of FT children, its influence, which has a long-lasting effect, is on sentence complexity. Taking into consideration that participation in games and routines also has an influence on the vocabulary development of PT children, this variable stands out as a predictor of later language development. Probably, this is so because in the situations of social games and daily routines, social interaction between the child and the adult is promoted, and the possibilities of the child being exposed to language and using language increase, which will promote language development. Therefore, the higher the participation in games and routines, the better the development of language, as other authors have pointed out (Camaioni and Laicardi, 1985; Tomasello, 2003; Dromi and Zaidman-Zait, 2011; Kuvač et al., 2014; Hsu and Iyer, 2016).

Another difference between the predictors of some grammar development measures in PT and FT children is related to the degree of specificity of the factors. In the case of PT children, MLU3 at 22 months of age is the factor with the highest impact on sentence complexity at 30 months of age, and MLU3 together with regular suffixes at 22 months are the factors which have a significant predictive value for regular suffixes at 30 months of age. Meanwhile, for the low risk PT children, word production at 22 months of age is the only significant predictor of sentence complexity and it is the predictor (together with games and routines) with the highest impact on regular suffixes at 30 months of age. These results indicate that later grammatical development in FT children seems to be more dependent on specifically grammatical antecedents than in PT children, whose grammatical development is more dependent on previous lexical development, with the exception of the MLU3 already commented on.

The use of two blocks of predictive variables for the linear regression analyses allowed us to identify the effect of variables taken at a longer distance (10 months of age), the effects of which would be difficult to detect if they were mixed with more proximal variables in the same block.

## 5. Conclusion

The results found in the ANOVA analyses clearly indicate that low risk preterm children do not seem to have lower performance than full term children in any of the communicative and linguistic measures obtained at any time. In addition, the developmental trajectories of lexical and morphosyntactic abilities followed by the FT and the low-risk PT groups are similar.

There is a difference between the FT and the PT children in the type of predictive variables of later vocabulary production, sentence complexity and regular morphemes used at 30 months of age. Although the use of first communicative gestures does not have effect on later vocabulary development of the PT children, participation in games and social routines does seem to have an influence.

Later grammatical development of the FT ad the low risk PT children seems to be influenced by different previous linguistic abilities, which tend to be more specifically grammatical in the case of FT children.

Obviously, the use of parental reports limits the type of analysis which can be performed, since no information on the frequency of the behaviors studied is possible, and the information provided is prefixed by the instrument. The use of observational methodology would probably report more detailed information on parents-child interactions.

More specific analyses of the type of gestures and the type of words or morphemes used would be necessary to go more in depth into the relationships between preverbal communicative abilities and later language development.

Another limitation of the study is the reduced number of participants in the FT sample, which makes the analyzes less powerful.

## Data availability statement

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

## Ethics statement

The studies involving human participants were reviewed and approved by the Galician Ethics Committee of Clinical Research. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

## Author contributions

MP-P was responsible for the conception of the study, data collection and analysis. AO and MP-P shared responsibility for drafting of the work and final approval of the version to be published.

## 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 MP-P). Funds for open access publication fees were received from the Consellería de Educación, Universidade e Formación Profesional -Xunta de Galicia.

## Acknowledgments

The authors wish to express their gratitude to the children and the families who participated in the study.

## Conflict of interest

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

## Editor's note

Melita Kovacevic edited the article in collaboration with Maria-José Ezeizabarrena, University of the Basque Country UPV/EHU, Spain.

## Publisher's note



## References

Bates, E. (1976). *Language and Context: the Acquisition of Pragmatics.* New York, NY: Academic Press.

Bates, E., Bretherton, I., and Snyder, L. (1988). *From first words to grammar. Individual differences and dissociable mechanisms.* Cambridge: MA: Cambridge University Press.

Bates, E., Camaioni, L., and Volterra, V. (1975). The acquisition of performatives prior to speech. *Merrill-Palmer Q.* 21, 205–226.

Bavin, E. L., Prior, M., Reilly, S., Bretherton, L., Williams, J., Eadie, P., et al. (2008). The early language in Victoria study: predicting vocabulary at age one and two years from gesture and object use. *J. Child Lang.* 35, 687–701. doi: 10.1017/s0305000908008726

Bruner, J. (1983). *Child's talk. Learning to use language*. Oxford: Oxford University Press.

Cadime, I., Silva, C., Santos, S., Ribeiro, I., and Viana, F. L. (2017). The interrelatedness between infants' communicative gestures and lexicon size: a longitudinal study. *Infant Behav. Dev.* 48, 88–97. doi: 10.1016/j.infbeh.2017.05.005

Camaioni, L., and Laicardi, C. (1985). Early social games and the acquisition of language. *Br. J. Dev. Psychol.* 3, 31–39. doi: 10.1111/j.2044-835X.1985.tb00952.x

Capirci, O., Iverson, J. M., Pizzuto, E., and Volterra, V. (1996). Gestures and words during the transition to two word speech. *J. Child Lang.* 23, 645–673. doi: 10.1017/S0305000900008989

Carpenter, M., Nagell, K., and Tomasello, M. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monogr. Soc. Res. Child Dev.* 63:176. doi: 10.2307/1166214

Carpenter, M., Tomasello, M., and Striano, T. (2005). Role reversal imitation and language in typically developing infants and children with autism. *Infancy* 8, 253–278. doi: 10.1207/s15327078in0803_4

Clark, R. A. (1978). The transition from action to gesture. In A. Lock (Ed.), *Action, gesture and symbol: The emergence of language*. London: (Academic Press), 231–257.

Crais, E., Day-Douglas, D., and Cox-Campbell, C. (2004). The intersection of the development of gestures and intentionality. *J. Speech Lang. Hear. Res.* 47, 678–694. doi: 10.1044/1092-4388(2004/052)

Dargue, N., and Sweller, N. (2020). Learning stories through gesture: Gesture's effects on child and adult narrative comprehension. *Educ. Psychol. Rev.* 32, 249–276. doi: 10.1007/s10648-019-09505-0

Dromi, E., and Zaidman-Zait, A. (2011). Interrelations between communicative behaviors at the outset of speech: parents as observers. *J. Child Lang.* 38, 101–120. doi: 10.1017/s0305000909990158

Fantasia, V., Fasulo, A., Costall, A., and Lopez, B. (2014). Changing the game: exploring infants' participation in early play routines. *Front. Psychol.* 16:9. doi: 10.3389/fpsyg.2014.00522

Farkas, C. (2007). Comunicación gestual en la infancia temprana: Una revisión de su desarrollo, relación con el lenguaje e implicancias de su intervención. *Psykhe* 16, 107–115. doi: 10.4067/S0718-22282007000200009

Fenson, L., Dale, P. S., Reznick, J. S., Thal, D., Bates, E., Hartung, J. P., et al. (1993). *MacArthur communicative development inventories. User's guide and technical manual.* San Diego: Singular Publishing Group.

Fenson, L., Marchman, V. A., Thal, D., Dale, P. S., Reznick, J. S., and Bates, E. (2007). *The mac Arthur-Bates communicative development inventories: User's guide and technical manual. 2*. Baltimore: Paul H. Brookes.

Gayraud, F., and Kern, S. (2007). Influence of preterm birth on early lexical and grammatical acquisition. *First Lang.* 27, 159–173. doi: 10.1177/0142723706075790

Goldin-Meadow, S., Levine, S. C., Hedges, L. V., Huttenlocher, J., Raudenbush, S. W., and Small, S. L. (2014). New evidence about language and cognitive development based on a longitudinal study: hypotheses for intervention. *Am. Psychol.* 69, 588–599. doi: 10.1037/a0036886

Goodwyn, S. W., Acredolo, L. P., and Brown, C. A. (2000). Impact of symbolic gesturing on early language development. *J. Nonverbal Behav.* 24, 81–103. doi: 10.1023/A:1006653828895

Hsu, H. C., and Iyer, S. N. (2016). Early gesture, early vocabulary, and risk of language impairment in preschoolers. *Res. Dev. Disabil.* 57, 201–210. doi: 10.1016/j.ridd.2016.06.012

Iverson, J. M., and Goldin-Meadow, S. (2005). Gesture paves the way for language development. *Psychol. Sci.* 16, 367–371. doi: 10.1111/j.0956-7976.2005.01542.x

Iverson, J. M., Longobardi, E., and Caselli, M. C. (2003). 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.* 38, 179–197. doi: 10.1080/1368282031000062891

Jackson-Maldonado, D. (2004). El retraso del lenguaje en niños mexicanos: vocabulario y gestos. *Anuario de psicología* 35:257. doi: 10.1344/%25x

Kuvač, J., Cepanec, M., and Simlesa, S. (2014). Gestual development and its relation to a child's early vocabulary. *Infant Behav. Dev.* 37, 192–202. doi: 10.1016/j.infbeh.2014.01.004

Laing, E., Butterworth, G., Ansari, D., Gsödl, M., Longhi, E., Panagiotaki, G., et al. (2002). Atypical development of language and social communication in toddlers with Williams syndrome. *Dev. Sci.* 5, 233–246. doi: 10.1111/1467-7687.00225

Nelson, K. (1985). *Making sense. The acquisition of shared meaning.* London: Academic Press.

Nelson, K. (2009). Wittgenstein and contemporary theories of word learning. *New Ideas Psychol.* 27, 275–287. doi: 10.1016/j.newideapsych.2008.04.003

Nicoladis, E., Mayberry, R., and Genesee, F. (1999). Gesture and early bilingual development. *Dev. Psychol.* 35, 514–526. doi: 10.1037//0012-1649.35.2.514

Özçalişkan, Ş., Gentner, D., and Goldin-Meadow, S. (2014). Do iconic gestures pave the way for children's early verbs? *Appl. Psycholinguist.* 35, 1143–1162. doi: 10.1017/S0142716412000720

Pérez-Pereira, M. (2008). "Early Galician/Spanish bilingualism: contrasts with monolingualism" in *A portrait of the young in the new multilingual Spain*. eds. C. Pérez-Vidal, M. Juan-Garau and A. Bel (Clevedon: Multilingual Matters), 39–62.

Pérez-Pereira, M. (2021). Prevalence of language delay among healthy preterm children, language outcomes and predictive factors. *Children* 8:282. doi: 10.3390/children8040282

Pérez-Pereira, M., and Cruz, R. (2018). A longitudinal study of vocabulary size and composition in low risk preterm children. *First Lang.* 38, 72–94. doi: 10.1177/0142723717730484

Pérez-Pereira, M., Fernández, P., Gómez-Taibo, M. L., and Resches, M. (2014). Language development of low risk preterm infants up to the age of 30 months. *Early Hum. Dev.* 90, 649–656. doi: 10.1016/j.earlhumdev.2014.08.004

Pérez-Pereira, M., and García Soto, X. R. G. (2003). El diagnóstico del desarrollo comunicativo en la primera infancia: adaptación de las escalas Mac Arthur al gallego. *Psicothema* 15, 352–361.

Pérez-Pereira, M., and Resches, M. (2011). Concurrent and predictive validity of the Galician CDI. *J. Child Lang* 38, 121–140. doi: 10.1017/s0305000909990262

Ratner, N., and Bruner, J. (1978). Games, social exchange and the acquisition of language. *J. Child Lang.* 5, 391–401. doi: 10.1017/S0305000900002063

Rowe, M. L., and Goldin-Meadow, S. (2009). Differences in early gesture explain SES disparities in child vocabulary size at school entry. *Science* 323, 951–953. doi: 10.1126/science.1167025

Rowe, M. L., Özçaliskan, S., and Goldin-Meadow, S. (2008). Learning words by hand: Gesture's role in predicting vocabulary development. *First Lang.* 28, 182–199. doi: 10.1177/0142723707088310

Salo, V. C., Debnath, R., Rowe, M. L., and Fox, N. A. (2023). Experience with pointing gestures facilitates infant vocabulary growth through enhancement of sensorimotor brain activity. *Dev. Psychol.* 59, 676–690. doi: 10.1037/dev0001493

Sansavini, A., Guarini, A., Alessandroni, R., Faldella, G., Giovanelli, G., and Salvioli, G. (2006). Early relations between lexical and grammatical development in very immature Italian preterms. *J. Child Lang.* 33, 199–216. doi: 10.1017/s0305000905007208

Sansavini, A., Guarini, A., and Savini, S. (2011a). Retrasos lingüísticos y cognitivos en niños prematuros extremos a los 2 años: ¿retrasos generales o específicos? [Linguistic and cognitive delays in very preterm infants at 2 years: general or specific delays?]. *Revista de Logopedia, Foniatría y Audiología* 31, 133–147. doi: 10.1016/S0214-4603(11)70182-6

Sansavini, A., Guarini, A., Savini, S., Broccoli, S., Justice, L., Alessandroni, R., et al. (2011b). Longitudinal trajectories of gestural and linguistic abilities in very preterm infants in the second year of life. *Neuropsychologia* 49, 3677–3688. doi: 10.1016/j.neuropsychologia.2011.09.023

Sansavini, A., Savini, S., Guarini, A., Broccoli, S., Alessandroni, R., and Faldella, G. (2010). The effect of gestational age on developmental outcomes: a longitudinal study in the first 2 years of life. *Child Care Health Dev.* 37, 26–36. doi: 10.1111/j.1365-2214.2010.01143.x

Silva, C., Cadime, I., Ribeiro, I., Acosta, V., Lima, R., and Viana, F. L. (2017). Communicative development of Portuguese infants aged between 8 and 15 months. *Revista de Logopedia, Foniatría y Audiología* 37, 121–129. doi: 10.1016/j.rlfa.2016.12.001

Stefanini, S., Bello, A., Caselli, M. C., Iverson, J. M., and Volterra, V. (2009). Co-speech gestures in a naming task: developmental data. *Lang. Cogn. Proc.* 24, 168–189. doi: 10.1080/01690960802187755

Stolt, S., Lehtonen, L., Haataja, L., and Lapinleimu, H. (2012). Development and predictive value of early vocalizations in very-low-birth-weight children: a longitudinal study. *Clin. Linguist. Phon.* 26, 414–427. doi: 10.3109/02699206.2011.648365

Stolt, S., Lind, A., Matomäki, J., Haataja, L., Lapinleimu, H., and Lehtonen, L. (2016). 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.* 61, 16–28. doi: 10.1016/j.jcomdis.2016.03.002

Stolt, S., Makila, A. M., Matomaki, J., Lehtonen, L., Lapinleimu, H., and Haataja, L. (2014). The development and predictive value of gestures in very-low-birth-weight children: a longitudinal study. *Int. J. Speech Lang. Pathol.* 16, 121–131. doi: 10.3109/17549507.2013.794861

Stolt, S., Matomäki, J., Haataja, L., Lapinleimu, H., and Lehtonen, L.PIPARI Study Group (2013). The emergence of grammar in very-low-birth-weight Finnish children at two years of age. *J. Child Lang.* 40, 336–357. doi: 10.1017/s0305000911000456

Suttora, C., Guarini, A., Zuccarini, M., Aceti, A., Corvaglia, L., and Sansavini, A. (2020). 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* 17:7684. doi: 10.3390/ijerph17207684

Suttora, C., and Salerni, N. (2012). Gestural development and its relation to language acquisition in very preterm children. *Infant Behav. Dev.* 35, 429–438. doi: 10.1016/j.infbeh.2012.02.008

Thal, D. J., Tobias, S., and Morrison, D. (1991). Language and gesture in late talkers: a 1-year follow-up. *Journal of Speech & Hearing Research* 34, 604–612. doi: 10.1044/jshr.3403.604

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge: MA: Harvard University Press.

Tomasello, M. (2008). *Origins of human communication*. Cambridge: MA: The MIT Press.

Tomasello, M., and Carpenter, M. (2007). Shared intentionality. *Dev. Sci.* 10, 121–125. doi: 10.1111/j.1467-7687.2007.00573.x

Tomasello, M., Kruger, A. C., and Ratner, H. H. (1993). Cultural learning. *Behav. Brain Sci.* 16, 495–511. doi: 10.1017/S0140525X0003123X

Varela-Moraga, V., Diethelm-Varela, B., and Pérez-Pereira, M. (2023). Effect of biomedical complications on very and extremely preterm children's language. *Front. Psychol.* 14. doi: 10.3389/fpsyg.2023.1163252

Vygotsky, L. S. (1978). *Mind in society: Development of higher psychological processes*. Cambridge, Massachusetts: Harvard University Press.

Zambrana, I. M., Ystrom, E., Schjolberg, S., and Pons, F. (2013). Action imitation at $1^{1/2}$ years is better than pointing gesture in predicting late development of language production at 3 years of age. *Child Dev.* 84, 560–573. doi: 10.1111/j.1467-8624.2012.01872.x

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 24 July 2023
DOI 10.3389/fpsyg.2023.1175658



OPEN ACCESS

EDITED BY
Maria-José Ezeizabarrena,
University of the Basque Country UPV/EHU, Spain

REVIEWED BY
Hila Gendler-Shalev,
University of Haifa, Israel
Philip Dale,
University of New Mexico, United States

*CORRESPONDENCE
Elisabeth Holm
✉ elisabeth.holm@oslomet.no

RECEIVED 27 February 2023
ACCEPTED 09 June 2023
PUBLISHED 24 July 2023

CITATION
Holm E, Hansen PB, Romøren ASH and Garmann NG (2023) The Norwegian CDI-III as an assessment tool for lexical and grammatical development in preschoolers.
*Front. Psychol.* 14:1175658.
doi: 10.3389/fpsyg.2023.1175658



# The Norwegian CDI-III as an assessment tool for lexical and grammatical development in preschoolers

Elisabeth Holm[1]*, Pernille Bonnevie Hansen[2], Anna Sara H. Romøren[1] and Nina Gram Garmann[1]

[1]Department of Early Childhood Education, Oslo Metropolitan University, Oslo, Norway, [2]Department of Scandinavian Languages and Literature, Inland Norway University of Applied Sciences, Hamar, Norway

Parental report instruments are a non-invasive way to assess children's language development and have proved to give both valid and reliable results when used with children under the age of 2;6 (and in some cases up to 3). In this study we examine the newly developed Norwegian edition of a language assessment tool for older preschoolers: *MacArthur-Bates Communicative Development Inventory III* (CDI-III), investigating whether this parental report tool can be used for assessing the language of monolingual Norwegian-speaking children between 2;6 and 4 years. NCDI-III results for 100 children between 2;6 and 4.0 are presented. All sections were significantly intercorrelated. All sections except *Pronunciation* showed growth with age. Internal consistency was measured both in terms of Cronbach's alpha and corrected item-scale correlation, and the results are discussed considering features of item difficulty distribution. Methodological considerations are discussed, as well as implications relevant both for possible later revisions and for CDI-III adaptations to new languages.

KEYWORDS

**language acquisition, parental report, CDI, vocabulary, assessment, preschoolers, item difficulty**

## 1. Introduction

Valid and reliable language assessment tools can be useful for a number of purposes, both for researchers and practitioners. The many language adaptations of the parental report tool *MacArthur-Bates Communicative Development Inventories* (CDI) have been used for, e.g., research on children's language acquisition in specific languages (e.g., Wehberg et al., 2008; Kristoffersen et al., 2012) and cross-linguistic research about children's language development (e.g., Bleses et al., 2008; Frank et al., 2021). Other examples of research are the use of CDIs to investigate how the language development of children with conditions such as cleft lip and palate, autism, impaired hearing, or language delay may differ from that of their typically developing peers (e.g., Scherer and D'Antonio, 1995; Charman et al., 2003). At the same time, CDIs are also used clinically or as screening tools by speech and language practitioners, child healthcare services and others. Some CDI adaptations have been validated with clinical use in mind (Heilmann et al., 2005; Thal et al., 2007; Urm and Tulviste, 2021).

The MacArthur-Bates CDI (Fenson et al., 2007) was constructed to capture reliable, precise and generalizable information about children's early communicative development through a report form filled in by parents (Fenson et al., 1994). The tool is widely recognized as an effective, cost-efficient and valid method for assessing a range of communicative skills in young children, yielding reliable measures of early language development across languages (Law and Roy, 2008). Originally, there were two questionnaires, *CDI I Words and gestures* for children aged 8 to 20 months, and *CDI II Words and sentences* for children up to 30 months. The former comprises a section on early communicative development, including questions about gestures and imitation as well as a vocabulary checklist of about 300 words. In the latter, the vocabulary checklist is twice as long, and includes a list of grammar questions concerning overgeneralizations and sentence complexity as well as a question about the child's three longest utterances. While some sections of the tool can be used to create compound scores, such as an estimated vocabulary size, the aim of its creators was to investigate language development in a broad sense, not to establish an overall measure of language development (Fenson et al., 1994).

Although the tool was meant to give an estimate of several aspects of children's language skills, each section was not meant to give an exhaustive overview. Hence, the CDI vocabulary checklist, despite being quite extensive, was never meant to provide a full overview of any one child's vocabulary, but rather to give an index of their vocabulary knowledge (Fenson et al., 1994). The introduction of shortforms (see, e.g., Jackson-Maldonado et al., 2013) further underlines this principle, as shorter versions of CDIs were developed to limit the workload involved in assessing children's knowledge.

Combining CDI data from multiple languages, researchers have been able to study children's language development across a wide variety of languages and societies. A recurring pattern is the great variation both in the pace of children's language acquisition and in their routes to language learning (Frank et al., 2017). Another common pattern is a clear gender effect: Girls tend to outperform boys on productive vocabulary, whereas this pattern is less clear for word comprehension (Eriksson et al., 2012; Frank et al., 2021). Effects of sociodemographic factors on vocabulary have also been described, but these effects vary between languages (Frank et al., 2021).

## 1.1. The development of CDI-IIIs for 3–4-year-olds

The first attempts at expanding the CDI methodology beyond 2;6 years were made for American English by researchers in the US (Dionne et al., 2003; Feldman et al., 2005; Fenson et al., 2007). They used the same categories of words as in CDI-II, keeping some of the CDI-II words and adding some new ones. This type of CDI-III forms for children up to 4 have subsequently been developed for (Mexican) Spanish (Pilot INV-III, Guiberson, 2008), Basque (Ezeizabarrena et al., 2013; Garcia et al., 2014), and Hungarian (Kas et al., 2022). Whereas CDI I and IIs have been found to represent a valid measure of children's language development of younger children, results from this type of CDI-IIIs have been more mixed, particularly concerning vocabulary (Eriksson, 2017). This was also the case for a previous version of the Norwegian CDI-III, where ceiling effects were found for the vocabulary section, and grammar items were found not to correlate with vocabulary (Sunde et al., 2014).

The requirements of the construction, validation and standardization of any assessment tool depend to a certain extent on what the main purpose of the instrument is intended to be: If it aims to capture the range of variation in language acquisition among children in its target population, the scales must have good discriminatory power across the full range in that population. As assessing vocabulary in older children necessarily entails capturing a very small subset of the children's actual vocabulary (compared to the case for very young children), selecting the appropriate set of words is by no means trivial. To cover the intended age ranges of the CDI-IIIs, the instruments need to include both words that are 'easy' enough to distinguish between children with a relatively small vocabulary and words that are 'sophisticated' enough to distinguish between children with a fairly large vocabulary. That is, they must contain an assembly of items that are easier, items that are harder and items that are somewhere in the middle. At the same time, the instrument should preferably not be too long and extensive to administer.

To handle this problem, Eriksson (2017) proposed that rather than selecting words from a wide range of semantic domains, as do the previous *MacArthur-Bates CDI* instruments, choosing a smaller set of pre-defined themes based on developmental literature and being relatively exhaustive within these topics might be a better approach (2017, p. 648). The Vocabulary section of the Swedish CDI-III adaptation (SCDI-III) is hence built around four domains believed to be central for children in general and where their vocabularies typically expand during the preschool years. These themes are *food words*, *body words*, *mental words* and *emotion words*. The *food* theme is selected because food is an essential part of life and words related to food are usually found in children's early vocabulary. Most of the words about food in the SCDI-III are verbs linked to cooking. The *body* words are selected to include both external and internal body parts, words for health conditions and body functions. Eriksson argues that children often begin to acquire words for external body parts during their second year of life, while words for internal organs, invisible to the child, are more demanding and are acquired at a slow pace. Being more abstract, *mental words* (words about thoughts) and *emotion words* (words about feelings) tend to be acquired from around the age of three, and children tend to acquire them at a slow pace during the preschool years.

The SCDI-III also introduced another invention: a *Metalinguistic awareness* section, included because these are skills generally acquired between 3 and 4 years while also being known to predict literacy (Eriksson, 2017, p. 648).

After conducting a validation and norming study, Eriksson (2017, p. 652) concluded that the structure of the Swedish version of the CDI-III could "well be integrated in similar instruments designed for other languages and cultures," and CDI-III adaptations based on the Swedish version have since been developed for several new languages: Norwegian (Garmann et al., 2019), Polish, Finnish (Stolt, 2023), Estonian (Tulviste and Schults, 2020), (Mexican) Spanish (Jackson-Maldonado et al., 2022), and most recently Ukrainian. Both the Estonian version and the Norwegian one are based directly on the Swedish approach and mostly use the same categories as Eriksson (2017), while still being adaptations, not translations. Table 1 gives a brief comparison of the structures of the American English CDI-III and the Swedish, Estonian, Finnish and Norwegian adaptations. Below, we will present the structures of the Norwegian NCDI-III and the Estonian ECDI-III with the Swedish SCDI-III as a starting point.

### 1.1.1. The structure of the Swedish CDI-III

The SCDI-III consists of sections asking about the child's general level of communication, vocabulary, grammar, metalinguistic

TABLE 1 Comparison of the number of items and score ranges (in parentheses) for the subscales of five CDI-III adaptations.

| CDI-III adaptation | US English | Swedish | Estonian | Finnish | Norwegian |
|---|---|---|---|---|---|
| General communication | N/A[a] | 1 (0–6) | 1 (0–6) | 6 (0–6) | 1 (0–6) |
| Vocabulary | 100 (0–12) | 100 (0–100) | 101 (0–101) | 100 (0–100) | 100 (0–100) |
| Grammar/syntax[b] | 12 (0–12) | 10 + 8 (0–36) | 7 (0–14) + 10 (0–20) | 8 (0–16) + 10 (0–20)[c] | 8 (0–24) + 10 (0–40) |
| Metalinguistic awareness | N/A | 7 (0–7) | 3 (0–3) + 4 (0–4) | 7 (0–7) | 7 (0–7) |
| Pronunciation/phonology | N/A | 1 item (0–2) | 6 (0–7) | 6 (0–7)[c] | 2 (0–5) |
| Semantics/comprehension | 12 (0–12) | N/A | N/A | N/A | N/A |

[a]In the original US English CDI-III form, Fenson et al. (2007) asked whether the child 'is combining words yet'. If not, assessment was aborted after the vocabulary checklist. This question did not contribute to any scale. [b]As discussed in sections 1.1.1–1.1.3, Eriksson (2017) combines his two grammar/syntax scales into one Swedish grammar score, while Tulviste and Schults (2020) and the current paper treat them as two separate scales for Estonian and Norwegian, respectively. [c]Stolt (2023) combines the two Finnish grammar scales *Morphology* and *Language complexity* with *Phonology* to a *Language structures* score (range: 0–43).

awareness, and pronunciation. The first section, *General level of communication*, consists of only one item: Parents are presented with a list of six alternative descriptions and asked to check the one that is true for their child. The alternatives range from 'My child does not speak yet' to 'My child often speaks in long sentences, like […]'. This section has a major ceiling effect and serves as a 'filter': If the parent indicates that the child does not speak, or is impossible to understand, they are not asked to answer any further questions (Eriksson, 2017, p. 649).

Next, there is a vocabulary section consisting of 100 words chosen from words belonging to four semantic domains: *food words* (16 items), *body words* (26 items), *mental words* (30 items), and *emotion words* (28 items). Parents are requested to indicate the words that they have heard the child say.

Two sections both address grammatical complexity: In the *Language complexity* section, respondents are asked to indicate which of two example sentences are more similar to the way their child speaks now. Each of the 10 items consists of one 'simple' and one 'complex' alternative, such as *Jag har choklad* ('I've got chocolate') versus *Jag har choklad på min glass* ('I've got chocolate on my ice cream'). For each item the parent indicates on a three-level scale whether the child mostly speaks in line with the simple or the complex example: 'always left' – 'equally often' – 'always right'. In the *Grammatical constructions* section, 8 items address various grammatical features, such as past tense morphology and passives, illustrated with examples and explanations. To reduce ceiling effects, Eriksson (2017, p. 648) decided to merge the two grammar scales, treating them as one broader *Syntax* scale with scores ranging from 0 to 36.

The SCDI-III *Pronunciation* section consists of only one item, asking the parent to compare the child's speech to that of children of the same age, indicating whether the child sounds a little younger than their peers, similar to their peers, or a little more advanced than most of them. The final section consists of 7 items concerning *Metalinguistic awareness*, where respondents are asked to indicate 'yes' or 'no' to questions concerning various phenomena linked to later literacy – such as whether the child notices similar-sounding words, imitates the way people speak, shows interest in letters, or can write letters or words.

### 1.1.2. The Estonian CDI-III

The Estonian CDI-III (ECDI-III) (Tulviste and Schults, 2020) has kept most of the structure in SCDI-III, with a few changes: The ECDI-III has 7 questions about grammatical constructions where the Swedish form has 8, and the Estonian vocabulary section consists of 101 words instead of 100. Furthermore, the *Metalinguistic awareness* scale is divided into two subscales: *Phonological awareness* (3 items) and *Orthographic awareness* (4 items). Lastly, the *Pronunciation* scale contains 5 new yes/no questions in addition to the original comparison item. These new questions concern "pronunciation difficulties" that the child may have and whether strangers can understand what the child says (Tulviste and Schults, 2020, p. 71).

### 1.1.3. The Norwegian CDI-III

The Norwegian CDI-III (NCDI-III), like the SCDI-III, consists of a 100 words vocabulary checklist in addition to sections with questions about the child's general level of communication, grammatical structures, sentence complexity, pronunciation, and metalinguistic awareness. Each section has a similar structure to that of the SCDI-III, with a few exceptions: The NCDI-III has a 4-level scale for the *Grammatical constructions* items and a 5-level scale for the *Sentence constructions* items, where their Swedish counterparts both have 3-level Likert scales. The NCDI-III scores thus range between 0 and 24 for *Grammatical constructions* and between 0 and 40 for *Sentence complexity*. The NCDI-III *Pronunciation* section consists of two items, where the Swedish has only one. As in the SCDI-III, there is one item asking parents to evaluate the child's speech relative to other children of the same age. In addition, like the Estonian version, the NCDI-III includes an item asking parents to indicate whether people who do not know the child have trouble understanding what the child says (4 levels: 'no, never' – yes, sometimes' – yes, often' – 'yes, always').

## 1.2. Validation studies of the new CDI-IIIs

The SCDI-III was normed and validated as an assessment tool to measure children's language skills by analysing data from 1,134 children aged 2;6 years to 4.0 (Eriksson, 2017). The ECDI-III and the NCDI-III have both been subjected to smaller evaluation studies based on data from 3-year-olds: The ECDI-III using data from 100 Estonian parents of 3-year-olds (Tulviste and Schults, 2020), while the NCDI-III has been piloted on data from parents of 28 Norwegian 3-year-olds (Garmann et al., 2019). However, neither of the two have yet been evaluated for the whole target age span.

All of the published reports have included assessments of the scales' internal consistency, the intercorrelation between scales, and how the scores may be related to demographic factors such as gender and parental education level. For the SCDI-III, Eriksson (2017) also presents age-based norms with percentile levels, as well as examining the dimensionality of each SCDI-III scale. Tulviste and Schults (2020) compare ECDI-III results from children with and without reported difficulties with language development, assessing the sensitivity and specificity of the instrument as a potential screening or diagnostic tool for children with language difficulties. Garmann et al. (2019) compare the NCDI-III reports from staff at the children's Early childhood education and care (ECEC) centers with those from parents, to see if reports from ECEC teachers combined with those from parents can be a reliable way to assess the linguistic development of bilingual children. None of the three studies, however, examine the structure of the *Vocabulary* scale (or any of the other scales) in terms of how easy or difficult the various items are.

### 1.3. The current study

In this study, we examine psychometric and linguistic properties of the NCDI-III based on parental report data covering the full CDI-III age span (2;6–4 years) to answer the following research questions:

1. Do the NCDI-III scales capture growth with age for children aged 2;6 to 4;0 years?
2. Which other demographic factors predict the children's scores?
3. Do the scales correlate with each other?
4. Is there internal consistency within each of the NCDI-III scales (*Vocabulary*, *Grammatical constructions*, *Sentence complexity* and *Metalinguistic awareness*)?
5. What is the pattern of difficulty distribution among the items in each scale of the NCDI-III, and in each thematic *Vocabulary* word group?

CDIs have been used for many purposes. While we do not rule out diagnostic use of the tool once norms are in place and the tool has been tried out also on clinical groups, our focus is in the current paper is descriptive. Based on previous research, we expect higher scores with age on the lexical, grammatical and metalinguistic subscales (question 1) and higher scores among girls than boys (question 2). Furthermore, we expect the strongest correlations between the two grammatical scales, and between these and the vocabulary scale (question 3). Concerning the items within each scale, we expect internal consistency within all scales, with the possible exception of the metalinguistic scale (question 4). None of the previous studies describe the dispersion of item difficulty within the CDI-III scales (question 5).

## 2. Methods and materials

### 2.1. Data

The data used in this paper were collected by student assistants. The participants were recruited via the students' own networks and by contacting ECEC teacher students and ECEC centers, and to a certain degree the recruitment was cumulative, with participants recruiting their own acquaintances (Flygstad and Milder, 2017). The NCDI-III and background forms were administered online, and parents were asked to fill them out digitally at their leisure. Inclusion criteria were that children were to be 'monolingual' (i.e., no household members with other first language than Norwegian), and that parents or ECEC staff reported no concerns regarding the child's language development. All participating children attended kindergarten. All the parents received written information about the project and signed consent forms. Methods for data collection and data processing were developed in line with guidelines from and evaluated by the Data Protection Services at Sikt – Norwegian Agency for Shared Services in Education and Research (formerly NSD – Norwegian Centre for Research Data).

In this paper, we analyze parent-reported NCDI-III data from three age groups: 2;6, 3, and 4 years (see Table 2 for information on number of children, age ranges and gender distributions). Two children were excluded from the data set: One because they did not belong to the age group from which the relevant data set was collected, the other because there seemed to be something wrong with the way the form was completed. The data from the 3-year-olds are also discussed in Garmann et al. (2019), and subsets of the collected data have been used in the student assistants' MA theses.

The background form showed that 34 of the 100 children had no siblings or younger siblings only, while 66 had older siblings or a twin. As for parental education level, there was a skewness toward several years of higher education. Of the 100 participating parents, only 2 reported primary school as their highest education, 14 had upper secondary education, while 3 had higher education of less than 3 years and 28 had 3 years of higher education. More than half the sample, 53 of the 100 parents, had *more than 3 years* of higher education. In comparison, as many as 52–55% of the general population between 20 and 49 years *have no* higher education, according to numbers from Statistics Norway (2022). Among those who do, about 2/3 have studied for 4 years or less [calculated from numbers provided by Statistics Norway (2022); Table 1].

TABLE 2 Age and gender distribution in the dataset.

| Age group | Boys: Girls | Mean age[a] (SD) | Median age (range) |
|---|---|---|---|
| 2;6-year-olds ($n = 36$) | 18:18 | 2;6.8 (45.1 days) | 2;6.13 (2;2.27–2;8.23) |
| 3-year-olds ($n = 28$) | 12:16 | 2;11.24 (28.0 days) | 2;11.23 (2;10.13–3;1.18) |
| 4-years-olds ($n = 36$) | 19:17 | 3;11.30 (27.3 days) | 3;11.30 (3;10.24–4;2.2) |
| Total ($n = 100$) | 49:51 | | |

[a]Age measured in days has been divided by 30.4 to calculate age in months.

Because parents with lower education levels were so few in our dataset, parental education was collapsed into two levels in the analyses, distinguishing only between parents who had completed at least 3 years of higher education (81 parents) and parents with lower levels of education (19 parents).

## 2.2. Analyses

Score distributions in the NCDI-III sections were calculated for each age group. In addition, regression models were used to investigate to what degree the variation in NCDI-III scores is predicted by age along with other demographic factors (gender, sibling status and parental education). There were minimum and/or maximum scores in most of the scales, and we see these as probable results of the scales' boundedness rather than real limits of the constructs measured: values outside the scope of the measure will appear as instances of the minimum value (left-censoring) or the maximum value (right-censoring). In our statistical analyses, we therefore used Tobit regression models (Tobin, 1958), as this method is suitable to estimate the relationship between variables when the dependent variable is censored.

Tobit regression models were used to investigate possible associations between demographic factors and children's NCDI-III scores for *Vocabulary*, *Sentence complexity*, *Grammatical constructions* and *Metalinguistic awareness*. Apart from age, the demographic factors that were investigated were gender, parental education level and sibling status. To check for possible interactions between age and gender, preliminary models were fitted with such interactions as a fifth predictor variable. No significant interaction was found for any of the scales, and we thus report models fitted without interaction. Age measured in days is the basis for the age variable and has been divided by 30.4 to calculate age in months. The parental education item in the questionnaire had five levels, but as the four lower levels were merged into one category, the parental education variable in the regression models is binary, and only distinguishes between parents with more than 3 years of higher education and parents with any lower levels of education.

Intercorrelations among the NCDI-III scales and sections were calculated using Pearson's product–moment correlation and Spearman's rank correlation rho, controlling that there were no large differences. In this paper we report Pearson's correlation coefficients; a comparison with other correlation coefficients can be found in Supplementary Appendix.

Cronbach's alpha is a common way to investigate a scale's internal consistency and thereby its reliability. Internal consistency was calculated in terms of Cronbach's alpha for all scales (*Vocabulary*, *Sentence complexity*, *Grammatical constructions*, and *Metalinguistic awareness*). As Cronbach's alpha is known to be strongly affected by the length of the scales (cf. Streiner, 2003), we also used corrected item-scale-correlation (see DeVellis, 2017) as an additional measure of internal consistency, to further investigate the internal consistency of the scales by a measure that is independent of the number of items in each scale.

The item difficulty distribution of each scale was analyzed, and for the *Vocabulary* section, we also examined the item difficulty profile of each of the four thematic word groups. We calculated difficulty values for each item; first globally, based on the whole sample of participants, and then for each age group separately. The item difficulty profiles of the four thematic word groups were based on the item difficulty values from the full sample.

For the dichotomous items in NCDI-III's *Vocabulary* and *Metalinguistic awareness* scales, item difficulty is reported in terms of proportion values (0.00–1.00), i.e., the proportion of participants indicating that their child had said the word or exhibited the characteristic asked about by the item in question. For non-dichotomous items such as those in the *Sentence complexity, Grammatical constructions* and *Pronunciation* sections, item difficulty is reported in terms of average response values. Mark that this makes the difficulty measure 'inverse', in the sense that a *low* proportion value or a *low* average response value indicate that an item is considered more difficult, as few children received high scores on those items. Correspondingly, items with *high* mean scores or proportion values are considered easy.

Statistical analyses were carried out in R version 4.1.0 (R Core Team, 2021) using RStudio version 1.4.1717 (RStudio Team, 2021). Tobit regressions were modeled using the *AER* package version 1.2–10 (Kleiber and Zeileis, 2008). Cronbach's alpha and item-rest correlations in the NCDI-III scales were calculated using the *psych* package (Revelle, 2021). Density plot and violin plots of item difficulty distributions were made with *ggplot2* (Wickham, 2016). Data editing and other analyses were performed with *dplyr* version 1.0.7 (Wickham et al., 2021) and the *base* package.

# 3. Results

The results are reported in the order of the research questions they address. Growth with age is first reported as seen in isolation and by age group, before age measured in days is presented as part of regression models along with gender, sibling status and parental education level. Correlations between sections are then analyzed, before the internal consistency of each scale is reported both in terms of Cronbach's alpha and in terms of corrected item-scale correlations. Finally, the item difficulty profile of each scale is described – first as calculated for the total sample (including calculations for each thematic word group in the *Vocabulary* scale), and then as calculated for each age group separately – followed by the relationship between item difficulty distribution and item-scale correlation.

## 3.1. Growth with age

All 4-year-olds, and a majority of the other children, reached a maximum score in the single-item *General level of communication* section. None of the children had minimum or maximum scores in the *Vocabulary* section, but in all other sections, there were one or more children with maximum scores. For *Sentence complexity* and *Metalinguistic awareness,* there were some children with zero-scores and some with maximum scores. All sections except *Pronunciation* show growth with age.

As Table 3 shows, there is growth with age in every section except for the *Pronunciation* section, where growth with age was only evident in one of the two items. This item asks whether or not strangers find it difficult to understand the child's speech, and has a score range from 0 ('yes, always') to 3 ('no, never'). Among the

TABLE 3 Median scores and score ranges of NCDI-III for each age group and across age groups.

| NCDI-III scale | 2;6-year-olds | 3-year-olds | 4-year-olds | Total |
|---|---|---|---|---|
| General level of communication (0–5) | 4.5 (3–5) | 5 (3–5) | 5 (5–5) | 5 (3–5) |
| Vocabulary (0–100) | 43 (15–74) | 55 (24–90) | 70.5 (44–93) | 59 (15–93) |
| Sentence complexity (0–40) | 14 (0–32) | 19 (1–40) | 25 (11–38) | 20.5 (0–40) |
| Grammatical constructions (0–24) | 12 (3–23) | 16 (4–24) | 19 (1–24) | 16 (1–24) |
| Pronunciation (0–5) | 4 (1–5) | 3.5 (1–5) | 4 (2–5) | 4 (1–5) |
| Metalinguistic awareness (0–7) | 2 (0–5) | 3 (0–6) | 5 (1–7) | 3 (0–7) |

2;6-year-olds, 5 children (13.9%) were reportedly *always* difficult to understand for strangers. The number was reduced to one child among the 3-year-olds and one among the 4-year-olds. Twenty six of the 2;6-year-olds (72.2%) were reported to be difficult to understand *sometimes*. This was the case for 17 (60.7%) of the 3-year-olds, and for 13 (36.1%) of the 4-year-olds. Only 5 of the 2;6-year-olds (13.9%) were *never* difficult to understand for strangers, while this was the case for 10 (35.5%) of the 3-year-olds, and a total of 22 of the 4-year-olds (61.1%). The answers to the second *Pronunciation* question, asking parents to compare their child's speech to that of other children of the same age, did not show any growth with age; rather, parents of 2;6-year-olds were the most likely to judge their child's speech as resembling that of 'slightly older children'.

## 3.2. NCDI-III results and demographic factors

For *Vocabulary*, both age and gender were significant predictors in the final regression model [$W(4) = 88.25$, $p < 0.001$], while neither parental education nor the existence of older siblings had a significant effect on vocabulary results (see Table 4). As shown in Figure 1, vocabulary scores overall increased with age (1.58 words per month), and girls scored higher than boys (with a gender effect of 8.27 words).

Table 5 shows the final Tobit regression model for *Sentence complexity*, where only age was a significant predictor [$W(4) = 36.96$, $p < 0.001$]. We observed a positive relationship between age and sentence complexity scores, with a sentence complexity increase of 0.65 per month, consistent with the pattern observed for vocabulary.

In the regression model for *Grammatical constructions* [$W(4) = 22.44$, $p < 0.001$], age was once more the only significant predictor (see Table 6). The score for grammatical constructions increased by 0.35 per month. That is, on average, every third month a new item is checked.

Table 7 presents the final regression model for *Metalinguistic awareness* [$W(4) = 80.56$, $p < 0.001$], parental education, age, sibling status and gender were all significant predictors. Age had a positive effect in the model, increasing the metalinguistic awareness score by 0.15 per month. So did parents' education, with children of parents who had completed at least 3 years of higher education scoring 1.18 higher than children whose parents had not. Sibling status had a negative effect of 0.98, meaning that children without older siblings scored higher, and gender had a positive effect of 0.71, meaning that being a girl was associated with higher scores. (The scale range is 0–7, and there were 9 left-censored and 3 right-censored observations).

TABLE 4 Regression model for overall Vocabulary score.

| Variable | B | SE | B | p |
|---|---|---|---|---|
| Constant | −7.399 | 8.443 | −0.876 | 0.381 |
| Education | −1.336 | 3.593 | −0.372 | 0.710 |
| Age (months) | 1.585 | 0.180 | 8.803 | <0.001 |
| Gender | 8.273 | 2.808 | 2.947 | 0.003 |
| Sibling status | 3.290 | 2.983 | 1.103 | 0.270 |
| Log (scale) | 2.630 | 0.071 | 37.187 | <0.001 |

Scale range: 0–100. No left-censored or right-censored observations.

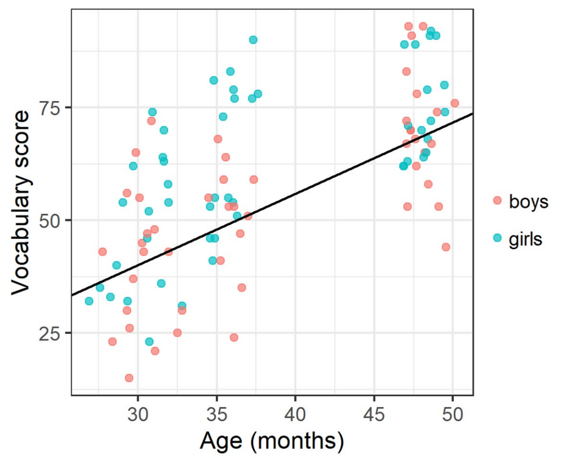


FIGURE 1
Scatterplot of vocabulary score by age and gender, with a fitted line for age (averaged between genders) based on the regression model in Table 4.

TABLE 5 Regression model for Sentence complexity.

| Variable | B | SE | B | p |
|---|---|---|---|---|
| Constant | −6.912 | 5.076 | −1.362 | 0.173 |
| Education | 1.202 | 2.160 | 0.557 | 0.578 |
| Age (months) | 0.650 | 0.108 | 6.016 | <0.001 |
| Gender | 1.874 | 1.686 | 1.112 | 0.266 |
| Sibling status | 0.264 | 1.788 | 0.147 | 0.884 |
| Log (scale) | 2.117 | 0.072 | 29.213 | <0.001 |

Scale range: 0–40. 2 left-censored and 1 right-censored observation.

For the NCDI-III sections *Pronunciation* and *General level of communication*, no regression model was fitted, as the former consists

TABLE 6 Regression model for Grammatical constructions.

| Variable | B | SE | β | p |
|---|---|---|---|---|
| Constant | 1.466 | 3.555 | 0.412 | 0.680 |
| Education | 0.372 | 1.522 | 0.245 | 0.807 |
| Age (months) | 0.353 | 0.076 | 4.647 | <0.001 |
| Gender | 1.246 | 1.186 | 1.051 | 0.293 |
| Sibling status | −0.057 | 1.259 | −0.045 | 0.964 |
| Log (scale) | 1.760 | 0.075 | 23.620 | <0.001 |

Scale range: 0–24. 0 left-censored and 7 right-censored observations.

TABLE 7 Regression model for Metalinguistic awareness.

| Variable | B | SE | β | p |
|---|---|---|---|---|
| Constant | −3.459 | 0.912 | −3.791 | <0.001 |
| Education | 1.174 | 0.389 | 3.021 | 0.003 |
| Age (months) | 0.153 | 0.019 | 7.935 | <0.001 |
| Gender | 0.717 | 0.298 | 2.406 | 0.016 |
| Sibling status | −0.978 | 0.316 | −3.099 | 0.002 |
| Log (scale) | 0.372 | 0.077 | 4.798 | <0.001 |

Scale range: 0–7. 9 left-censored and 3 right-censored observations.

of only two questions and the latter has only one major ceiling effect. One of the two questions in the *Pronunciation* section furthermore differs from the rest of the tool by relating to development relative to the child's age. Here, parents are asked to judge whether the child's speech sounds most like that of younger children, children of the same age, or older children.

TABLE 8 Correlations between the scales of the NCDI-III.

| | Sent. comp. | Gram. constr. | Pronun. | Metaling. |
|---|---|---|---|---|
| Vocabulary | 0.70*** | 0.67*** | 0.36*** | 0.52*** |
| Sentence complexity | | 0.66*** | 0.42*** | 0.44*** |
| Grammatical constructions | | | 0.46*** | 0.45*** |
| Pronunciation | | | | 0.27** |

Pearson's correlation coefficient. ** $p < 0.005$; *** $p < 0.001$.

TABLE 9 Internal consistency of NCDI-III scales as measured by corrected item-scale correlations and Cronbach's *α*.

| | Item-scale correlations | | | α |
|---|---|---|---|---|
| | <0.30 | 0.30–0.50 | >0.50 | |
| Vocabulary (100) | 14 | 37 | 49 | 0.97 |
| Sentence complexity (10) | 0 | 0 | 10 | 0.92 |
| Grammatical constructions (8) | 0 | 1 | 7 | 0.89 |
| Metalinguistic awareness (7) | 2 | 3 | 2 | 0.66 |

The number of items within each scale found within each correlation range. Total number of items for each scale listed in parenthesis.

## 3.3. Correlation between scales

All sections correlated significantly with each other ($p = 0.006$ for the correlation between *Pronunciation* and *Metalinguistic awareness*, $p < 0.001$ for all others). As shown in Table 8 below, the strongest correlations were the ones between the *Vocabulary* section and the *Sentence complexity* ($r = 0.70$), and *Vocabulary* and *Grammatical constructions* ($r = 0.67$), as well as the two grammar scales' correlation with each other ($r = 0.66$). The weakest correlations were found to be those of *Pronunciation* with *Vocabulary* ($r = 0.36$) and *Metalinguistic awareness* ($r = 0.27$) respectively.

## 3.4. Internal consistency

As shown in Table 9, all scales had alpha scores above 0.65. The *Metalinguistic awareness* scale had a lower alpha score than the other scales; only 0.66 [though still 'adequate', in terms of Eriksson, 2017]. The two grammar scales both had high alpha levels: 0.92 and 0.89. The *Vocabulary* section had the highest alpha, at 0.97.

Using corrected item-scale-correlations as a measure of internal consistency, the *Sentence complexity* and *Grammatical constructions* scales were still found to be highly consistent, but the *Vocabulary* section came out as far less consistent than suggested by the alpha score alone. Table 9 shows the distributions of corrected item-scale correlation coefficients for the items in each scale as well as each scale's Cronbach's alpha value.

## 3.5. Difficulty of items

All of the 100 words in the vocabulary section were checked at least once by one of the parents, but none of the words were checked by all parents. Median item difficulty was 0.66, range: 0.04–0.99. (Note that the item difficulty measure is 'inverse', meaning that higher values indicate 'easier' items and vice versa.) The variance in answers was, however, not symmetrically distributed, as shown in Figure 2.

Only 50 of the words had a difficulty in the range between 0.21 and 0.80, meaning that the other half of the words had difficulty levels closer to the ends of the range. Thirty of the 100 words were reported to be said by either 0–10% of the children or 91–100%. Most of these words were found in the easy end of the spectrum, with as many as 23 words reported to be said by 91% or more. Overall, the items of the *Vocabulary* section of the NCDI-III ranged from very easy to very difficult, but with fewer words near the middle range and more words near the poles.

As Figure 3 shows, the four thematic groups of words showed differing difficulty profiles in the NCDI-III results. In general, the *body words* and the *food words* tended to be easier than the words related to thoughts and feelings, with many of the body words concentrated near the easiest limit of the range. The *emotion words* and the *mental words* both had a more balanced distribution between easy and difficult words, but while the mental words showed an almost seamless gradient covering the whole difficulty span, the emotion words were

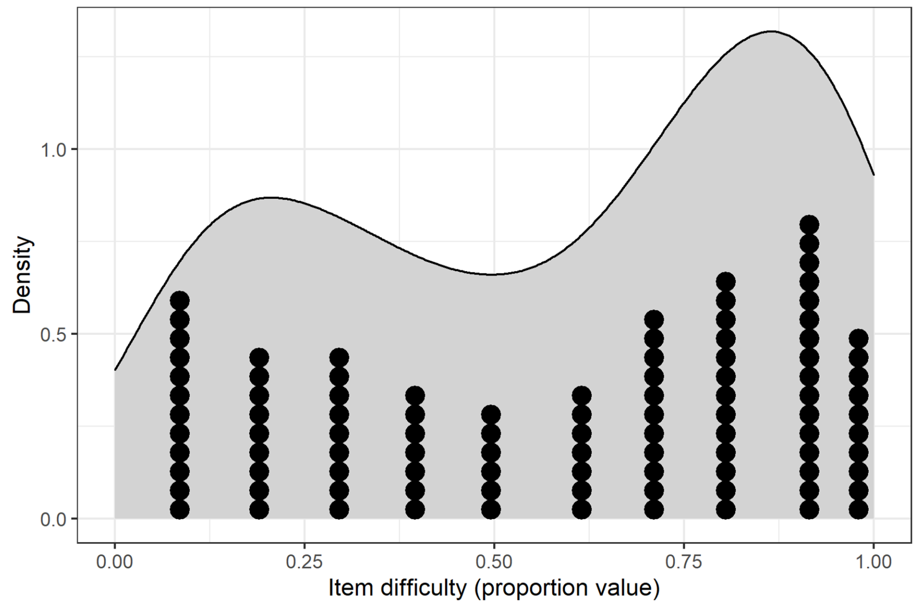


FIGURE 2
Density plot of the levels of item difficulty (proportion values) in the NCDI-III *Vocabulary* section (easier items have higher proportion values), with dots marking each observation. Bin width: 0.1.

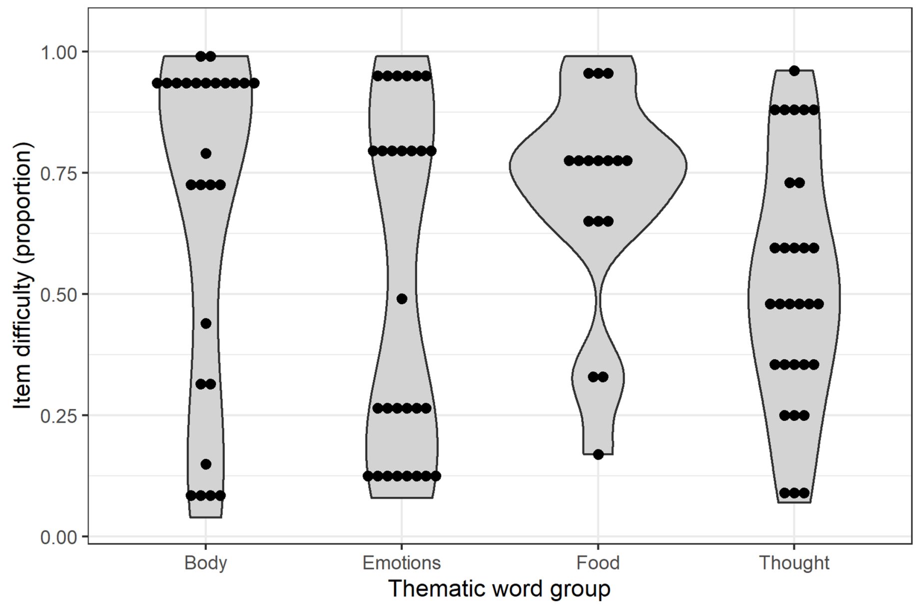


FIGURE 3
Violin plot of difficulty (proportion values) for each word in NCDI-III, by thematic word group (easier items have higher values). Bin width: 0.1.

divided into one set of easier and one set of more difficult words, with only one single word in the center of the scale. Thus, apart from the mental words, we found a split in all word groups, with easy and hard words, and few or none in between.

Among the items in the two grammar scales, there was less variance when it comes to levels of difficulty. The overall easiest item of the *Sentence complexity* scale (item response range: 0–4) had a mean response value of 3.14, and the other 9 items had mean values ranging from 1.56 to 2.29. The median item difficulty of the *Sentence complexity* scale was a mean response value of 1.89. In the *Grammatical constructions* section, the items (response range: 0–3) mainly covered the area from moderate to easy, with the most difficult item having a mean score of 1.34 and the easiest a mean score of 2.39 (median: 2.02).

The *Metalinguistic awareness* section's dichotomous items, on the other hand, seemed to have more dispersed difficulty levels, especially among the older children. One item was exceptionally difficult (0.04) and the rest had proportion values ranging from 0.29 to 0.69 (median: 0.5). The most difficult item – asking whether the child can write some words on their own – was only checked for four children; all of them were 4-year-olds. As was to be expected, the distribution of item difficulty varied between the age groups. Among the 2;6-year-olds, the median *Vocabulary* item difficulty was 0.46 (range: 0–1), among the 3-year-olds it was 0.71 (range: 0–1) and among the 4-year-olds it was 0.86 (range: 0.08–1). The bimodal distribution with relatively few words in the middle of the scale observable in Figure 2 were apparent in all three age groups. For the 2;6-year-olds, there was a skewness toward words being difficult, whereas the opposite was true for the 3-and 4-year-olds.

In each age group, there were words that were reported to be said by every child, and among the 2;6-year-olds and among the 3-year-olds, there were also some words not marked as said by any child in the age group. Among the 2;6-year-olds ($N=36$), 28 words were checked by 10% or less. On the other hand, 11 words were checked by more than 90%. Among the 3-year-olds ($N=28$), 15 words were produced by 10% or less, and 26 words were reported to be produced by more than 90%. Among the 4-year-olds ($N=36$), as many as 43 words were checked by more than 90%. In this age group, only 2 words were checked by 10% or less.

Among the items in the two grammar scales, the item difficulty variance was larger among the youngest children than the older ones. In the *Sentence complexity* scale, the gap between the easiest item and the rest seemed to shrink with age: Among the 2;6-year-olds, the easiest item had a mean response value of 2.86, while the other items' values ranged between 0.83 and 1.83. Among the 3-year-olds, this item's mean response value was 3.29, with the others ranging from 1.43 to 2.29, and among the 4-year-olds, the other items (2.03–2.92) had more or less caught up with it (3.31). The median item difficulty of the *Sentence complexity* scale increased from 1.21 among the 2;6-year-olds to 1.93 among the 3-year-olds and 2.54 among the 4-year-olds. In the *Grammatical constructions* section, the median grew from 1.5 among the 2;6-year-olds (range: 0.89–2.11) to 2.21 among the 3-year-olds (range: 1.68–2.50), and 2.38 among the 4-year-olds (range: 1.47–2.58).

The median proportion value of the *Metalinguistic awareness* scale grew from 0.31 among the 2;6-year-olds (range: 0–0.47) to 0.46 among the 3-year-olds (range: 0–0.79) and 0.72 among the 4-year-olds (range: 0.11–0.94). One of the items, however, asking whether the child divides words into syllables (e.g., *ba-de, sko-le, le-ke*), did not get higher proportion values with age.

### 3.6. The relationship between item difficulty and item-total correlation

For the *Vocabulary* section, we found that more words were located at the easy and difficult ends of the scale than near the center, even though the four thematic word groups had differing difficulty profiles. The relationship between item difficulty and corrected item-scale correlation for all the *Vocabulary* items is shown in Figure 4. There is an abundance of very easy items at the right-hand side of the figure. Those are words that were checked by nearly all participants, and consequentially, they tend to have very low corrected item-scale correlation coefficients.

## 4. Discussion

This paper set out to examine psychometric and linguistic properties of the NCDI-III through statistical analyses of data from

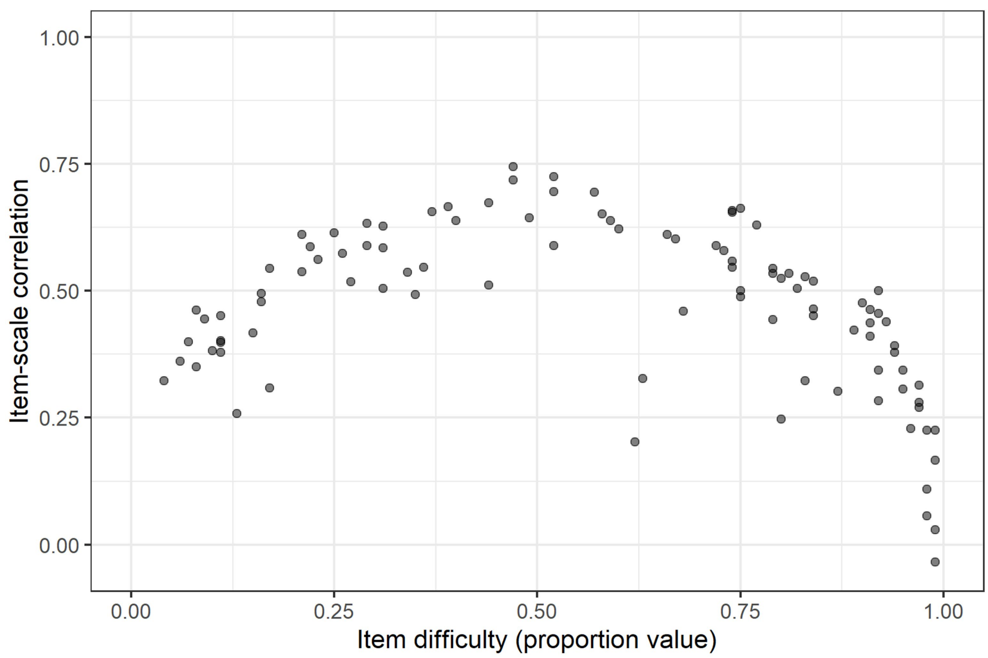


FIGURE 4
Scatterplot of corrected item-scale correlation coefficients for each word in NCDI-III presented by difficulty (easier items to the right, as they have a higher proportion value).

100 children. In line with our expectations, we found higher scores with age, an expected gender effect within some scales, strong correlations between the two grammatical scales, and between these and the vocabulary scale. Furthermore, in line with previous research, there was a high internal consistency as measured with Cronbach's alpha. Regarding dispersion of item difficulty within the vocabulary checklist, we uncovered a bimodal distribution with few words in the mid-difficulty range; globally, within all three age groups and within three of the four thematic word groups. Below, we will discuss our findings in light of previous research, before we elaborate on the bipolar dispersion of difficulty and possible consequences thereof.

## 4.1. Demographics, correlations and consistency

In the *General level of communication* section, all the 4-year-olds, and a majority of the other children, were reported to *often talk in long sentences*, and all the children had a score of 3 or more. This was to be expected, given that no child was included in the study if there was concern about their language development. These results align well with those of (Eriksson, 2017), who reports that 81% of the 1,134 children in the Swedish study "were reported to *talk in long and complicated sentences*" (p. 650), but contrast with findings from Estonian: Only half of Tulviste and Schults' (2020) 3-year-olds reached max score, compared to 25 of the 28 Norwegian 3-year-olds. This difference is striking, even when we take into account that 20 of the Estonian participants were "described by their parents as experiencing difficulties with language development" (p. 69). There may be several possible explanations for the difference. There could be real differences between languages or between populations. Another possible cause is the slightly differing wording of the most advanced alternative, with 'long sentences' in the NCDI-III versus 'complex sentences' in the Estonian counterpart. It is possible that parents' threshold for describing their children's sentences as 'complex' may be higher than the threshold for describing them as 'long.'

There was growth with age in both of the NCDI-III grammar scales, but the *Grammatical constructions* scale had a considerable ceiling effect both among the 3-year-olds and 4-year-olds. This is in line with Tulviste and Schults' (2020) findings from Estonian three-year-olds where there was a ceiling effect in the grammatical constructions scale. The SCDI-III grammar scores were merged as a means to resolve issues with ceiling effects (Eriksson, 2017, p. 648). This approach seems reasonable also for Norwegian, but it presupposes weighting two differently scored grammar scales against each other. Whereas the two grammar scales in the Swedish CDI-III have the same answer structure and almost the same maximum scores (20 and 16), the two grammar scales in the Norwegian version differ in answer structures. Consequently, there is a gap in maximum scores, so that the NCDI-III *Sentence complexity* scale's maximum is 40 while the *Grammatical constructions* scale has a maximum score of only 24. Analyses from an ongoing project on the relationship between vocabulary and grammar in 1–4-year-olds indicate that transforming and merging the two grammar scales resolves ceiling issues also for Norwegian (Holm and Hansen, in progress).

The two items in the NCDI-III *Pronunciation* section also have different measurement scales: The first item's scale is absolute, asking whether strangers find the child's speech difficult to understand. As expected, we found growth of age in the answers to this question. The second question factors in age, and thus no growth with age was expected: Parents are asked to *compare the child's speech to the speech of other children of the same age*. Interestingly, 49 per cent of both Norwegian and Swedish parents report their child to sound like slightly older children, while only 10 per cent in both groups report their child to sound a little younger. Eriksson (2017, p. 652) points out that parental overestimation of their child's pronunciation may represent a familiarity effect: The parents are used to their own child's way of speaking, and thus find them easier to understand than other children of the same age.

In line with the Estonian and Swedish results, the Norwegian girls outperformed the boys on *Vocabulary* and *Metalinguistic awareness*. We did however not find any effects of gender on grammar, in contrast to Eriksson (2017). A recent review on gender effects in early language acquisition suggests that gender differences may differ across ages and language domains (Rinaldi et al., 2021); to determine whether gender effects differ between lexical and grammatical knowledge in Norwegian-speaking children, a larger sample might be needed. Like Eriksson (2017) found for the SCDI-III, both birth order and parental education level predicted *Metalinguistic awareness,* with firstborns outperforming laterborns and children with higher-educated parents scoring higher than children with lower-educated parents. *Metalinguistic awareness* thus seems to be sensitive to all demographic variables studied here. Given the skewness toward higher education in the dataset, we cannot conclude that parental education level does not have any influence on the other scales – only that these scales do not appear to distinguish between children of parents with more and less than 3 years of higher education.

Assuming that language skills consist of several different types of abilities where some are more closely interrelated than others, we should expect some parts of the NCDI-III to be more strongly intercorrelated than others. In line with Eriksson (2017) and Tulviste and Schults (2020) as well as a vast literature on younger children, we expected a stronger correlation between vocabulary and grammar. Our findings met our expectations: No correlations were stronger than those between *Vocabulary* and the two grammar scales, and there were only weak correlations between *Pronunciation* and *Vocabulary*, as well as between *Pronunciation* and *Metalinguistic awareness*.

All NCDI-III scales had Cronbach's alpha values within Eriksson's (2017) suggested adequacy threshold of >0.65, although *Metalinguistic* awareness only barely so. As the alpha is affected by length, the very high *Vocabulary* score (0.97) could be at least partially attributed to the fact that it is very long (100 items). While a high alpha score is often considered 'excellent' and not discussed further, Streiner (2003) points out that a very high alpha could be a sign of redundancy. A different picture appeared when investigating internal consistency through item-scale correlations: In terms of corrected item-scale correlations, there is lower internal consistency within *Vocabulary* than within *Grammatical constructions* and *Sentence complexity*, even if *Vocabulary* had by far the highest alpha.

The *Metalinguistic awareness* section showed a relatively low degree of internal consistency, both in terms of Cronbach's alpha and in terms of corrected item-total correlations. Eriksson (2017) points out that his results suggest that the *Metalinguistic awareness* scale "taps into a slightly different set of knowledge" than the vocabulary and grammar scales (p.652). This holds also for our Norwegian results, as the scale shows a limited correlation with other sections and is the only scale predicted by all demographic variables in our model. Furthermore, the lower internal consistency suggests that the *Metalinguistic awareness* section is better treated as an assembly of useful questions than as a scale. Each question may still give valuable information about a child's metalinguistic and pre-literacy development.

## 4.2. Difficulty dispersion

The difference noted for *Vocabulary* between the two consistency measures is connected to its difficulty distribution. As the *Vocabulary* list is fixed while children's vocabularies grow with age, the word list needs to include both words that are easy enough to distinguish between the youngest children and words that are sufficiently difficult to distinguish between the 4-year-olds. There is thus an inevitable tradeoff between each item's overall discriminatory power and the total scale's ability to capture variance across the whole age span. Very easy and very difficult items will necessarily have a low response variance and thereby a low overall discriminatory effect and weak item-total correlation. As pointed out by deVellis (2017, p. 143), "an item that does not vary cannot covary." Hence, some redundancy in the scale was expected, especially among the youngest and oldest children. More surprising was the overall bimodal difficulty distribution, with few intermediately difficult words both within and across age groups, and a large number of items near the poles of the difficulty continuum. Particularly in the 'easy' end of the range, there were many words with a very weak item-scale correlation.

The high Cronbach's alpha combined with the low consistency in terms of item-scale correlations suggest that the *Vocabulary* scale may be unnecessarily long, and that excessive words could be removed from the easy end of the scale without damaging the instrument's ability to distinguish between children. Alternatively, one could remove some of the easiest and maybe also some of the hardest words and replace them with words in the medium difficulty range. However, as the sample of participants in this study was skewed toward higher levels of education, and as children whose caregivers were concerned about their linguistic development were excluded from the sample, these tendencies might not be as strong in a more representative sample. A recent response to the issue of redundancy in CDI word lists is the development of adaptive versions based on existing CDI data. Here, parents respond to a dynamic word list that adjusts to their responses, meaning that the researcher achieves the wanted information about each child through far fewer questions (e.g., Mayor and Mani, 2019; Kachergis et al., 2022; Mieszkowska et al., 2022). Such adaptive versions are built on data on large numbers of words from a substantial pool of children. Static forms such as NCDI-III, despite their higher level of redundancy, still offer the advantage of being less resource-intensive in development and standardization, and may be administered without access to digital technology.

The most frequent words in a given language are the ones that tend to be acquired earlier. These constitute a much smaller set of words than the vast amount of less frequent words. As the NCDI-III *Vocabulary* scale is meant to be a proxy measure for a continuous underlying construct – children's vocabulary size – this poses a challenge: Each of the more 'sophisticated' words would generally be sampled from larger and more diverse possible vocabularies than the easy ones, making the use or lack of use of any specific difficult word less representative of a child's total vocabulary. The NCDI-III follows Eriksson's (2017) response to this issue, sampling words from a smaller set of topics based on developmental literature. Our finding that the NCDI-III's body words and food words generally tended to be easier than the words related to thoughts and feelings agrees well with his assumptions. Further, when it comes to body words, most of the NCDI-III words for internal body parts were found among the difficult words.

## 4.3. Limitations

Ideally, a validation or norming sample should resemble the population for which the instrument is meant to be used, and recruitment methods and criteria for inclusion of participants in validation and norming studies thus have consequences for the appropriateness of a tool (see, e.g., Friberg, 2010) In the current study, there is an unusually large overrepresentation of children from families with high education levels. Furthermore, children about whom language concerns had been expressed were left out of the study entirely. These biases may influence our results: The fact that parental education did not appear as a significant predictor of children's results in most of the NCDI-III scales should not be taken as evidence that parental education cannot predict NCDI-III scores in the general population. To see how well the NCDI-III captures the full breadth of variation found in the Norwegian child population from 2;6 to 4 years, will require further research with a wider selection of participants. Our conclusions regarding the distribution of item difficulty must also be made with the caveat that levels of difficulty calculated from a more representative sample of participants may differ somewhat from what we present in this study.

In the current study, Cronbach's alpha is used as a measure of the internal consistency –and thereby reliability – of the NCDI-III scales. However, Cronbach's alpha alone does not give us the *dimensionality* of a scale. It is perfectly possible for a scale measuring two or more underlying constructs to obtain a high alpha, especially if the scale has many items (Streiner, 2003). In order to establish the dimensionality of the NCDI-III scales, a larger study with more participants will be required. However, as Norwegian and Swedish are very closely related languages, and as the Norwegian *Vocabulary* scale is closely modeled on the Swedish one, there is reason to believe that its dimensionality is close to what Eriksson (2017) found for the Swedish adaptation.

The CDI III has so far focused on monolingual children in Western, industrialized, rich and democratic countries. Data from other populations are necessary to assess if generalizations about children's language acquisition using parental reports with children between 2;6 and 4 are valid also outside of the WEIRD context (Henrich et al., 2010).

## 4.4. Summary

In this paper, we have evaluated the NCDI-III based on a sample of monolingual children between 2;6 and 4 years of age, finding psychometric properties quite similar to what Eriksson (2017) and Tulviste and Schults (2020) have reported for Swedish and Estonian respectively: There is growth with age, and girls outperform boys in the *Vocabulary* and *Metalinguistic awareness* results. There was adequate internal consistency within all scales in terms of Cronbach's alpha, although less so for *Metalinguistic awareness*. Ceiling effects in the grammar scales could possibly be amended by merging the two scales, but one would have to decide on how to weigh the items from each grammatical scale against the other.

Based on findings from our analyses of item difficulty distribution and internal consistency, we suggest that subsequent CDI-III adaptations may benefit from paying attention to the difficulty profiles of the scales, preferably avoiding items too close to the poles of the difficulty range.

## Data availability statement

The dataset underlying the analyses in this article will be made available through the Wordbank database (http://wordbank.stanford.edu/data).

## Ethics statement

The studies involving human participants were reviewed and approved by the Norwegian Data Protection Services, SIKT (Norwegian Agency for Shared Services in Education and Research, formerly NSD, Norwegian Centre for Research Data) under the number 788858. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

## Author contributions

NG and AR had the idea for an evaluation of the NCDI-III. EH was recruited to carry out the study. The data collection was initiated by the research project at OsloMet, led by NG with AR and PH as close collaborators, and carried out by master's students. EH organized the last round of data collection, and proposed and conducted the analyses of item difficulty profiles, along with using Tobit regression for the modeling of relationships between NCDI-III results and demographic factors. EH, PH, and AR wrote summaries of background literature for the other co-authors. EH prepared and carried out the statistical analyses, discussing them with the others, especially PH, in the process. PH contributed substantially to the analyses, with practical knowledge and in discussions of methods, interpretations of results, and preparation of figures and tables. EH drafted most of the article text, but all co-authors have drafted paragraphs. All co-authors have discussed and revised the text in collaboration, in several rounds. All authors have suggested literature, participated in discussions about possible designs for the manuscript and research questions, and read through the final draft of the article before submitting it for peer review.

## Funding

The research has been mainly funded by Faculty of Education and International Studies (LUI), OsloMet - Oslo Metropolitan University, and publishing is funded by Department of Research and Development, OsloMet - Oslo Metropolitan University.

## Acknowledgments

First of all, we wish to thank the families who have participated in this study. We would also like to thank Mia Helene Aarre, Stine Helen Nævra Bakke, Malene Bøyum, Nina Flygstad, Marthe Gullvik, Mari Huse Heggedal, Grethe Johanne Imislund, Anette Løkkevol, Nina Milder, Runa Brevik Mykland, Christine Fenne Nielsen, Ingrid L. S. Pedersen, Synnøve Pettersen, Camilla Nordahl Sagen, and Lene Støvern for collecting the data for this study. We are also grateful for Eirik Tengesdal and Malene Bøyum who worked with the technical solutions on the questionnaires. Eirik also helped out with the organization of the data. We appreciate all the helpful discussions with and comments from the members of the SAMBA research group at OsloMet, particularly Ove Edvard Hatlevik, Joakim Evensen Hansen, Maren Meyer Hegna, Elisabeth Bjørnestad, and Hanne Næss Hjetland. We would also like to thank the audience at the EUNM-CDI seminar in 2022 for helpful comments on our work, and the two reviewers and editor who carefully read and gave feedback on the current paper.

## Conflict of interest

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

## Editor's note

Maria-José Ezeizabarrena edited the article in collaboration with Melita Kovacevic, University of Zagreb, Zagreb, Croatia.

## Publisher's note



## Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2023.1175658/full#supplementary-material

## References


Bleses, D., Vach, W., Slott, M., Wehberg, S., Thomsen, P., Madsen, T. O., et al. (2008). Early vocabulary development in Danish and other languages: a CDI-based comparison. *J. Child Lang.* 35, 619–650. doi: 10.1017/S0305000908008714

Charman, T., Drew, A., Baird, C., and Baird, G. (2003). Measuring early language development in preschool children with autism spectrum disorder using the MacArthur communicative development inventory (infant form). *J. Child Lang.* 30, 213–236. doi: 10.1017/s0305000902005482

DeVellis, R. F. (2017). *Scale development: Theory and applications*, vol. *26. 4th Edn*. Los Angeles: SAGE.

Dionne, G., Dale, P. S., Boivin, M., and Plomin, R. (2003). Genetic evidence for bidirectional effects of early lexical and grammatical development. *Child Dev.* 74, 394–412. doi: 10.1111/1467-8624.7402005

Eriksson, M. (2017). The Swedish communicative development inventory III: parent reports on language in preschool children. *Int. J. Behav. Dev.* 41, 647–654. doi: 10.1177/0165025416644078

Eriksson, M., Marschik, P. B., Tulviste, T., Almgren, M., Pérez Pereira, M., Wehberg, S., et al. (2012). Differences between girls and boys in emerging language skills: evidence from 10 language communities. *Br. J. Dev. Psychol.* 30, 326–343. doi: 10.1111/j.2044-835X.2011.02042.x

Ezeizabarrena, M.-J., Barnes, J., García, I., Barreña, A., and Almgren, M. (2013). "Using parental report assessment for bilingual preschoolers: the Basque experience" in *Solutions for the assessment of bilinguals*. ed. V. C. M. Gathercole doi: 10.21832/9781783090150

Feldman, H. M., Campbell, T. F., Kurs-Lasky, M., Rockette, H. E., Dale, P. S., Colborn, D. K., et al. (2005). Concurrent and predictive validity of parent reports of child language at ages 2 and 3 years. *Child Dev.* 76, 856–868. doi: 10.1111/j.1467-8624.2005.00882.x

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., and Pethick, S. J. (1994). Variability in early communicative development. *Monogr. Soc. Res. Child Dev.* 59:i. doi: 10.2307/1166093

Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, J. S., and Bates, E. (2007). *MacArthur-Bates communicative development inventories*. Baltimore, MD: Paul H. Brookes Publishing Company.

Flygstad, N., and Milder, N. (2017). Utprøving av CDI-III for norske enspråklige treåringer: En pilotstudie. Master's thesis, University of Oslo.

Frank, M. C., Braginsky, M., Yurovsky, D., and Marchman, V. A. (2017). Wordbank: an open repository for developmental vocabulary data. *J. Child Lang.* 44, 677–694. doi: 10.1017/S0305000916000209

Frank, M. C., Braginsky, M., Yurovsky, D., and Marchman, V. A. (2021). *Variability and consistency in early language learning: The Wordbank project*. Cambridge, MA: MIT Press.

Friberg, J. C. (2010). Considerations for test selection: how do validity and reliability impact diagnostic decisions? *Child Lang Teach Ther* 26, 77–92. eric. doi: 10.1177/0265659009349972

Garcia, I., Barreña, A., Ezeizabarrena, M.-J., Almgren, M., Arratibel, N., and Barnes, J. (2014). Haur euskaldunen komunikazio-garapena neurtzen 30-50 hilabete bitartean: MacArthur-Bates CDI-III tresnaren euskal bertsioa [assessing the communicative development of 30 to 50-month old Basque children: the Basque version of the MacArthur-Bates CDI-III]. *Uztaro* 88, 33–72. doi: 10.26876/uztaro.88.2014.3

Garmann, N. G., Romøren, A. S. H., Flygstad, N., Milder, N., Pedersen, I., Simonsen, H., et al. (2019). Språkkartlegging i norske barnehager - en pilotundersøkelse av foreldre og barnehageansattes rapportering av norsktalende tre-åringers språkferdigheter ved hjelp av CDI III. *RASK: Internationalt Tidsskrift for Sprog Og Kommunikation*, 49, 87–103.

Guiberson, M. (2008). Validity of a parent vocabulary checklist for young Spanish speaking children of Mexican immigrants. *Int. J. Speech Lang. Pathol.* 10, 279–285. doi: 10.1080/17549500802216763

Heilmann, J., Ellis Weismer, S., Evans, J., and Hollar, C. (2005). Utility of the MacArthur-Bates communicative development inventory in identifying language abilities of late-talking and typically developing toddlers. *Am. J. Speech Lang. Pathol.* 14, 40–51. doi: 10.1044/1058-0360(2005/006)

Henrich, J., Heine, S. J., and Norenzayan, A. (2010). The weirdest people in the world? *Behav. Brain Sci.* 33, 61–83. doi: 10.1017/S0140525X0999152X

Holm, E., and Hansen, P. (in progress). *The relationship between vocabulary and grammar in Norwegian 1;6-4-year-olds (manuscript in preparation)*.

Jackson-Maldonado, D., Marchman, V. A., Dale, P., and Rubio-Codina, M. (2022). *The MacArthur-Bates CDI-III for Spanish-speaking children*.

Jackson-Maldonado, D., Marchman, V. A., and Fernald, L. C. H. (2013). Short-form versions of the Spanish MacArthur–Bates communicative development inventories. *Appl. Psycholinguist.* 34, 837–868. doi: 10.1017/S0142716412000045

Kachergis, G., Marchman, V. A., Dale, P. S., Mankewitz, J., and Frank, M. C. (2022). Online computerized adaptive tests of Children's vocabulary development in English and Mexican Spanish. *J. Speech Lang. Hear. Res.* 65, 2288–2308. doi: 10.1044/2022_JSLHR-21-00372

Kas, B., Jakab, Z., and Lőrik, J. (2022). Development and norming of the Hungarian CDI-III: a screening tool for language delay. *Int. J. Lang. Commun. Disord.* 57, 252–273. doi: 10.1111/1460-6984.12686

Kleiber, C., and Zeileis, A. (2008). Applied econometrics with {R}. Springer-Verlag. Available at: https://CRAN.R-project.org/package=AER

Kristoffersen, K. E., Simonsen, H. G., Eiesland, E. A., and Henriksen, L. Y. (2012). Utvikling og variasjon i kommunikative ferdigheter hos barn som lærer norsk – en CDI-basert studie. *Norsk Tidsskrift Logopedi* 58, 34–43.

Law, J., and Roy, P. (2008). Parental report of infant language skills: a review of the development and application of the communicative development inventories. *Child Adolesc. Mental Health* 13, 198–206. doi: 10.1111/j.1475-3588.2008.00503.x

Mayor, J., and Mani, N. (2019). A short version of the MacArthur–Bates communicative development inventories with high validity. *Behav. Res. Methods* 51, 2248–2255. doi: 10.3758/s13428-018-1146-0

Mieszkowska, K., Krajewski, G., Król, P., Krysztofiak, M., and Haman, E. (2022). Developing adaptive CDIs in polish and the CDI-online web app. In *Proceeding of EUNM-CDI 2022: The eighth European network meeting on communicative development inventories*, Dubrovnik, Croatia.

R Core Team. (2021). R: A language and environment for statistical computing [manual]. R Foundation for Statistical Computing. Available at: https://www.R-project.org/

Revelle, W. (2021). Psych: procedures for psychological, psychometric, and personality research. Northwestern University. Available at: https://CRAN.R-project.org/package=psych

Rinaldi, P., Pasqualetti, P., Volterra, V., and Caselli, M. C. (2021). Gender differences in early stages of language development. Some evidence and possible explanations. *J. Neurosci. Res.* 101, 643–653. doi: 10.1002/jnr.24914

RStudio Team. (2021). RStudio: Integrated development environment for R [manual]. RStudio, PBC. Available at: http://www.rstudio.com/

Scherer, N., and D'Antonio, L. (1995). Parent questionnaire for screening early language development in children with cleft palate. *Cleft Palate Craniofacial J* 32, 7–13. doi: 10.1597/1545-1569(1995)032<0007:PQFSEL>2.3.CO;2

Statistics Norway. (2022). Educational attainment of the population. Available at: https://www.ssb.no/en/utdanning/utdanningsniva/statistikk/befolkningens-utdanningsniva

Stolt, S. (2023). Internal consistency and concurrent validity of the parental report instrument on language in pre-school-aged children – The Finnish Communicative Development Inventory III. *First Language*, 01427237231167301. doi: 10.1177/01427237231167301

Streiner, D. L. (2003). Starting at the beginning: an introduction to coefficient alpha and internal consistency. *J. Pers. Assess.* 80, 99–103. doi: 10.1207/S15327752JPA8001_18

Sunde, J. A., Erikstad, N. H., and Andersen, K. (2014). *Validering av MacArthur-Bates Communicative Development Inventories III [CDI-III] for norsk enspråklige og engelsk-norsk tospråklige barn mellom tre og fire år – En pilotstudie* Master's thesis. Bergen: University of Bergen.

Thal, D., DesJardin, J. L., and Eisenberg, L. S. (2007). Validity of the MacArthur-Bates communicative development inventories for measuring language abilities in children with Cochlear implants. *Am. J. Speech Lang. Pathol.* 16, 54–64. doi: 10.1044/1058-0360(2007/007)

Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica* 26, 24–36. doi: 10.2307/1907382

Tulviste, T., and Schults, A. (2020). Parental reports of communicative development at the age of 36 months: The Estonian CDI-III. *First Language*, 40, 64–83. doi: 10.1177/0142723719887313

Urm, A., and Tulviste, T. (2021). Toddlers' early communicative skills as assessed by the short form version of the Estonian MacArthur-Bates communicative development inventory II. *J. Speech Lang. Hear. Res.* 64, 1303–1315. doi: 10.1044/2020_JSLHR-20-00201

Wehberg, S., Vach, W., Bleses, D., Thomsen, P., Madsen, T. O., and Basbøll, H. (2008). Girls talk about dolls and boys about cars? Analyses of group and individual variation in Danish children's first words. *First Lang.* 28, 71–85. doi: 10.1177/0142723707081729

Wei, T., and Simko, V. (2021). *R package 'corrplot': visualization of a correlation matrix*. Available at: https://github.com/taiyun/corrplot

Wickham, H. (2016). ggplot2: elegant graphics for data analysis. Available at: https://ggplot2.tidyverse.org

Wickham, H., François, R., Henry, L., and Müller, K. (2021). dplyr: a grammar of data manipulation. Available at: https://CRAN.R-project.org/package=dplyr

frontiers | Frontiers in Psychology

TYPE Brief Research Report
PUBLISHED 27 July 2023
DOI 10.3389/fpsyg.2023.1205294

Check for updates

OPEN ACCESS

EDITED BY
Melita Kovacevic,
University of Zagreb, Croatia
Maria-Jose Ezeizabarrena,
University of the Basque Country UPV/EHU,
Spain, in collaboration with editor MK

REVIEWED BY
Maja Cepanec,
University of Zagreb, Croatia
Silvia Nieva,
Complutense University of Madrid, Spain
Angela Grimminger,
University of Paderborn, Germany

*CORRESPONDENCE
Sara Feijoo
sfeijoo@ub.edu

RECEIVED 13 April 2023
ACCEPTED 14 July 2023
PUBLISHED 27 July 2023

CITATION
Feijoo S, Amadó A, Sidera F,
Aguilar-Mediavilla E and Serrat E (2023)
Language acquisition in a post-pandemic context: the impact of measures against COVID-19 on early language development.
*Front. Psychol.* 14:1205294.
doi: 10.3389/fpsyg.2023.1205294



# Language acquisition in a post-pandemic context: the impact of measures against COVID-19 on early language development

Sara Feijoo[1]*, Anna Amadó[2], Francesc Sidera[3], Eva Aguilar-Mediavilla[4] and Elisabet Serrat[3]

[1]Department of Llengües i Literatures Modernes i Estudis Anglesos, Universitat de Barcelona, Barcelona, Spain, [2]Department of Psicologia, Sociologia i Treball Social, Universitat de Lleida, Lleida, Spain, [3]Department of Psicologia, Universitat de Girona, Girona, Spain, [4]Department of Pedagogia Aplicada i Psicologia de l'Educació, Institut de Recerca i Innovació Educativa (IRIE), Universitat de les Illes Balears, Palma, Spain

Language acquisition is influenced by the quality and quantity of input that language learners receive. In particular, early language development has been said to rely on the acoustic speech stream, as well as on language-related visual information, such as the cues provided by the mouth of interlocutors. Furthermore, children's expressive language skills are also influenced by the variability of interlocutors that provided the input. The COVID-19 pandemic has offered an unprecedented opportunity to explore the way these input factors affect language development. On the one hand, the pervasive use of masks diminishes the quality of speech, while it also reduces visual cues to language. On the other hand, lockdowns and restrictions regarding social gatherings have considerably limited the amount of interlocutor variability in children's input. The present study aims at analyzing the effects of the pandemic measures against COVID-19 on early language development. To this end, 41 children born in 2019 and 2020 were compared with 41 children born before 2012 using the Catalan adaptation of the MacArthur-Bates Communicative Development Inventories (MB-CDIs). Results do not show significant differences in vocabulary between pre- and post-Covid children, although there is a tendency for children with lower vocabulary levels to be in the post-Covid group. Furthermore, a relationship was found between interlocutor variability and participants' vocabulary, indicating that those participants with fewer opportunities for socio-communicative diversity showed lower expressive vocabulary scores. These results reinforce other recent findings regarding input factors and their impact on early language learning.

KEYWORDS

**language acquisition, expressive vocabulary, parental questionnaires, pandemic, child development**

## 1. Introduction

It is a well-known fact that linguistic input is multimodal in nature, and that language learners make use of multimodal cues. While sound is the most obvious cue to speech comprehension in oral languages, since the McGurk effect was first described (McGurk and MacDonald, 1976), decades of research have shown that visual cues have an important impact in speech comprehension, and that infants use such cues from a very early age (Hollich et al., 2005; Bahrick and Lickliter, 2012; Kawase et al., 2014; Astor et al., 2021; Çetinçelik et al., 2021). For instance, young infants pay more attention to new vowel contrasts when these are presented in an audiovisual modality rather than when they are presented in an audio-only or a visual-only modality (Ter Schure et al., 2016). Also, young children show a preference for the speaker's mouth rather than other areas of the face (Tenenbaum et al., 2015; Tsang et al., 2018).

Furthermore, research also shows that access to input variability seems fundamental for young learners, in order for them to acquire the patterns and rules of the language they are exposed to (Perry et al., 2010; Rowe, 2012; Jones and Rowland, 2017; Serrat et al., 2021; Kartushina et al., 2022). In particular, talker variability has been described to facilitate linguistic development (Richtsmeier et al., 2009; Rost and McMurray, 2009; Rojas et al., 2016). Richtsmeier et al. (2009) showed that exposure to nonwords spoken by 10 different talkers resulted in faster and more accurate production among young children than exposure to the same nonwords spoken by a single talker. In a similar line, Rojas et al. (2016) found that preschoolers' expressive language skills benefit from access to input by different interlocutors, particularly interactions with older siblings and peers. Rost and McMurray (2009) also found that infants exposed to multiple speakers showed higher word discrimination rates than infants in a single-speaker condition.

The COVID-19 pandemic context and, especially, the measures against the virus adopted worldwide might have had an impact on children's language development. On the one hand, the generalised use of masks diminishes the quality of linguistic input, since masks distort the acoustic speech signal and, besides, they reduce visual cues to speech. On the other hand, restrictions regarding social gatherings as well as frequent lockdown episodes might have altered the variability of input to which children were exposed.

The data available so far is controversial (LoBue et al., 2023). On the one hand, some studies do suggest significant developmental differences between babies born during the pandemic and babies born before. For instance, Shuffrey et al. (2022) found that pandemic infants had significantly lower scores for gross motor, fine motor, and personal-social skills. In their study, none of the participant mothers or babies had been infected with the virus. Thus, the authors claim that the developmental differences found are not due to the virus itself, but to the social measures adopted against the virus, and the environment that was created as a result. Deoni (2022) also found that children born during the pandemic had significantly lower verbal, non-verbal, and cognitive performance compared to pre-pandemic children. Furthermore, Deoni's study also showed that SES, birth weight and gestation duration were protective factors, since children with lower SES, lower weight and/or shorter gestation were more affected. In a similar line, Frota et al. (2022) found that post-pandemic children exhibited lower performance at word segmentation tasks than pre-pandemic children.

On the other hand, some other studies have found no differences between pre- and post-Covid children (Wermelinger et al., 2022; Sperber et al., 2023). For instance, Mitsven et al. (2022) found that post-Covid children's language production was unaffected by mask use in the preschool classroom, and that children could benefit from the language they were exposed to despite their teachers' mask. In a similar line, Singh et al. (2021) also found no differences in children's ability to locate a target word referent when the target word was presented by a speaker wearing a mask, compared to a speaker without mask.

Given the controversy of the existing data, the present study aims at further exploring the effects that measures against COVID-19 might have had on children's lexical development. While most data available so far comes from experimental studies (Singh et al., 2021; Deoni, 2022) or classroom settings (Mitsven et al., 2022), the present study uses parental questionnaires to assess vocabulary growth among young learners in the pandemic context. The use of such an instrument will grant access to children's linguistic production in a number of different communicative situations, often only available to parents, especially during the pandemic restrictions. Furthermore, this method will easily allow comparisons between data collected during the pandemic and normative data collected before the pandemic with exactly the same instrument. In particular, the present analysis aims at answering the following research questions:

> RQ1: Will children born in the pandemic context show lower expressive vocabulary scores than children born before the pandemic?
>
> RQ2: Will measures against COVID-19 (i.e., mask use and restrictions in terms of social interaction) relate to children's expressive vocabulary scores as measured by parental questionnaires?

## 2. Method

### 2.1. Participants

The study included 82 participants (38 girls) between 8 and 30 months (M age = 19.83 months; *SD* = 5.26) from Catalan-speaking families. Participants were divided into two groups: the pre-Covid group (41 participants born before 2012) and the post-Covid group (41 participants born between 2019 and 2020).

The pre-Covid group was created by selecting a sub-sample from the normative sample of the Catalan MB-CDI (Serrat et al., 2022). To create this control group, we randomly selected those children who matched the post-Covid children in the following variables: (a) age; (b) sex; (c) prematurity status; (d) linguistic context (i.e., degree of exposure to languages other than Catalan, see Serrat et al., 2021); (e) birth order; and (f) education of mothers (see Table 1 for sociodemographic variables).

The participants in each of these two groups (pre- and post-Covid) were divided into two subgroups according to their chronological age. Thus, their families were given a different version of the instrument to complete: families of children between 8 and 18 months answered the Catalan adaptation of the McArthur-Bates CDI inventory 1 (i.e., words and gestures, MB-CDI: WG), and families of children between 16 and 30 months answered the inventory 2 (i.e., words and sentences, MB-CDI: WS). In this way, a total of 32 participants (16 pre- and 16 post-Covid) completed the MB-CDI: WG (M age = 14.39 months; *SD* = 2.120) and 50 participants (25 pre- and 25 post-Covid) completed the MB-CDI: WS (M age = 23.2 months;

TABLE 1 Sociodemographic variables for pre- and post-Covid groups.

| Variables | Pre-Covid | Post-Covid | Differences between groups |
|---|---|---|---|
| | *M* (*SD*) | *M* (*SD*) | *U* Mann Whitney or $\chi^2$ |
| *N* | 41 | 41 | |
| Age (months) | 19.83 (5.26) | 19.83 (5.26) | $U = 840$, $p = 1.0$ |
| Gender | | | |
| Male | 22 | 22 | $\chi^2 = 0.000$, $p = 1.00$ |
| Female | 19 | 19 | |
| Prematurity | | | |
| Yes | 0 | 0 | - |
| No | 41 | 41 | |
| Weigh at birth | 3.32 (0.4) | 3.27 (0.48) | $U = 606$, $p = 0.676$ |
| Birth order | | | |
| 1st | 32 | 31 | |
| 2nd | 9 | 9 | $\chi^2 = 1.016$, $p = 0.602$ |
| 3rd | 0 | 1 | |
| Education of mothers[a] | | | |
| Primary | 0 | 0 | |
| Secondary | 5 | 4 | $\chi^2 = 0.188$, $p = 0.655$ |
| University | 34 | 37 | |
| Bilingualism | | | |
| No | 18 | 17 | $\chi^2 = 0.345$, $p = 0.842$ |
| Familiar bilingualism (only one Catalan-speaking parent) | 6 | 8 | |
| Other contacts | 17 | 16 | |
| Otitis[a] | | | |
| Yes | 8 | 9 | $\chi^2 = 0.046$, $p = 0.829$ |
| No | 32 | 32 | |
| Previous language difficulties | | | |
| Yes | 1 | 0 | $\chi^2 = 1.012$, $p = 0.314$ |
| No | 40 | 41 | |

[a]Lost data.

*SD* = 3.452). Regarding the SES of the families, measured on the basis of maternal education, 6.1% of the sample completed secondary studies, 3.7% post-secondary studies, and 90.2% university studies.

## 2.2. Instruments

The data for this study was obtained using the MacArthur-Bates Communicative Development Inventories (MB-CDIs) adapted to Catalan (Serrat et al., 2022). Of all the sections included in this instrument, the vocabulary section was considered, since it is common to both questionnaires. Parents had to indicate their child's ability to understand or say a series of words in the case of the MB-CDI: WG, or just the ability to say the words in the case of the MB-CDI: WS. The MB-CDI: WG lists 423 words which are grouped into 19 categories, while the MB-CDI: WS lists 678 words which are grouped into 22 categories.

For the gathering of the child's personal and socio-demographic data, the last part of the questionnaire was used. For the specific purpose of the present study, in order to analyse the impact of measures against COVID-19 on vocabulary development, the following two questions were added:

- How would you define your child's variety of sociocommunicative interaction in the last 3 months?

  (1) Very little (2) Little (3) Average (4) Quite a lot (5) A lot.
- How often has your child been in contact with interlocutors wearing a mask since the beginning of COVID?

  (1) Never (2) Hardly ever (3) Sometimes (4) Often (5) Always.

## 2.3. Procedure

In order to administer the questionnaires to the families of the post-Covid group, 3 early childhood educational centers in Catalonia (Spain) were contacted during June and July 2021. Several waves of lockdowns and restrictions of different types (e.g., mobility, social gatherings…) had occurred in this area since the beginning of the pandemic. All educational centers were closed from March to September of 2020. When they reopened, mask use was compulsory for teachers at all educational levels until April 2022. In addition, lockdown episodes occurred whenever positive cases emerged within a group and, consequently, all students were sent home for quarantine.

We contacted the directing teams of the educational centers through email. We described the objectives of the study and the characteristics of the target sample and asked them to give a document with information about the study and a consent form to the families. Parents who gave written consent received the questionnaires in written format and were asked to return them within a week. Most children from the post-Covid group attended educational centers (73.1%). The rest were personal and professional contacts of the authors. All questionnaires were filled by either the child's mother or by both parents.

The total vocabulary scores for each child were used to calculate their percentile of expressive vocabulary according to the normative scores of the test, and this percentile of vocabulary was taken as a dependent variable. Data were analyzed with SPSS version 25. The nonparametric Mann–Whitney U test was used to compare independent groups, as dependent variables (Total expressive vocabulary and Percentile of expressive vocabulary) did not show a normal distribution (Shapiro–Wilk $W = 0.796$, $p < 0.001$ and $W = 0.946$, $p = 0.002$, respectively). A Chi-square approach was used to compare the number of participants classified as "high vocabulary level" (percentile equal to or over 75), "typical development" (percentile between 26 and 74), and "low vocabulary level" (percentile equal to or lower than 25). Finally, two linear regression analyses, one introducing the "diversity of communicative interaction" and the other one "face mask use" were performed over the dependent variable "total expressive vocabulary" in the sample of post-Covid children.

## 3. Results

Results show equal exposure to face masks between children studied with MB-CDI: WG and MB-CDI: WS ($U = 779.5$, $p = 0.833$), and between children born in 2019 and 2020 ($U = 195$, $p = 0.803$). However, older children received more diverse social interactions than younger children ($U_{2019\text{-}2020} = 118$, $p = 0.017$; $U_{CDI.WG\text{-}CDI.WS} = 115$, $p = 0.023$).

Although the means of expressive vocabulary were lower in the post-Covid group than in the pre-Covid group, results did not show significant differences between both groups, neither in the total expressive vocabulary ($U = 796$, $p = 0.680$), nor in the percentile of expressive vocabulary ($U = 712.5$, $p = 0.234$) (see Table 2).

Also, the distribution of children in three groups (i.e., low vocabulary level, typical development, and high vocabulary level) through standardized data (percentile) of their total expressive vocabulary did not show significant differences (see Table 3). Nevertheless, the distribution approximates significance when only two groups were considered ($\chi^2 = 2.53$, $p = 0.099$) showing more children with high vocabulary level in the pre-Covid group and more children with low vocabulary level in the post-Covid group.

In order to know which variables affected the total expressive vocabulary of the children in the post-Covid group, we performed two regression analyses, one introducing "diversity of communicative interaction" and the other introducing "face mask use" as independent variables over the dependent variable "total expressive vocabulary."

As can be seen in Table 4, face mask use cannot explain differences in the total expressive vocabulary. Nevertheless, diversity of communicative interaction explains 16.6% of the variability of the total expressive vocabulary.

## 4. Discussion

The present study employed parental questionnaires to assess children's vocabulary development during the pandemic context. The instrument provided detailed and accurate data of children's vocabulary knowledge, which allowed for direct comparison with similar data from pre-pandemic children.

In terms of this comparison, as expressed in our first research question, the analysis found no differences between children in the pre-Covid group, and the post-Covid group as far as expressive vocabulary is concerned. Although there seems to be a tendency to a distribution of children with higher scores in the pre-Covid group and lower scores in the post-Covid group, these differences did not reach statistical significance. Thus, the results obtained from parents' questionnaires in a domestic context are similar to the results found in a preschool classroom context by Mitsven et al. (2022), who also found that pandemic children benefit from teacher's input as well as pre-pandemic children. Therefore, as suggested by LoBue et al. (2023), measures against COVID-19 might have had an impact on caregivers' socioemocional behaviour, but the measures seem to have had little or no effect on infants' development. Alternatively, it might be the case that post-pandemic children are not developing worse than pre-pandemic children, but simply differently, and that those differences are reflected in other areas of language development (Frota et al., 2022; Shuffrey et al., 2022). It should also be born in mind that additional factors such as SES have been described as protective factors in the pandemic (Deoni, 2022), since children with higher SES outperform children with lower SES. Given the fact that most of the children in our sample belong to a high SES, this might have weakened the differences between our post-pandemic group and our pre-pandemic group. Unfortunately, the limitations of the present study and the size of the present sample did not allow for proper comparisons of groups with different SES. Our sample size did not allow for within-group comparisons considering other factors either, nor did it provide widely generalizable data. However, these results might be an important contribution to the field, given the uniqueness of the circumstances in which these data were collected.

Regarding our second research question, the present analysis found that interlocutor variability was related to children's expressive vocabulary, which indicates that those participants with more frequent lockdown episodes and less opportunities for socio-communicative diversity showed lower expressive vocabulary scores. As the literature suggests, exposure to vocabulary items

TABLE 2 Mean, standard deviation and independent group comparison between the pre- and post-Covid groups.

| Variables | Pre-Covid | | Post-Covid | | Differences between groups |
|---|---|---|---|---|---|
| | *M* (*SD*) | Range | *M* (*SD*) | Range | *U* Mann Whitney or $\chi^2$ |
| Total expressive vocabulary | 155.5 (182.7) | 0–571 | 140.4 (174.1) | 0–653 | $U = 796$, $p = 0.680$ |
| Percentile expressive vocabulary | 51.8 (29.3) | 5–99 | 43.8 (27.3) | 5–95 | $U = 712.5$, $p = 0.234$ |

TABLE 3 Distribution of children regarding expressive vocabulary performance in the pre- and post-Covid groups.

| Variables | Pre-Covid | Post-Covid | Differences between groups |
|---|---|---|---|
| | *n* | *n* | $\chi^2$ |
| Total participants | 41 | 41 | |
| Low vocabulary level (percentile <25) | 11 | 15 | $\chi^2 = 2.721$, $p = 0.257$ |
| Typical vocabulary development (percentile 26–74) | 18 | 20 | |
| High vocabulary level (percentile >75) | 12 | 6 | |

*n*, number of children.

TABLE 4 Regression analyses for measures against COVID-19 over total expressive vocabulary.

| | Total expressive vocabulary | | | |
|---|---|---|---|---|
| Predictor | $\beta$ | $\Delta R^2$ | F or t | p |
| Model 1 | | 0.166 | 8.946 | 0.005 |
| Diversity of communicative interaction | 0.432 | | 2.99 | 0.005 |
| Model 2 | | −0.011 | 0.118 | 0.732 |
| Face mask use | −0.038 | | −0.344 | 0.732 |

spoken by different talkers results in faster and more accurate development than exposure to the same items spoken by a single talker (Richtsmeier et al., 2009). In the same line, Rojas et al. (2016) or Serrat et al. (2021) also found that preschoolers with access to input by different interlocutors show higher rates of expressive language skills, due to the wider range of topics, referents and vocabulary items that children are exposed to and, subsequently, acquire. However, an important limitation regarding the present study lies in the way sociocommunicative diversity was measured, given the fact that the question that was addressed to parents in the questionnaire might have been interpreted differently by different participants. Given the importance of sociocommunicative diversity and the relationship it seems to have with linguistic development, further research should explore this relationship with a more accurate and objective operationalization of the variable related to communicative diversity.

Regarding the use of mask, there seems to be no relationship between this measure against COVID-19 and children's vocabulary development. Therefore, the present results are in line with those obtained by Singh et al. (2021), who also found no differences in terms of word-object identification between speech with mask or without mask. As some researchers have claimed (Pycha et al., 2022; Wermelinger et al., 2022), speakers might modify their language production in the presence of a physical barrier, namely a face mask, in order to make their speech more intelligible. In fact, previous findings have already shown that speakers tend to increase their speech quality (i.e., speech rate, pitch, length of words, etc.) while wearing a mask, in order to compensate for difficult communicative situations (Crimon et al., 2022). Additionally, non-verbal cues such as co-speech gestures, beats or iconic gestures have also been said to compensate for speech degradation in a number of contexts (Drijvers and Özyürek, 2017; Crimon et al., 2022). Therefore, it is possible that the use of alternative communication strategies might have neutralized the possible negative effects that masked speech might have produced otherwise.

## 5. Conclusion

The present study made use of parental questionnaires in order to assess expressive vocabulary development among children born within the COVID-19 pandemic context. As an instrument, the parental questionnaire provided valuable evidence of linguistic development from a sample in a context that was otherwise very difficult to obtain. At the same time, it allowed for direct comparisons with normative data obtained from children born before the pandemic. The main findings of such comparison revealed no significant differences between pre- and post-pandemic children in terms of expressive vocabulary. Nevertheless, further analyses within the post-pandemic group indicated that, despite mask-use had no effect on vocabulary development during the pandemic, restrictions on social gatherings did, given that lower interlocutor variability scores among post-pandemic children correlated with lower expressive vocabulary scores. Given the results obtained, future studies should further explore the relationship between interlocutor variability and early language development in order to confirm this finding.

## Data availability statement

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

## Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements at the moment of the first data collection. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

## Author contributions

SF: data collection, analysis and writing, and edition of the final paper. AA and FS: data collection, analysis, and writing of the final paper. EA-M and ES: data collection, analysis and writing of the final paper. All authors contributed to the article and approved the submitted version.

## Funding

This work was supported by grant 2021 SGR 00303 from the Department of Research and Universities of the Government of Catalonia.

## Conflict of interest

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

## Publisher's note

# References

Astor, K., Thiele, M., and Gredebäck, G. (2021). Gaze following emergence relies on both perceptual cues and social awareness. *Cogn. Dev.* 60:101121. doi: 10.1016/j.cogdev.2021.101121

Bahrick, L. E., and Lickliter, R. (2012). "The role of intersensory redundancy in early perceptual, cognitive, and social development" in *Multisensory development*. eds. A. Bremner, D. J. Lewkowicz and C. Spence (Oxford: Oxford University Press), 183–205.

Çetinçelik, M., Rowland, C. F., and Snijders, T. M. (2021). Do the eyes have it? A systematic review on the role of eye gaze in infant language development. *Front. Psychol.* 11:589096. doi: 10.3389/fpsyg.2020.589096

Crimon, C., Barbir, M., Hagihara, H., de Araujo, E., Nozawa, S., Shinya, Y., et al. (2022). Mask wearing in Japanese and French nursery schools: the perceived impact of masks on communication. *Front. Psychol.* 13:874264. doi: 10.3389/fpsyg.2022.874264

Deoni, S. (2022). Impact of the COVID-19 pandemic environment on early child brain and cognitive development. *Biol. Psychiatry* 91:S26. doi: 10.1016/j.biopsych.2022.02.082

Drijvers, L., and Özyürek, A. (2017). Visual context enhanced: the joint contribution of iconic gestures and visible speech to degraded speech comprehension. *J. Speech Lang. Hear. Res.* 60, 212–222. doi: 10.1044/2016_JSLHRH-16-0101

Frota, S., Pejovic, J., Cruz, M., Severino, C., and Vigário, M. (2022). Early word segmentation behind the mask. *Front. Psychol.* 13:879123. doi: 10.3389/fpsyg.2022.879123

Hollich, G., Newman, R. S., and Jusczyk, P. W. (2005). Infants' use of synchronized visual information to separate streams of speech. *Child Dev.* 76, 598–613. doi: 10.1111/j.1467-8624.2005.00866.x

Jones, G., and Rowland, C. (2017). Diversity not quantity in caregiver speech: using computational modeling to isolate the effects of the quantity and the diversity of the input on vocabulary growth. *Cogn. Psychol.* 98, 1–21. doi: 10.1016/j.cogpsych.2017.07.002

Kartushina, N., Rosslund, A., and Mayor, J. (2022). Toddlers raised in multi-dialectical families learn words better in accented speech than those raised in monodialectal families. *J. Child Lang.* 49, 1093–1118. doi: 10.1017/S0305000921000520

Kawase, S., Hannah, B., and Wang, Y. (2014). The influence of visual speech information on the intelligibility of English consonants produced by non-native speakers. *J. Acoust. Soc. Am.* 136, 1352–1362. doi: 10.1121/1.4892770

LoBue, V., Pérez-Edgar, K., Kirkham, N., and Herbert, J. (2023). The impact of COVID-19 on infant development: a special issue of infancy. *Infancy* 28, 4–7. doi: 10.1111/infa.12528

McGurk, H., and MacDonald, J. (1976). Hearing lips and seeing voices. *Nature* 264, 746–748. doi: 10.1038/264746a0

Mitsven, S. G., Perry, L. K., Jerry, C. M., and Messinger, D. S. (2022). Classroom language during COVID-19: associations between mask-wearing and objectively measured teacher and preschooler vocalizations. *Front. Psychol.* 13:874293. doi: 10.3389/fpsyg.2022.874293

Perry, L. K., Samuelson, L. K., Malloy, L. M., and Schiffer, R. N. (2010). Learn locally, think globally: exemplar variability supports higher-order generalization and word learning. *Psychol. Sci.* 21, 1894–1902. doi: 10.1177/0956797610389189

Pycha, A., Cohn, M., and Zellou, G. (2022). Face-masked speech intelligibility: the influence of speaking style, visual information, and background noise. *Front. Commun.* 7:874215. doi: 10.3389/fcomm.2022.874215

Richtsmeier, P. T., Gerken, L., Goffman, L., and Hogan, T. (2009). Statistical frequency in perception affects children's lexical production. *Cognition* 111, 372–377. doi: 10.1016/j.cognition.2009.02.009

Rojas, R., Iglesias, A., Bunta, F., Goldstein, B., Goldenberg, C., and Reese, L. (2016). Interlocutor differential effects on the expressive language skills of Spanish-speaking English learners. *Int. J. Speech Lang. Pathol.* 18, 166–177. doi: 10.3109/17549507.2015.1081290

Rost, G. C., and McMurray, B. (2009). Speaker variability augments phonological processing in early word learning. *Dev. Sci.* 12, 339–349. doi: 10.1111/j.1467-7687.2008.00786.x

Rowe, M. L. (2012). A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child Dev.* 83, 1762–1774. doi: 10.1111/j.1467-8624.2012.01805.x

Serrat, E., Aguilar-Mediavilla, E., Sanz-Torrent, M., Andreu, L., Amadó, A., Badia, I., et al. (2022). Inventaris del desenvolupament d'habilitats comunicatives MacArthur-Bates en català. Guia d'ús i manual tècnic. Editorial UOC.

Serrat, E., Amadó, A., Bonet, A., Feijóo, S., and Aguilar-Mediavilla, E. (2021). Growing up in a socially bilingual environment: simultaneous acquisition of a typologically close language (Crecer en un entorno social bilingüe: adquisición de una lengua de forma simultánea a otra lengua tipológicamente similar). *J. Stud. Educ. Dev.* 44, 336–369. doi: 10.1080/02103702.2021.1888490

Shuffrey, L. C., Firestein, M. R., Kyle, M. H., Fields, A., Alcántara, C., Amso, D., et al. (2022). Association of birth during the COVID-19 pandemic with neurodevelopmental status at 6 months in infants with and without in utero exposure to maternal SARS-CoV-2 infection. *JAMA Pediatr.* 176:e215563. doi: 10.1001/jamapediatrics.2021.5563

Singh, L., Tan, A., and Quinn, P. C. (2021). Infants recognize words spoken through opaque masks but not through clear masks. *Dev. Sci.* 24:e13117. doi: 10.1111/desc.13117

Sperber, J. F., Hart, E. R., Troller-Renfree, S. V., Watts, T. W., and Noble, K. G. (2023). The effect of the COVID-19 pandemic on infant development and maternal mental health in the first two years of life. *Infancy* 28, 107–135. doi: 10.1111/infa.12511

Tenenbaum, E. J., Sobel, D. M., Sheinkopf, S. J., Malle, B. F., and Morgan, J. L. (2015). Attention to the mouth and gaze following in infancy predict language development. *J. Child Lang.* 42, 1173–1190. doi: 10.1017/S0305000914000725

Ter Schure, S., Junge, C., and Boersma, P. (2016). Discriminating non-native vowels on the basis of multimodal, auditory or visual information: effects on infants' looking patterns and discrimination. *Front. Psychol.* 7, 1–11. doi: 10.3389/fpsyg.2016.00525

Tsang, T., Atagi, N., and Johnson, S. P. (2018). Selective attention to the mouth is associated with expressive language skills in monolingual and bilingual infants. *J. Exp. Child Psychol.* 169, 93–109. doi: 10.1016/j.jecp.2018.01.002

Wermelinger, S., Moersdorf, L., and Daum, M. M. (2022). How experience shapes infants' communicative behaviour: comparing gaze following in infants with and without pandemic experience. *Infancy* 27, 937–962. doi: 10.1111/infa.12488

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 28 July 2023
DOI 10.3389/fpsyg.2023.1179999

Check for updates

OPEN ACCESS

EDITED BY
Melita Kovacevic,
University of Zagreb, Croatia

REVIEWED BY
Virginia A. Marchman,
Stanford University, United States
Natalia Gagarina,
Leibniz Center for General Linguistics (ZAS), Germany

*CORRESPONDENCE
Tiia Tulviste
✉ Tiia.Tulviste@ut.ee

RECEIVED 05 March 2023
ACCEPTED 14 July 2023
PUBLISHED 28 July 2023

CITATION
Tulviste T and Schults A (2023) How congruent are parent reports on 3–4-year-old children's language skills with other sources of data? *Front. Psychol.* 14:1179999.
doi: 10.3389/fpsyg.2023.1179999



# How congruent are parent reports on 3–4-year-old children's language skills with other sources of data?

Tiia Tulviste* and Astra Schults

University of Tartu, Tartu, Estonia

**Background:** Parental report measures such as the MacArthur-Bates Communicative Development Inventories (CDIs) are frequently used to study communicative skills of children under 3 years of age. Less is known about the usability of such reports for assessing communication skills in older children due to their advanced language skills, and a higher variety of communicative partners and communication contexts.

**Aims:** To assess the concurrent and predictive validity of the Estonian (E) CDI-III at ages 3;0 and 4;0 years. The first research goal was to examine its concurrent variability—associations with teacher reports and directly measured language skills. The second goal of the study was to investigate the predictive validity of parent reports—the degree to which parent-and teacher-reported language scores for children at age 3;0 are useful for predicting examiner-administered language comprehension and production scores 1 year later.

**Methods:** Estonian monolingual children were investigated longitudinally at ages 3;0 ($n = 104$; M age = 35.77 months, $SD = 0.84$; 42% males) and 4;0 ($n = 87$; M age = 48.18 months, $SD = 1.16$; 42% males) years. Children were assessed with the parent-reported ECDI-III, with teacher-reported assessments on children's talkativeness, vocabulary size and grammatical skills, and the examiner-administered New Reynell Developmental Language Scales IV (NRDLS).

**Results:** Results indicated significant positive relationships between the ECDI-III total scores, teacher reports, and directly measured language comprehension and production scores, demonstrating concurrent validity of parental reports of children language skills at both ages. When controlling for mothers' education, children's gender, and reported language difficulties, parental and teacher reports were predictive of language production scores, whereas only parental reports predicted comprehension scores 1 year later. None of the controls was predictive of later language comprehension and production scores.

**Conclusion:** In sum, good concurrent and predictive validity of the ECDI-III shows that the instrument is a valid tool for assessing communicative skills in Estonian children. Results suggest that parent reports can offer useable information also about communicative skills of children older than three years.

KEYWORDS

CDI, parental reports, teacher report, language development, communicative development, parental education, Estonian

## 1. Introduction

Parental report measures such as the MacArthur-Bates Communicative Development Inventories (MB-CDIs) are widely used instruments for estimating language skills of infants and toddlers (Fenson et al., 2007). Parent reports are time-and cost-effective in obtaining a picture of child early development and allowing to gather data on large samples. Unlike direct testing, parent report instruments do not require a well-trained estimator, and children do not need to communicate with an unfamiliar adult and to solve tasks that may be decontextualized and novel for them (Fenson et al., 2007). Parents are good reporters on children's language skills likely due to the possibility of observing children communicating in various situations and knowing what children are able to say. Their reports are not influenced by the child's current mood, health, attention state or temperament (e.g., shyness) like direct assessment. Being their children's first teachers, parents stimulate their development, and CDIs could serve as a tool for monitoring children's language learning. Because formal testing is difficult to conduct in small children, parent reports are especially suitable for estimation of communicative abilities below 3 years of age (Fenson et al., 2007).

A number of studies show the utility, validity, and reliability of parent reports on infants' and toddlers' language skills (see Fenson et al., 2007; Law and Roy, 2008 for reviews). Significant correlations have been found between parental reports, concurrent spontaneous speech measures and direct assessments of child language skills (Pan et al., 2004). Recent studies show strong predictive validity of parents reports on children's early language skills. For example, Bleses et al. (2016) indicated that early expressive vocabulary predicts reading and math outcomes 10 years later. Less is known about parental reports as a source of information about older children's language skills. Unlike younger children, they have better communicative skills.

There is a growing body of studies addressing the utility and validity of parent reports for assessing communication skills in children over 3 years of age (Dionne et al., 2003; Eriksson, 2017). Several adaptations of the CDI-III have been developed based on the original version of the instrument (Dionne et al., 2003), for example for Basque, Norwegian, and Spanish (see Kas et al., 2022). Studies using the CDI-III original version reflect to the ability of parents to provide valid estimation of children's language skills also at ages 30–37 month of age (Fenson et al., 2007). Validation studies have found ceiling effects in "Syntactic complexity" and "Uses of language" subscales of the original version after 33 months (Fenson et al., 2007) and in all subscales of the Basque version after 42 months (see Kas et al., 2022). Other adaptations have been used the Swedish version of the CDI-III (Eriksson, 2017) designed for children from 30 months to 48 months, as for example Estonian (Tulviste and Schults, 2020), Hungarian (Kas et al., 2022) and Portuguese CDI-IIIs (Cadime et al., 2021). The Swedish version covers a longer period of time than the original version of the CDI-III. Eriksson (2017) found a slight ceiling effect in the syntactic complexity and metalinguistic awareness subscales after 45 months.

All versions of the CDI-IIIs are relatively new and only a few studies have focused on its concurrent validity. Odeskog and Stenberg reported low correlations between the Swedish CDI-III and the Peabody Picture Vocabulary Test and the Boston Naming Test in 44 children aged 36–47 months (see Kas et al., 2022). Tulviste and Schults (2020) found medium correlations between the vocabulary scores of the ECDI-III and directly measured Reynell Language Comprehension and Production Scale scores in 100 children at the age of 3 years. Cadime et al. (2021) showed moderate correlations with the Vocabulary total score and the Syntax score of the European Portuguese CDI-III and the language score of the Griffiths Mental Development Scales in 23 children aged 30–48 months. Studies using standardized tests to investigate the concurrent validity of CDI-IIIs suffer from limited age range and small sample size. There is a lack of studies on the predictive validity of CDI-IIIs. Thus, more studies are needed to address the concurrent and predictive validity of parent report in children aged 3;0 years and older.

Given that nowadays most children beyond 3;0 years of age are enrolled in kindergarten and spend long days in child-care settings, kindergarten teachers also play an important role in facilitating child development. Teachers are expected to monitor child language acquisition and identify children with speech and communication problems, because early intervention is more efficient than later intervention (e.g., Dale et al., 2003). Accordingly, some researchers have started to use kindergarten teachers as a source of information about children's language skills via CDIs (Vagh et al., 2009; Bleses et al., 2018; Cadime et al., 2021). Teachers are seen as good judges of child language abilities, working frequently with groups of same-age children. That provides them plenty of opportunities to compare communicative abilities of children in similar age. Kindergarten teachers also have the opportunity to observe children interacting with different communicative partners in different interactional contexts, despite the range of contexts being limited. Moreover, during teacher training, they have studied child development milestones, including their communicative development, and how to stimulate child development.

Some authors suggest to use multiple reporters for estimating children's language skills, considering the possibility that children may talk about somewhat different topics with different conversational partners in and outside home, and in case of bilingual children, also involve different languages. Many parents of bilingual children may not be able to report children's non-native language abilities (Vagh et al., 2009). However, De Houwer et al. (2005) found in a study with monolingual children that although there are significant correlations among estimations about children's language skills done by different reporters (mothers, fathers and the third person) via CDI, they assess language skills of the same child rather differently, especially in case of older children whose language skills are relatively high.

Thus, there are some concerns about the use of reports with older children due to their increased communicative abilities as well as a higher number of conversational partners and interactional contexts. Therefore, studies providing more information about the utility of using reports as a source of information about children's language skills after 3 years, are needed. The current study assessed the concurrent and predictive validity of the ECDI-III – the Estonian adaption of the Swedish CDI III (Eriksson, 2017), using the data gathered at two timepoints: at children's age of 3;0 and 4;0 years.

The first aim of the present study was to test the concurrent validity of parent reports on general communicative skills (the total score of the ECDI-III) as compared with teacher reports, and directly assessed 3-and 4-year-old children's language comprehension and production scores by a standardized examiner-administered language assessment New Reynell Developmental Language Scales (NRDLS) (Edwards et al., 2011). Although CDIs have been used to assess various aspects of child early language, most previous studies have focused on infants' and toddlers' vocabulary skills and checked the validity and reliability of the vocabulary list (Pan et al., 2004; Fenson et al., 2007; Tulviste and Schults, 2020). Some other studies have explored both vocabulary and grammatical development, since multiword sentences, basic sentence structure and inflections of the native language are also good indicators of the rate of language development (Fenson et al., 2007). The children participating in our study were at ages 3;0 and 4;0 years when grammatical and phonological skills are also indicative about the level of their language skills. Therefore, in addition to vocabulary scores, we also used scores from other subscales, and calculated total ECDI-III scores to serve as an indicator of more general language skills of the child. Another reason for using total scores instead of only vocabulary was that the NRDLS assesses general language comprehension and production skills. As teacher reports we used compound teacher ratings of their answers to three questions about children's communicative abilities: teachers' evaluations of child's talkativeness, size of vocabulary, and complexity of sentences compared to the child's age mates.

The second aim of the study was to investigate to what extent parent-and teacher reported language skills have substantial predictive validity, evaluating the utility of both sources of reports around the time of their third birthday for predicting language skills around their 4th birthday. To explore how well parent and teacher reports predict future language skills, we also considered parental education and child gender, established as important predictors of child language development (Fenson et al., 2007). Moreover, education might also affect how adequate the reports are. As pointed out by Stiles (1994), the CDIs place high demands on parents to reflect on different aspects of child communication. It is likely that parents with higher educational level manage better in filling out the questionnaire as they have better knowledge about what children are able to say (Fenson et al., 2007). Plenty of studies mostly based on parental reports have found gender differences in children's communicative skills. Girls have demonstrated to have larger vocabularies and quicker rates of grammatical development than boys (Bornstein et al., 2004; Eriksson et al., 2012; Simonsen et al., 2014; Urm and Tulviste, 2016).

Thus, the study addressed the following questions:

1. Are parent reports at ages 3;0 and 4;0 valid estimators of Estonian children's language skills when compared with teacher reports and experimenter-measured language skills (language comprehension and production via the NRDLS)?
2. To what extent do earlier parent and teacher reports predict children's language skills 1 year later, when controlling for mother's education and child's gender?

# 2. Methods

## 2.1. Participants

As part of a larger research project, „The role of early social contexts in supporting the development of language skills: A way to close the academic achievement gap", led by the first author of the current paper, a longitudinal study to validate ECDI-III was carried out. The first gathering of data was around children's third birthday for 104 children (44 boys, 60 girls, age range from 2;10 to 3;3, $M = 35.77$ months, $SD = 0.84$). The second gathering of data was around children's fourth birthday ($M = 48.18$, $SD = 1.16$, age range from 3;10 to 4;2) for 87 of the original participants. At first gathering of data 20 children (12 boys and 8 girls) were identified by their parents as experiencing difficulties with language development, 17 of them (9 boys and 8 girls) participated also at the second gathering of data. According to the parents, the children were otherwise healthy. According to parental reports, Estonian was the dominant language in the families, although 12 children had a parent or grandparents who sometimes (less often than daily, for a couple of hours at a time) spoke another language with the child. None of the participants were excluded due to reported difficulties with language development nor due to exposure to another language. Most participants were from middle or higher SES homes with mothers having completed upper secondary (33%) or university (54%) education, 7% of the participants had parents with a lower secondary education, and parental education data were not available for 6% of children. The income for the family was more than one average wage for 82% of the participants. One hundred and one of the participating children were attending kindergarten or a playgroup regularly.

## 2.2. Procedure

Participants were recruited through kindergartens and child care centers. We sent an invitation to participate to all kindergartens and child care centers in the cities of Tartu and Pärnu, Estonia, where there were groups for 3-year-olds. If the head of the institution agreed to take part in the study, they asked the teachers of three-year-olds to hand out the invitations to the families. An invitation to participate was sent shortly before the child's third birthday to 207 families. Roughly half of the invited families agreed to participate, signing the informed consent form. The teachers who had children from their group participating in the study were asked to fill in Social Skills Questionnaires (Häidkind et al., 2018) on paper. As there were two teachers per group, they decided themselves which one of them would fill in the Social Skills Questionnaire for each participating child. Most of the teachers filled in one or two questionnaires, maximum number of Social Skills Questionnaires filled in by one teacher was four. Trained research assistants visited families at home on two occasions (around child's third and fourth birthday) and administered the Estonian version of NRDLS (Edwards et al., 2011), first the Comprehension Scale and then the Production Scale. If a child did not comply to take both scales of the NRDLS during one visit (e.g., being fussy, tired), the assistants visited the family again. Five children did not comply the Production Scale during the second visit either, and NRDLS was left uncompleted. At both visits, the assistants asked the parents to complete the questionnaires (subject information sheet, ECDI-III, and Social Skills Questionnaire) within the next couple of days. The parents could choose if they preferred to fill in the questionnaires online or on paper. A paper version of the questionnaires was handed out to them with a prepaid return envelope. Seventy eight of the parents completed the questionnaires online and 26 on paper. We sent gentle reminders about the questionnaires waiting to be completed to those families who had agreed to participate in the study but who had not completed the questionnaires in 2 weeks after having received either the link to the questionnaire or the questionnaire on paper. Still, five parents completed only the vocabulary section of the ECDI-III. Written feedback on the child's language results was sent to the parents. Day-care teachers who provided reports on children's communicative skills were provided gift cards for their help, as were families who participated at both times of data gathering. Children received stickers as presents.

The Research Ethics Committee of the University of Tartu approved the study. The CDI Advisory Board approved the development of adaptations of CDI-III to Estonian, based on the work already authorized and done for Swedish.

## 2.3. Materials

### 2.3.1. ECDI-III

The ECDI-III (Tulviste and Schults, 2020) is the Estonian adaptation of the CDI-III developed for Swedish by Eriksson (2017), designed for children 30 to 48 months old and consists of 6 subscales. First, in the level of communication section parents have to indicate if their child can speak and how complex their child's speech is (6 alternative items). The parents are asked to continue with filling in the rest of the checklist only if they have marked an alternative indicating that their child uses at least one-word utterances.

Second, in a 100 item vocabulary list the parents have to indicate words (from the list of 100 words, mainly verbs and adjectives) that their child produces in four themes: food words (16 items), body words (26 items), mental words (30 items), and emotion words (28 items).

Third, in the syntax section the parents are asked about their child's grammar usage and sentence complexity. Grammar usage lists 7 items including the plural, comparisons, past tense, and conjunctions. The parents are asked to indicate for each item if their child has never used a particular example of grammar (scored 0), has used it several times (scored 1), or uses it on a daily basis (scored 2). Thus, the possible score for grammar usage ranged from 0 to 14. Sentence complexity consists of 10 pairs of sentences that consists of a short sentence with simple grammar and a complex, more elaborated sentence, both expressing the same main meaning. Regarding the pairs of simple and complex sentences the parents had to indicate for each pair if their child currently uses the simpler one (scored 0), alternates between simple and complex sentences (scored 1), or currently uses the more complex one (scored 2). The maximum score of sentence complexity is 20. The maximum score for syntax section is 34.

Fourth, in the metalinguistic awareness section the parents assess phonological awareness and orthographic awareness of the children. For phonological awareness (3 items), the parents have to indicate whether their child is able (scored 1) or unable (scored 0) to notice rhymes, to break words into syllables, and to understand that some people speak a foreign language. For orthographic awareness (4 items), the parents have to indicate whether their child is engaged in activities related to letters (scored 1 or 0 respectively) such as being interested in letters, recognizing some letters, writing some letters, and writing some short familiar words. The maximum score for metalinguistic awareness is seven.

Fifth, in the pronunciation section the parents are asked how their child's speech sounds compared to other children of the same age, and if their child has pronunciation difficulties. For five of the listed items parents are asked to indicate if their child has difficulties (scored 0) or not (scored 1) with the pronunciation of more difficult phonemes (*r*-sound and *s*-sound), changing the form of words as they are produced, and if strangers are able to understand the child. The final item asked if the child's speech resembles that of a younger child (scored 0), an age mate (scored 1), or an older child (scored 2). The maximum score for the pronunciation section is seven. All subscale scores were summed (max = 154).

### 2.3.2. The New Reynell developmental language scales

Children's language comprehension and production skills were tested using NRDLS (Edwards et al., 2011). This is the most recent version of the well-known structured tests—the Reynell Developmental Language Scales. The scales test vocabulary and grammar: the comprehension and production of single words (nouns and verbs) as well as of simple and complex sentences with easiest items at the beginning and most difficult in the end. Objects, pictures and variety of testing procedures are used to maintain the attention of children. First the comprehension tasks and then the production tasks were administered to each child individually by a research assistant during home visits. The Comprehension Scale consists of 72 items and the Production Scale of 64 items. An adapted version for Estonian children has the same number of items in both scales, but wording of some items in the pronouns, complex sentences, and grammatical judgment sections have been changed because of language differences between Estonian and English. Estonian is an agglutinative language, characterized by a large number of cases (14 productive cases), no grammatical gender (either of nouns or personal pronouns), and no articles. In the Estonian pronouns section, *ennast* "himself/herself" and *teda* "him/her" have been used. The complex sentences section assesses the child's comprehension of passive sentences, and the thematic roles expressed by the passive sentences are reversable. The child is expected to show the picture that goes with what is said, e.g., to show the picture of a baby being fed by the mother after the experimenter said "The mother is fed by the baby." Because these passive sentences from the original English versions were not translatable into Estonian, sentences in the active voice (e.g., "Tita annab emale süüa") are used and the child has to work out who is doing what to whom. The study has preliminary norms only for 3–4-year-old children based on 255 children in the age range from 34 to 50 months (Tulviste, unpublished data). In the present study, Cronbach's alpha was used to access the internal consistency of items within the scales. These were 0.93 for the Comprehension scale and 0.96 for the Production scale. At both ages, the two scales correlated highly, $r = 0.74$ as the children were three and $r = 0.84$ as the children were four.

### 2.3.3. Teacher reports

From Social Skills Questionnaire (Häidkind et al., 2018) we included three items to the analyses. Social Skills Questionnaire (SSQ) is based on social skills classification (Merrell and Gimpel, 1998) as well as on the Estonian curriculum of preschool childcare institutions. The questionnaire is designed to be filled in by the kindergarten or playgroup teachers who have many opportunities to observe the children in everyday social situations. Three items from Social Skills Questionnaire included to this study were teachers' evaluations of child's talkativeness, size of vocabulary, and complexity of sentences compared to the child's age mates at the first data collection. These evaluations were included in the analyses as these give an indication for teachers' experience with child's language production. The evaluations were given for each of the items as 1 point if the child was at a lower level, 2 points if the child was on bar, and 3 points if the child was at a higher level compared to the age mates. As each of these items was positively correlated with the other two ($rs = 0.51$ to 0.80) we combined the scores of these three into one sum showing teacher's general evaluation of child's language skills.

### 2.3.4. Data analysis

All the answers given by the parents in ECDI III and teachers in SSQ as well as NRDLS test results for the children were included in the data set. Two multiple regression analyses were conducted to determine the extent to which the ECDI-III total score and teacher reports at age 3;0 predict comprehension and production scores at age 4;0, controlling for maternal education (with vs. without higher education), child gender and reported language difficulties (with vs. without language difficulties).

# 3. Results

## 3.1. Internal consistency of ECDI-III

Cronbach's α for the whole list of words was $\alpha = 0.97$ (standardized $\alpha = NA$) as the children were 3 years old. As the children were four there were too many items with null variance to calculate the Cronbach's α for the whole list of words. Cronbach's α for the syntax section were $\alpha = 0.92$ (standardized $\alpha = 0.92$) both as the children were three and four, pronunciation accuracy as the children were three $\alpha = 0.72$ (standardized $\alpha = 0.71$) and as the children were four $\alpha = 0.75$ (standardized $\alpha = 0.75$), metalinguistic awareness as the children were three $\alpha = 0.66$ (standardized $\alpha = 0.63$) and as the children were four $\alpha = 0.57$ (standardized $\alpha = 0.57$).

## 3.2. Variability in children's language measures at both data collection times

As shown in Table 1, children's language skills varied greatly at both time points, regardless of the assessment tool used. Furthermore, in 1 year all language scores central to the study increased significantly.

### 3.2.1. ECDI-III scores

#### 3.2.1.1. Level of communication

Around the third birthday three of the 104 participants were reported by the parents as not yet producing one-word utterances. At the same time 14 of the children were using short utterances and 86 were using sentences. A year later all of the 87 participants were reported by their parents to be using at least one-word utterances. Four of them were using short utterances and 75 were using sentences. Descriptive statistics of the subscales (Vocabulary, Syntax, Metalinguistic skills, Pronunciation) and the total scores of the ECDI-III at two data gatherings are presented in Table 1.

### 3.2.2. The New Reynell developmental language scales scores

The maximum score for language comprehension scale was 72, for language production scale 64, and total maximums score was 136. At 3 years of age the average score for language comprehension was around 49, the average score for language production was 33, and the average total score was 83. At 4 years of age the average score for language comprehension had increased for 10 points, being 59, the average score for language production had increased for 13 points, being 46, and the average total score had increased for 20 points, being 105.

### 3.2.3. Teacher reports

As the children were 3 years old, we asked for teachers' evaluations of child's talkativeness, size of vocabulary, and complexity of sentences compared to the child's age mates with resulting maximum score being 9. At the age of three average score of teachers' evaluations was around 6. See Table 2 for descriptive statistics and for differences between scores from two data collections.

TABLE 1 Descriptive statistics of language measures and differences in scores from two data collections.

| | First data collection at age 3;0 | | | | | Second data collection at age 4;0 | | | | $t$ |
|---|---|---|---|---|---|---|---|---|---|---|
| | $N$ | $M$ | $SD$ | Range | | $N$ | $M$ | $SD$ | Range | |
| ECDI-III | | | | | | | | | | |
| Syntax | 99 | 17.52 | 8.71 | 0–34 | | 78 | 24.56 | 7.20 | 0–34 | 11.48 |
| Pronunciation | 98 | 3.85 | 1.91 | 0–7 | | 79 | 4.38 | 2.10 | 0–7 | 4.40 |
| Metalinguistic skills | 99 | 2.33 | 1.68 | 0–6 | | 79 | 4.33 | 1.65 | 0–7 | 12.90 |
| Vocabulary | 103 | 52.83 | 21.82 | 0–92 | | 79 | 73.18 | 17.91 | 12–100 | 13.40 |
| Total | 98 | 83.94 | 29.05 | 12–138 | | 78 | 113.01 | 24.24 | 27–151 | 16.31 |
| NRDLS | | | | | | | | | | |
| Comprehension | 99 | 48.74 | 10.87 | 9–67 | | 87 | 58.92 | 8.99 | 27–72 | 13.08 |
| Productive | 94 | 33.03 | 13.32 | 2–61 | | 87 | 46.72 | 12.20 | 9–64 | 14.91 |
| Total | 94 | 83.00 | 21.05 | 28–128 | | 87 | 105.64 | 20.33 | 39–136 | 17.24 |
| Teacher report | 84 | 6.26 | 2.21 | 3–9 | | | | | | |

$T$-tests on each subscale and scale were significant at $p < 0.001$.

TABLE 2 Correlations of language measurements at both data collections.

| | First data collection at age 3;0 | | | | | | | | | Second data collection at age 4;0 | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Syntax | Pron | Meta | Vocab | Total | Compr | Prod | Total | | Syntax | Pron | Meta | Vocab | Total | Compr | Prod |
| 1st data collection | | | | | | | | | | | | | | | | |
| ECDI-III | | | | | | | | | | | | | | | | |
| Pron | **0.53** | | | | | | | | | | | | | | | |
| Meta | **0.32** | *0.19* | | | | | | | | | | | | | | |
| Vocab | **0.70** | **0.39** | **0.38** | | | | | | | | | | | | | |
| Total | **0.85** | **0.52** | **0.44** | **0.97** | | | | | | | | | | | | |
| NRDLS | | | | | | | | | | | | | | | | |
| Compr | **0.61** | **0.40** | **0.45** | **0.63** | **0.69** | | | | | | | | | | | |
| Prod | **0.61** | **0.50** | **0.37** | **0.62** | **0.68** | **0.74** | | | | | | | | | | |
| | **0.63** | **0.49** | **0.43** | **0.67** | **0.73** | **0.91** | **0.96** | | | | | | | | | |
| TE | **0.47** | **0.51** | *0.13* | **0.42** | **0.46** | **0.43** | **0.39** | **0.41** | | | | | | | | |
| 2nd data collection | | | | | | | | | | | | | | | | |
| ECDI-III | | | | | | | | | | | | | | | | |
| Syntax | **0.74** | **0.46** | **0.22** | **0.67** | **0.73** | **0.59** | **0.63** | **0.60** | **0.44** | | | | | | | |
| Pron | **0.45** | **0.73** | *0.13* | **0.45** | **0.49** | **0.39** | **0.39** | **0.39** | **0.51** | **0.56** | | | | | | |
| Meta | **0.37** | **0.24** | **0.63** | **0.51** | **0.46** | **0.40** | **0.33** | **0.36** | **0.29** | **0.39** | **0.40** | | | | | |
| Vocab | **0.49** | **0.27** | **0.31** | **0.75** | **0.69** | **0.54** | **0.48** | **0.50** | **0.45** | **0.68** | **0.44** | **0.43** | | | | |
| Total | **0.65** | **0.42** | **0.34** | **0.77** | **0.79** | **0.62** | **0.60** | **0.61** | **0.48** | **0.85** | **0.56** | **0.48** | **0.96** | | | |
| NRDLS | | | | | | | | | | | | | | | | |
| Compr | **0.66** | **0.39** | **0.34** | **0.67** | **0.67** | **0.77** | **0.66** | **0.73** | **0.48** | **0.64** | **0.45** | **0.47** | **0.66** | **0.68** | | |
| Prod | **0.59** | **0.42** | **0.33** | **0.65** | **0.63** | **0.71** | **0.77** | **0.79** | **0.58** | **0.60** | **0.52** | **0.50** | **0.62** | **0.64** | **0.84** | |
| Total | **0.65** | **0.43** | **0.35** | **0.69** | **0.68** | **0.77** | **0.76** | **0.81** | **0.56** | **0.65** | **0.51** | **0.51** | **0.66** | **0.69** | **0.95** | **0.97** |

All statistically significant correlations at $p < 0.05$ are in bold. Pron, Pronunciation; Meta, Metalinguistic skills; TE, Teacher report; Compr, Comprehension; Prod, Production; Vocab, Vocabulary; Total, RCDI-III Total.



10.3389/fpsyg.2023.1179999

## 3.3. Relations between the language measures at both data collections

The correlations for all of the language measures at both ages and between two ages are presented in Table 2. The significant positive correlations of all of the language measures between the first and the second data collection ranged from $r=0.22$ to $r=0.81$, $p<0.05$. The only correlations that were not significant were between metalinguistic awareness at 3 years of age and pronunciation at 4 years of age.

The significant correlations of parental reports of syntax, pronunciation, metalinguistic awareness, and vocabulary with teacher reports of child's language skills ranged from $r=0.41$ to $r=0.51$, $p<0.05$ as the children were 3 years of age. The only correlation that was not significant was between metalinguistic awareness and the teacher report of child's language skills at that age. The correlations of parental reports of syntax, pronunciation, metalinguistic awareness, and vocabulary from the second time of data collection with teacher reports of child's language skills ranged from $r=0.29$ to $r=0.58$, $p<0.05$.

The correlations of ECDI-III subscales with NRLDS subscales ranged from $r=0.37$ to $r=0.69$, $p<0.05$ as the children were 3 years old, from $r=0.45$ to $r=0.66$, $p<0.05$ as the children were 4 years old, and from $r=0.33$ to $r=0.67$, $p<0.05$ between two ages. The correlations of ECDI-III and NRLDS total scores were $r=0.73$ to, $p<0.05$ at 3 years of age, $r=0.69$, $p<0.05$ at 4 years of age, and $r=0.68$, $p<0.05$ between two ages.

## 3.4. Mother and teacher reports as predictors of later language skills

In Tables 3, 4 we provide findings from two separate multiple regression analyses showing to what extent the ECDI-III total scores and teacher reports at age 3;0 predict children's comprehension and production scores measured by the NRDLS at age 4;0, controlling for mothers' education, child gender, and reported language difficulties. In both analyses, we entered the ECDI-III total score first (Model 1). Then we explored teacher reports as the predictor (Model 2). Next, we entered the ECDI-III total score and teacher reports in one model to investigate their combined effect (Model 3). Then we added mothers' education (Model 4), and finally in Model 5 also child gender and reported language difficulties. As shown in Table 3, ECDI-III total score alone explains approximately 45%, and teacher reports alone 23% of the variance in comprehension scores. Together they explained

**TABLE 3** Regression models predicting children's comprehension scores at age 4;0 (NRDLS) on the basis of child language measures at age 3;0.

| Predictors | NRDLS comprehension β-coefficient (standard error) | | | | |
|---|---|---|---|---|---|
| | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 |
| Intercept | 41.50*** (2.27) | 46.64*** (2.86) | 38.47*** (2.74) | 38.56*** (2.70) | 38.06*** (3.77) |
| ECDI total | 0.21*** (0.03) | | 0.18*** (0.03) | 0.17*** (0.03) | 0.17*** (0.03) |
| Teacher report | | 1.96*** (0.43) | 0.90* (0.40) | 0.75 (0.40) | 0.64 (0.44) |
| Maternal education[a] | | | | 2.91 (1.67) | 2.85 (1.70) |
| Gender | | | | | 0.91 (1.66) |
| Reported language difficulties[b] | | | | | −0.68 (2.22) |
| $R^2$ | 0.45 | 0.23 | 0.49 | 0.51 | 0.52 |
| $F$ | 65.54*** | 20.63*** | 32.14*** | 23.09*** | 13.60*** |

ECDI-III—ECDI-III total score. *$p<0.05$, ***$p<0.001$.
[a]Maternal education was represented as a dummy variable with no university education as the reference category. [b]Child's language difficulties was represented as a dummy variable with no language difficulties serving as the reference category.

**TABLE 4** Regression models predicting children's productive scores at age 4;0 (NRDLS) on the basis of child language measures at age 3;0.

| Predictors | NRDLS production β-coefficient (standard error) | | | | |
|---|---|---|---|---|---|
| | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 |
| Intercept | 24.60*** (3.24) | 26.72*** (3.62) | 17.77*** (3.67) | 17.82*** (3.69) | 16.64*** (5.09) |
| ECDI-III | 0.26*** (0.04) | | 0.19*** (0.04) | 0.19*** (0.04) | 0.18*** (0.04) |
| Teacher report | | 3.19*** (0.55) | 2.04*** (0.53) | 1.95*** (0.55) | 1.67** (0.60) |
| Maternal education[a] | | | | 1.59 (2.28) | 1.40 (2.31) |
| Gender | | | | | 2.38 (2.25) |
| Reported language difficulties[b] | | | | | −1.96 (3.00) |
| $R^2$ | 0.39 | 0.34 | 0.50 | 0.54 | 0.52 |
| $F$ | 52.14*** | 34.33*** | 33.87*** | 26.15*** | 13.79*** |

ECDI-III—ECDI-III total score. **$p<0.01$, ***$p<0.001$.
[a]Maternal education was represented as a dummy variable with no university education as the reference category. [b]Child's language difficulties was represented as a dummy variable with no language difficulties serving as the reference category.

49% of the variance. When adding mothers' education, the R-squared statistics increases 2%, and only the ECDI-III total score remained the significant predictor. When adding child gender and reported language difficulties only the ECDI-III total score remained the significant predictor. Neither teacher reports nor the control predicted comprehension scores significantly.

As shown in Table 4, the ECDI-III total alone predicted 39%, and teacher reports alone 35% of the variability in production scores. When combined they predicted 50% of the variance in production scores and both remained significant predictors. When mother education was added in model already containing the ECDI-III total score and teacher reports (Model 3), the R-squared statistic increased to 54%, but mother education was not a significant predictor. Adding child gender and reported language difficulties decreased approximately 2% the predictability of production scores, the ECDI-III and teacher reports remained predictive. None of the controls was predictive of production scores at age 4;0.

# 4. Discussion

Numerous studies have demonstrated that early language skills are good indicators of children's concurrent and future development and adjustment. However, some researchers have questioned the concurrent and predictive validity of parent report measures as assessment tools of language skills of children after their first years of life, because with growing age children become more communicative, having more conversational partners and interactional contexts than during their first years of life. Therefore, the present study sets out to compare parent reports with two other sources of information about Estonian children's language skills—teacher reports and experimenter assessments—at ages 3;0 and 4;0 years.

## 4.1. Concurrent correlations between the ECDI-III, teacher reports and directly measured language scores

The first aim of the study was to explore the concurrent validity and utility of report measures in estimating child language skills at ages 3;0 and 4;0, taking measures of language comprehension and production administered by an expert examiner via the NRDLS as a golden standard. Results of correlational analysis indicated significant positive correlations of acceptable magnitude ($rs = 0.64–0.69$) between the ECDI-III total scores and with directly measured language comprehension and production scores. The strongest correlations of directly measured language comprehension and production scores were with vocabulary and syntax scores of the ECDI-III. The finding suggested that vocabulary and grammar development were the most indicative CDI measures of children's language skills also in the age period studied in our study. Other aspects of language development (i.e., pronunciation and metalinguistic abilities) provided only some additional information. The results are in line with most validation studies with younger children, where only the vocabulary list or in some studies vocabulary and grammar sections were addressed (Fenson et al., 2007).

The study found lower and moderate ($rs = 0.37–0.44$ at Wave 1, and $rs = 0.48–0.58$ at Wave 2), albeit significant correlations between scores reported by teachers and those obtained by parent report or direct assessments. Correlations of teacher reports with direct language measures were lower than those of parent reports, likely in part, since teachers were asked only 3 questions – to estimate child's talkativeness, size of vocabulary, and complexity of sentences compared to the child's age mates. A reason for low correlations between two report measures may also lie in good communicative abilities of children at this age that makes it difficult for a reporter to capture all of what children are able to say. Moreover, parents and teachers observe children communicating in different interactional contexts and with different communicative partners (De Houwer et al., 2005). Keeping this in mind, our results suggest that teachers are capable of reporting on 3-and 4-year-old children's communication skills and teacher ratings are a good source of information about children's language skills. Both parent and teacher reports were congruent with direct assessments. Differently from parents, teachers have the privilege to observe and compare language skills of many same-age children (Vagh et al., 2009). Despite of this, investigating the validity of reports made by parents and teachers against the direct measure of child language comprehension and production, parents turned to be better reporters than teachers.

## 4.2. Mother and teacher reports as predictors of later language skills

Our second aim was to find out how well two different sources of information—parent and teacher reports—predict future language abilities, considering also mother's education, child gender and reported language difficulties. Results revealed that parental reports on children's earlier language skills (ECDI-III total scores) and teacher reports were important predictors of language comprehension and production scores assessed by standardized language measures 1 year later. At the same time, the ECDI-III total score predicted later language skills, especially comprehension scores, better than teacher reports. Thus, parental report measure showed in addition to concurrent validity also good predictive validity. This is consistent with previous findings of the validity of CDIs, suggesting that parents are well-informed about their children's communicative skills (Pan et al., 2004; Fenson et al., 2007).

Furthermore, although inclusion of mothers' education in the models already containing parent-and teacher-reported language skills at age 3;0 explained variance in future comprehension and production scores significantly better than previous language skills alone, education turned out to be a nonsignificant predictor. The finding did not confirm the effect of parental education on children's language skills, although this has been frequently reported in the literature (Fenson et al., 2007). Furthermore, subsequent language skills were not predicted by gender. These findings contradict several previous studies reporting gender-differences in language development (Bornstein et al., 2004; Eriksson et al., 2012). A possible explanation may be the relatively high educational level of mothers who participated in our study, as gender differences in language development have been found to be larger in lower SES compared to upper SES families (Barbu et al., 2015).

Most previous research on predictors of language skills has focused more on expressive vocabulary than on other dimensions of infants' and toddlers' communicative abilities measured by the CDIs (Fenson et al., 2007). Some of our findings that differ from previous studies (e.g., no effect of mothers' education, child gender, and reported language difficulties) can be attributed also to older age of children who participated in our study and that we addressed more general language skills.

The use of parent reports with children older than three has been a concern because their communicative abilities have grown and they spend more time out of their homes, being exposed to various conversational partners and interactional contexts. The practical importance of our study is that it proved that parents of children at 3;0 and 4;0 years of age provide adequate information about their children's language skills and that they are still best reporters on these skills. There remains a need for more information on how good estimators of children language skills teachers are. The utility and validity of teachers as reporters of child language skills is particularly pressing as teachers should identify children with language problems as early as possible. Significant, although modest correlations between mother and teacher reports albeit teacher ratings based only on 3 items provided evidence for the utility of teacher reports to receive useful information about children's language skills. The study pointed out that in order to understand whether teacher reports are in accordance with parent reports, it is important to compare their ratings by using the same report instrument (e.g., CDIs). Of course, it is time consuming and a burden for teachers to report on each child in the group of many children. Until now, there is only one CDI study comparing parents and teachers as reporters, but it has been done with bilinguals from lower-SES families (De Houwer et al., 2005).

A limitation of the study is that the predictive validity of reports was not studied over a longer interval than 1 year. Longitudinal studies are needed to find out how well the language skills reported for the age group predict their skills in the long term. Moreover, the usability of parent reports of children's language skills also needs to be investigated for children over 4;0 years of age.

### 4.3. Conclusion

The current study showed that parent reports on children's language skills at 3.0 and 4;0 years of age are indicative of concurrent language skills, and valid predictors of subsequent language skills 1 year later. There is a need for more information about how good reporters of children's language skills teachers are when using an assessment tool such as the CDI-III. The knowledge is useful for practice as might reduce negative consequences of language problems by timely identification and support.

## Data availability statement

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

## Ethics statement

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

## Author contributions

TT formulated the research questions, supervised the data collection, and wrote the first draft of the manuscript. AS organized the dataset, participated in data analysis and interpretation of the data, and participated in writing the method section of the manuscript. All authors approved the submitted version.

## Funding

This study was supported by the Estonian Research Council (grant number: PRG1761).

## Conflict of interest

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

## Editor's note

Melita Kovacevic edited the article in collaboration with Maria-Jose Ezeizabarrena. Faculty of Arts, University of the Basque Country UPV/EHU Vitoria/Gasteiz, Spain.

## Publisher's note



## References

Barbu, S., Nardy, A., Chevrot, J.-P., Guellaï, B., Glas, L., Juhel, J., et al. (2015). Sex differences in language across early childhood: family socioeconomic status does not impact boys and girls equally. *Front. Psychol.* 6:1874. doi: 10.3389/fpsyg.2015.01874

Bleses, D., Jensen, P., Højen, A., and Dale, P. S. (2018). An educator-administered measure of language development in young children. *Infant Behav. Dev.* 52, 104–113. doi: 10.1016/j.infbeh.2018.06.002

Bleses, D., Makransky, G., Dale, P., Højen, A., and Ari, B. (2016). Early productive vocabulary predicts academic achievement 10 years later. *Appl. Psycholinguist.* 37, 1461–1476. doi: 10.1017/S0142716416000060

Bornstein, M. H., Hahn, C.-S., and Haynes, O. M. (2004). Specific and general language performance across early childhood: stability and gender considerations. *First Lang.* 24, 267–304. doi: 10.1177/0142723704045681

Cadime, I., Santos, A. L., Ribeiro, I., and Viana, F. L. (2021). Parental reports of preschoolers' lexical and syntactic development: validation of the CDI-III for European Portuguese. *Front. Psychol.* 12:677575. doi: 10.3389/fpsyg.2021.677575

Dale, P. S., Price, T. S., Bishop, D. V., and Plomin, R. (2003). Outcomes of early language delay: I. predicting persistent and transient language difficulties at 3 and 4 years. *J. Speech Lang. Hear. Res.* 46, 544–560. doi: 10.1044/1092-4388

De Houwer, A., Bornstein, M. H., and Leach, D. B. (2005). Assessing early communicative ability: a cross-reporter cumulative score for the MacArthur CDI. *J. Child Lang.* 32, 735–758. doi: 10.1017/S0305000905007026

Dionne, G., Dale, P., Boivin, M., and Plomin, R. (2003). Genetic evidence for bidirectional effects of early lexical and grammatical development. *Child Dev.* 74, 394–412. doi: 10.1111/1467-8624.7402005

Edwards, S., Letts, C., and Sinka, I. (2011). *The New Reynell developmental language scales*. London: GL Assessment Ltd.

Eriksson, M. (2017). The Swedish communicative development inventory III: parent reports on language in preschool children. *Int. J. Behav. Dev.* 41, 647–654. doi: 10.1177/0165025416644078

Eriksson, M., Marschik, P. B., Tulviste, T., Almgren, M., Pérez Pereira, M., Wehberg, S., et al. (2012). Differences between girls and boys in emerging language skills: evidence from 10 language communities. *Br. J. Dev. Psychol.* 30, 326–343. doi: 10.1111/j.2044-835X.2011.02042.x

Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, J. S., and Bates, E. (2007). *MacArthur-Bates communicative development inventories: user's guide and technical manual*. Baltimore, MD: Paul H. Brookes Publishing Co.

Häidkind, P., Schults, A., and Palts, K. (2018). 3-4aastaste eesti laste üldoskuste tase kolme hindamisvahendi alusel. *Estonian J. Educ.* 6, 179–214. doi: 10.12697/eha.2018.6.1.08

Kas, B., Jakab, Z., and Lőrik, J. (2022). Development and norming of the Hungarian CDI-III: a screening tool for language delay. *Int. J. Lang. Commun. Disord.* 57, 252–273. doi: 10.1111/1460-6984.12686

Law, J., and Roy, P. (2008). Parental report of infant language skills: a review of the development and application of the communicative development inventories. *Child Adolesc. Mental Health* 13, 198–206. doi: 10.1111/j.1475-3588.2008.00503.x

Merrell, K., and Gimpel, G. (1998). "Social Skills of Children and Adolescents," in *Conceptualization, Assessment, Treatment.* (Mahwah, NJ: Lawrence Erlbaum Associates).

Pan, B. A., Rowe, M., Spier, E., and Tamis-Lemonda, C. (2004). Measuring productive vocabulary of toddlers in low-income families: concurrent and predictive validity of three sources of data. *J. Child Lang.* 31, 587–608. doi: 10.1017/S0305000904006270

Simonsen, H. G., Kristoffersen, K. E., Bleses, D., Wehberg, S., and Jorgensen, R. N. (2014). The Norwegian communicative developmental inventories: reliability, main developmental trends and gender differences. *First Lang.* 34, 3–23. doi: 10.1177/0142723713510997

Stiles, J. (1994). On the nature of informant judgments in inventory measures:… and so what is it you want to know? *Monogr. Soc. Res. Child Dev.* 59, 180–185.

Tulviste, T., and Schults, A. (2020). Parental reports of communicative development at the age of 36 months: the Estonian CDI-III. *First Lang.* 40, 64–83. doi: 10.1177/0142723719887313

Urm, A., and Tulviste, T. (2016). Sources of individual variation in Estonian toddlers' expressive vocabulary. *First Lang.* 36, 580–600. doi: 10.1044/2020_JSLHR-20-00201

Vagh, S. B., Pan, B. A., and Mancilla-Martinez, J. (2009). Measuring growth in bilingual and monolingual children's English productive vocabulary development: the utility of combining parent and teacher report. *Child Dev.* 80, 1545–1563. doi: 10.1111/j.1467-8624.2009.01350.x

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 03 August 2023
DOI 10.3389/fpsyg.2023.1214518

Check for updates

OPEN ACCESS

EDITED BY
Melita Kovacevic,
University of Zagreb, Croatia
Maria-Jose Ezeizabarrena,
University of the Basque Country UPV/EHU,
Spain, in collaboration with editor MK

REVIEWED BY
Natalie Ann Munro,
The University of Sydney, Australia
Barbara Blaha Degler,
National Autonomous University of Mexico,
Mexico

*CORRESPONDENCE
Daniela Gatt
✉ daniela.gatt@um.edu.mt

RECEIVED 29 April 2023
ACCEPTED 17 July 2023
PUBLISHED 03 August 2023

CITATION
Gatt D, Camilleri L and Grech C (2023) Word usage as measured by parental checklists and language samples: trends, comparisons, and implications.
*Front. Psychol.* 14:1214518.
doi: 10.3389/fpsyg.2023.1214518



# Word usage as measured by parental checklists and language samples: trends, comparisons, and implications

Daniela Gatt[1]*, Liberato Camilleri[2] and Chloe Grech[1]

[1]Department of Communication Therapy, Faculty of Health Sciences, University of Malta, Msida, Malta,
[2]Department of Statistics and Operations Research, Faculty of Science, University of Malta, Msida, Malta

**Background:** Although parental checklists are well-known for their potential in indexing young children's lexicon size, they can also be used to track children's acquisition of individual words. Word-level data can be used to identify the checklist words most and least commonly employed across groups of children. Like parent-completed vocabulary checklists, samples of spontaneous language use collected from multiple children can also generate measures of word commonality, concerned with the numbers of children producing individual words. To our knowledge, comparisons of word usage as determined by parental checklist and language sample data obtained in parallel from the same children have not been carried out. Also scarce in the empirical literature are item-level analyses of early bilingual lexicons that explore word usage across two emerging languages. The present study aimed to contribute toward bridging both gaps through the analysis of data generated by a bilingual Maltese-English adaptation of the vocabulary checklist of the MacArthur Communicative Development Inventories: Words and Sentences (CDI: WS) and spontaneous language samples for the same children. An additional objective was to derive implications for revising the current version of the vocabulary checklist, in preparation for its eventual standardization.

**Materials and methods:** For 44 Maltese children aged 12, 18, 24, and 30 months, the words reported by their main caregivers on the vocabulary checklist were identified, along with their respective semantic categories. For the same children, 20-min language samples obtained during free play with the caregiver were transcribed orthographically. Words identified through parental report and language sampling were analyzed for commonality, i.e., the number of children producing each word.

**Results:** Comparison of the word usage patterns obtained through both methods indicated differences in the words most commonly sampled and those most commonly reported, particularly in relation to grammatical categories. Notwithstanding these differences, positive and significant correlations emerged when considering all grammatical categories and languages across commonality levels.

**Discussion:** The commonality scores based on parental checklist data have implications for reconsidering the length and language balance of the Maltese-English adaptation of the CDI: WS vocabulary checklist. Sampled word usage patterns can contribute additional objectivity in updating the reporting instrument in preparation for its eventual standardization.

KEYWORDS
vocabulary checklist, parental report, language sampling, word frequency, vocabulary acquisition, bilingual acquisition

## 1. Introduction

There is a wealth of evidence in favor of using parent report tools to measure young children's language skills for research and clinical purposes. When parents and primary caregivers are asked to describe their children's emergent language through interviews, questionnaires and checklists, they are known to impart reliable and valid information. This is because, typically, primary caregivers are attentive to their children's early language milestones, monitoring their emergent skills closely across daily settings (Fenson et al., 1994). In particular, parental report enables the collection of comprehensive vocabulary data from extensive samples of children, enhancing the recognition of universal trends and natural variability in language acquisition through a relatively undemanding method (Frank et al., 2017). Reported vocabulary data also contribute to a better understanding of the developmental trajectories of individual words (Fenson et al., 1994) and how these compare across different languages (Frank et al., 2021).

### 1.1. The MacArthur-Bates Communicative Development Inventories

The MacArthur-Bates Communicative Development Inventories (CDIs) are among the parent-report instruments most widely used to measure children's early language skills, including vocabulary. Adaptations of the original U.S. English version span several languages (see Dale and Penfold, 2011). Importantly, CDI vocabulary measures are not intended as an exhaustive inventory of the words known by the child, for comparison with measures generated by other tools (Frank et al., 2021). Rather, they function as an index of children's lexical abilities relative to their peers', measured through the same instrument (Fenson et al., 1994).

Although vocabulary checklists tap strategically into parents' familiarity with their children's early lexicons, they are also subject to reporting biases. Parental estimates of children's early vocabulary and emergent grammar skills may be inaccurate, particularly if lower income and educational levels are present (e.g., Roberts et al., 1999; Feldman et al., 2000). Socioeconomic variables are less consistently associated with early language difficulties than neurobiological factors (Rescorla and Dale, 2013), suggesting that lower parent-based scores among disadvantaged groups may be attributed more directly to incomplete reporting than to the impact of unfavorable environmental conditions. Besides, details reported by parents would have been filtered through their own subjective perceptions (Stiles, 1994). In fact, vocabulary checklists attempt to optimize parents' reporting ability by prompting them to record their children's current and newly emerging lexical skills through a recognition format (Dale, 1996). The addition of words recollected by parents from memory may be relevant during the earlier stages of instrument design (e.g., Fenson et al., 1994). Reliance on a recognition format also means that the extensiveness of parent-reported data is regulated by the specific reporting opportunities, that is, the words available on the checklist for ticking. Nonetheless, the wealth of in-depth and dependable vocabulary data yielded by parental report goes a long way in mitigating its methodological bias. In fact, the CDI vocabulary checklists have an impressive track record of eliciting reliable and valid measures of children's expressive lexicon size (e.g., Galeote et al., 2016; Frank et al., 2021; de Anda et al., 2022).

### 1.2. Comparisons between parental report and language sample measures

The validation of newly-developed parent report instruments requires comparability to direct assessment measures. Given the scarcity of norm-referenced tests for young children, concurrent measures for establishing validity are often obtained through language sampling and informal structured assessment. Parental estimates of children's vocabularies have been compared to the lexical skills emerging spontaneously during naturalistic observation, or elicited through structured testing, with positive and significant correlations often resulting (see Fenson et al. (2007) for a review). Moderate to high correlations with sampled vocabulary measures have also been reported for CDI adaptations to numerous languages representing a range of language families, e.g., Spanish (Jackson-Maldonado et al., 1993); Danish (Bleses et al., 2008); Kigiriama and Kiswahili (Alcock et al., 2015), as well as bilingual adaptations for children exposed to specific language pairs (see Gatt et al. (2014) for Maltese-English; O'Toole and Fletcher (2010) for Irish-English), as well as language adaptations for children with specific disorders (see, e.g., Galeote et al. (2016) for the Spanish adaptation for children with Down Syndrome), substantiating the concurrent validity of CDI-based vocabulary measures.

Comparisons of parent-reported and sampled vocabulary scores have also served the purpose of establishing whether the presence of a noun bias in children's early lexicons is subject to methodological influences. Substantial cross-linguistic evidence points toward a general mechanism in vocabulary composition, whereby young children's expressive vocabularies start off with a predominance of social words (e.g., sound effects and routine words) which gives way to nouns, followed by a subsequent emphasis on predicates, comprising main verbs and adjectives, and eventually culminates in function words. Earlier findings in the field drew on the original CDI (Bates et al., 1994) and its adaptations to other languages (e.g., Caselli et al. (1999) for Italian). Although a noun bias has also been identified through spontaneous language sampling (see Bassano (2000) for French; Kauschke and Hofmeister (2002) for German), conflicting evidence (e.g., Tardif et al., 1999) prompted consideration of methodological factors potentially impacting on children's vocabulary composition. A small body of research has therefore employed a combination of caregiver report and sample measures to minimize methodological bias, while keeping language constant. For example, when employing reported and sampled vocabulary measures in parallel, Salerni et al. (2007) consistently identified a predominance of nouns in Italian's children's 200- and 500-word vocabularies. However, significant differences across methods emerged in grammatical category proportions.

Such findings underscore the fact that the choice of method for documenting early vocabulary skills may influence the measures obtained. However, the parameters of this methodological distinction may appear ambiguous, particularly since the empirical literature shows parental report and language sampling to concur time and again in their measurement of expressive lexicon size. Yet, they are intrinsically different. Checklist measures of vocabulary span various daily settings in which parents observe their children's available language skills. In contrast, sampled vocabulary production draws on a limited window in which the child's compliance, interlocutor's input and context of interaction bear directly on the amount and representativeness of data obtained (Bates et al., 1988; Frank et al., 2021). Observed language behaviors have been fittingly described as 'sporadic' (Fenson et al., 1994: 11).

Despite differences in absolute scores, reported and sampled vocabulary size for the same children tend to rank similarly. However, the occurrence of specific words and their frequency of production are known to be highly sensitive to sampling parameters, such as the toys employed during free play (Frank et al., 2021). In contrast, words reported by parents draw on the child's participation in various interactional exchanges and do not incorporate information on frequency of occurrence. This difference was aptly synthesized by Bates et al. (1988), who contrasted vocabulary checklists' potential to identify the words children *know* with language samples' potential to capture the words children *use*. In terms of actual word usage, the two methods are therefore expected to be discordant, suggesting that differences between methods at the item level may be more pronounced than for composite counts of words. This could partly explain the marked absence of comparisons between checklist and sample word-level data from the research literature.

The purpose of the present study was twofold. First, it identified the more commonly produced words identified through each method for a single cohort and investigated the extent to which these reported and sampled words overlapped. Second, it sought to derive guidelines for a subsequent iteration of the Maltese-English vocabulary checklist adaptation, in preparation for norming. The CDI Advisory Board recommends that CDI adaptations are piloted extensively, with detailed item-level analyses of checklist data and language sampling being requisites for arriving at a final set of words that is amenable to norming. A careful, data-driven, approach is particularly critical to the development of parent report instruments intended to measure early bilingual vocabularies, since extensive individual variability stemming from language exposure variables is expected (Weisleder et al., 2022). Revising the current checklist in light of this study's findings would be a crucial step toward eventually standardizing the Maltese-English CDI adaptation. A new revision of the Maltese-English vocabulary checklist adaptation would comprise the fourth iterative cycle of the instrument.

The methodological issues addressed by this study also have theoretical ramifications that stem from its focus on an under-researched language pair. Recent years have seen more publications on Maltese children's bilingual acquisition of Maltese and English in prominent language acquisition journals than in the past (Kidd and Garcia, 2022). However, documentary evidence is still largely lacking, despite the fact that Maltese and English are two languages with highly dissimilar typologies, making for more valuable comparisons across them (Slobin and Bowerman, 2007). Moreover, the normative bilingual context in which these two languages are acquired (Gatt and Dodd, 2019) adds to the relevance of documenting Maltese children's bilingual acquisition in detail, particularly since the study of normative bilingualism holds immense theoretical potential (Montanari and Nicoladis, 2016).

The nature and scope of word usage data, from the methodological perspective of parental report and language sampling, is reviewed in the next section. In the present text, we use the term 'word usage' to refer to the occurrence of individual words in vocabulary data, so that our primary focus is on item-level trends in children's expressive vocabularies rather than on aggregated scores of vocabulary size.

## 1.3. Word-level vocabulary checklist measures

Vocabulary checklists can never claim to include all the words that young children understand and/or produce as their language skills emerge. The diversity of words that young children accumulate through interactions in specific language-learning environments, together with the rapidity with which their vocabularies grow, imply that beyond the stage of children's first words, an exhaustive vocabulary checklist is barely conceivable. Its sheer length would also make it unwieldy and daunting to complete. Vocabulary checklists therefore seek to present parents, or primary caregivers, with a sample of words that realistically represent children's lexical repertoires (Fenson et al., 1994). Arriving at a set of words that best characterizes typically-developing children's varying levels of typical lexical development is one objective of item-level vocabulary measures. For example, the current version of the U.S. English CDI vocabulary checklists was developed from several iterations based on data collected through parental questionnaires that preceded the CDIs (Fenson et al., 1994). Scrutiny of these data shed light on the psychometric properties of the instruments. In particular, information on the rate and pattern of growth shown by individual words delineate their sensitivity to developmental change. Usage of individual words reported across children plays a role in informing decisions on which items to discard and retain, so that composite scores across all words can then reflect vocabulary development tendencies (Fenson et al., 1994). Importantly, checklist versions in other languages are emphasized to be adaptations, rather than translations. The original list of words should be assessed for cultural and linguistic relevance to the population of interest. The MacArthur-Bates Advisory Board encourages a similar distribution of words across difficulty levels and grammatical categories in CDI adaptations as in the original (Frank et al., 2021). Studies evaluating the psychometric properties of newly-adapted instruments may therefore resort to analyses of item-level performance. For example, Weber et al. (2018) examined children's responses to individual items on the Wolof adaptation of the CDI:WS short form, to ascertain that the set of words captured varying levels of vocabulary ability for Wolof-learning children in the target age range.

## 1.4. Word-level language sample measures

In the analysis of child language samples, two expressive vocabulary measures that feature often are type and token counts, representing the number of different words and total number of words produced, respectively. Although their computation involves scrutiny of individual words produced to distinguish the unique and recurring ones, the resulting measures are broad-based rather than focused on individual items. A body of research, however, has gone beyond the identification and counting of early words emerging in naturalistic contexts, zooming in on the occurrence of each word in terms of commonality and frequency. In a landmark study by Beukelman et al. (1989), every word occurring in classroom language samples obtained from six typically-developing children was examined for commonality, or consistency of use across the group (i.e., the number of children employing it), as well as frequency of use by each child and across participants. These measures, obtained from 3- and 4-year-olds, were intended to guide the choice of vocabulary for non-verbal classmates using augmentative and alternative means of communication (AAC). The most commonly used words, referred to as core vocabulary, were also those occurring most frequently in the samples. In contrast, fringe vocabulary consisted of words showing limited usage. Fringe word repertoires are much larger than core vocabularies, reflecting personalized interests and routines (Trembath et al., 2007). Studies measuring word usage by children aged 3 and younger are very limited. This could be partly due to the delicate task of assigning word status to early productions that are partly or largely unintelligible (see Vihman and McCune (1994) for a detailed discussion of early word identification). Also conspicuous is the paucity of sampled word usage investigations in bilingual contexts. To our knowledge, only Robillard et al. (2014) have addressed bilingual children's word usage, with a sub-group of their school-aged participants being French-English speakers. Moreover, lengthy and labor-intensive transcription procedures likely explain the compromise between number of participants and sampling duration required in study designs. For example, Banajee et al. (2003) analyzed word usage for a sample of 50 typically-developing English-speaking children aged 24–36 months. Analysis drew on the first 150 utterances produced during two daily activities in nursery and daycare settings over three days. Trembath et al.'s (2007) study of word usage focused on a sample of six typically-developing Australian children including 3-year-olds (range = 3–5 years). For each child, analysis was based on a sample of 3,000 words collected in their preschool classroom. In a narrative review of research literature in the field, Laubscher and Light (2020) flagged the considerable variation across published word lists, attributing this to differences across studies in methods, contexts of sampling and criteria for defining words. Beyond these differences, function words are consistently prominent in core word lists, with nouns featuring rarely (see Beukelman et al., 1989; Banajee et al., 2003; Trembath et al., 2007). More recently, Frick Semmler et al. (2023) examined seven published core vocabulary lists, five of which also featured in Laubscher and Light's (2020) review. While highlighting the general predominance of verbs, findings also revealed that none of the listed words appeared before the age of 25 months in typically-developing children.

## 1.5. Comparisons between reported and sampled measures at the word-level

It is noteworthy that item-level comparisons of expressive vocabularies as documented by caregiver report and direct assessment methods have rarely been reported. Among these few investigations is Dale's (1991) comparison between words elicited from English-speaking 24-month-olds on a standardized picture naming test and CDI items reported by their parents. On the items common to both instruments, average agreement was 72.5%, supporting the CDI's validity. Most mismatches resulted from words reported by parents not emerging on direct assessment, with factors such as children's compliance considered as likely contributors. Ring and Fenson (2000) compared 40 toddlers' CDI expressive vocabulary scores to parents' estimates of the children's picture naming skills and their actual naming performance. The purposely-designed picture booklet contained images of 35 words available for reporting on the CDI. Mean naming judgment scores were higher than actual naming performance for all but three items having moderate to difficult levels. Positive and significant correlations between CDI-based scores, parental naming estimates and child naming scores, even when checklist scores were based on a random sample of items not represented on the naming task, were taken as support for checklist validity. Gatt et al. (2014) investigated the correspondence between individual checklist items reported by parents of Maltese children aged 12–30 months and words produced by the children themselves on an informal picture naming task, reporting a percentage agreement in the range of 78–84%. Beyond the matching of checklist words and lexical items elicited through structured vocabulary tasks, there does not seem to be research comparing the individual words reported by parents to those sampled naturalistically. In the study reported here, we address this gap in the research literature.

Prior to evaluating trends in the usage of reported and sampled words, the parameters of the comparison need to be established. A fundamental consideration is the divergences between parental report and language sampling in the terminology adopted and the nature of the data yielded. First, checklist and sample data construe frequency of word usage differently. Word frequencies derived from CDI data specify the numbers of children reported to use each checklist entry (see Fenson et al., 1994; Frank et al., 2021). In sample data, frequency quantifies word occurrences in a snapshot of naturalistic language use, at individual or group level. The number of times a word occurs in sample data collected from different individuals may be referred to as 'composite frequency' (e.g., Trembath et al., 2007). Commonality, defined as the number of participants using a particular word during sampling (e.g., Banajee et al., 2003), is akin to word frequencies derived from checklist data. Further, the frequency of production of individual words generated by sampled vocabulary data cannot be gleaned from completed checklists. To avoid confusion, in our measures and results we use the term 'commonality' to refer to the number of children using a specific word in both checklist and sample data; we take 'frequency' to denote the number of times an item occurs in participant samples considered collectively. Second, sampled 'core' word data and parent-reported vocabulary inventories are also intrinsically different in layout and purpose. Core vocabularies are relatively short, functioning as "a framework for functional language use" (Banajee et al., 2003, p. 68) and remaining consistent across settings and individuals. In contrast, parent-based vocabularies bring together a range of words commonly known by young, typically-developing children. This is because they draw on the contents of the CDI vocabulary checklists, which contain words most commonly understood and produced by typically-developing children within the specified age bracket. Thus, a child's lexicon size can be estimated by the parent and, depending on the availability of norming data, compared to standardized measures obtained for the instrument. Interestingly, Laubscher and Light (2020) point out that checklist items are very different from the words ranking highly in language sample corpora, according to their review of various core vocabulary lists in the empirical literature. These lists consistently feature a preponderance of function words, along with a scarcity of nouns and social words, that together represent only around one fifth of the items listed in the CDI vocabulary inventories. Laubscher and Light's (2020) comparative tabulations revealed that words available for reporting in CDI instruments coincided minimally with sampled core nouns, while pronouns, question words, prepositions and other function words on core lists show the highest percentage overlap. Limited overlap between sampled core words and CDI checklist items was also reported by Frick Semmler et al. (2023). Both studies attributed this disparity to CDI checklists capturing individualized components of young children's typical vocabularies, which core vocabulary lists are unable to detect due to sampling constraints.

### 1.6. Research questions

To our knowledge, there have been no studies that compared word usage as documented through parental report and language sampling for the same children. The present study addresses this evidence gap. It investigates word usage in a cohort of 12-30-month-olds predominantly exposed to Maltese. Measures obtained in a naturalistic setting were compared to those derived from caregiver report, which was based on a Maltese-English adaptation of the CDI: WS vocabulary checklist. The research questions addressed are the following:

1. Which words were most commonly reported by caregivers for Maltese children aged 12, 18, 24, and 30 months, in terms of grammatical category and language?
2. Which words were most commonly used by the same children spontaneously during free play with their caregivers, in relation to grammatical category and language? Which words were most frequently used?
3. How did the most commonly reported words compare to the most commonly and the most frequently sampled words, for the same children? Were there similarities when word class and language were considered?

## 2. Materials and methods

### 2.1. Participants

The participants were 44 typically-developing Maltese children aged 12, 18, 24, and 30 months. Each age group consisted of approximately equal numbers of boys and girls (12 months: 6 boys, 5 girls; 18 months: 5 boys, 7 girls; 24 months: 5 boys, 6 girls; 30 months: 5 boys, 5 girls). The main caregivers of all participants were mothers, except for one 30-month-old boy who was mostly cared for by his grandmother. While all the children's parents had a secondary level of education, 17 mothers and 16 fathers had pursued their studies to post-secondary level. Eleven parents were in possession of a university degree, with one mother and one father also having a postgraduate qualification. Seven children were randomly selected from the National Register of Births in Malta. The remaining children were identified through snowball sampling. None of the participants manifested features that clearly impaired their language development at the time of data collection. In the absence of norms for early language acquisition based on Maltese children, participants were judged to be developing typically by the speech-language pathologist collecting the data, the first author. Data from all potential participants were collected by the same person, so clinical judgment was applied uniformly. No significant medical conditions were reported for any of the children. Preterm birth at 32 and 34 weeks ($N=2$), occurrence of middle ear infections ($N=10$) and the presence of speech or language difficulties in an older sibling ($N=2$) (none reported in the parents) were not considered as exclusionary criteria. Since the two participants born prematurely were healthy preterm infants, they were likely to have better language outcomes than preterm infants with identified medical conditions (see Loeb et al., 2020). Roberts et al.'s (2004) meta-analysis identified very small associations, if at all, between a history of otitis media in early childhood and later language outcomes. Moreover, preterm birth and middle ear infections do not feature in models of strongly weighted risk factors for later language disorder (e.g., Ellis and Thal, 2008; Zambrana et al., 2014; Bishop et al., 2016). Fisher's (2017) systematic review and meta-analysis reported that the presence of a speech or language disorder, or learning disability, in parents and/or siblings, was not a significant predictor of expressive language outcomes in late talkers. In view of these research findings, children were not excluded from the study on the basis of preterm birth, middle ear infections or speech or language difficulties in older siblings, since these factors did not appear to impact language skills at the time of data collection and were not necessarily predictive of later language difficulties. The study was approved by the University of Malta Research Ethics Committee. Primary caregivers gave informed proxy consent for their children's participation in the study and consented to their own involvement.

Each child was exposed primarily to Maltese within the home context. In Malta, both Maltese and English carry the status of official languages, with bilingualism being widespread. Maltese, the national language, is essentially Semitic in origin but incorporates Romance and English borrowing (Hoberman, 2007). Typologically, Maltese is very different from English, a Germanic language. Among the characteristics of the Maltese language are its rich inflectional and derivational morphology, its optional subject forms, made possible by the person, number and gender inflections coded on the verb, and its free word order (Borg and Azzopardi-Alexander, 1997). Free and suffixed pronouns are marked for first, second and third person, with singular and plural distinctions also coded for each person. Pronominal suffixes attached to nouns mark possession, to verbs, where they mark direct and indirect objects, and to prepositions as their objects (Borg and Azzopardi-Alexander, 1997; Hoberman, 2007).

Maltese is the dominant language of most Maltese individuals (National Statistics Office, 2007; Vella, 2013; National Council for the Maltese Language, 2021). Since Maltese and English exist in close proximity with each other, they exert a degree of cross-linguistic influence on each other. The variety of English spoken in Malta is often referred to as Maltese English, in recognition of the fact that it is influenced by the pronunciation, intonation, grammar and vocabulary of Maltese (Borg, 1988; Brincat, 2011; Krug and Sönning, 2018). On the other hand, spoken Maltese regularly features the use of English words, phrases, sentences and stretches of discourse (Borg, 1988; Camilleri Grima, 2013).

Monolingual input is highly unlikely for Maltese children (Vella, 2013). Stable bilingualism at a societal level and extensive language contact mean that children receive both Maltese and English input from a very early age, with amount and timing of exposure varying across households (Camilleri, 1995; Gatt and Dodd, 2019). Adults speaking Maltese to their young children often prefer English words or phrases over their Maltese equivalents (Borg, 1988). This 'functional borrowing' pattern characteristic of child-directed language use is potentially explained by the relatively simpler phonotactic structure of English, despite Maltese and English have similar consonantal phonetic inventories (Grech and Dodd, 2008; Galea and Ussishkin, 2018). This language choice pattern, specific to adult-child dyads, supplements the established borrowings, core borrowings and single-word code-switches from English expected in spoken Maltese (Gatt et al., 2011; see also Myers-Scotton, 2002, 2006). While established English borrowings compensate for lexical gaps in Maltese, e.g., *stiker* (sticker), core borrowings are English words employed predictably instead of available Maltese equivalents, as in the case of *toys* typically being preferred to *ġugarelli* in both adult- and child-directed contexts. Functional borrowing is thus a form of core borrowing specific to child-directed language use. On the other hand, single-word switches involve the sporadic preference of an English word to a Maltese equivalent. In this study, the participants' Maltese-dominant home language exposure was established upon initial telephone contact with the primary caregiver and confirmed by the latter through completion of a language background questionnaire.

## 2.2. Language sampling

A naturalistic 20-min sample of children's utterances was obtained as they engaged with their main caregivers in free play. A standard set of toys was provided to enhance replication of the sampling context across children. This consisted of a telephone, camera, abacus, stacking cups, cars, baby doll and baby care items, kitchen set, farm animals, tool set, insert puzzles and a pop-up cause-and-effect toy. The range of play materials was purposely chosen to cater for the varying levels of cognitive skill expected in the 12-30-month age range, besides taking children's different toy preferences into account. Play interactions were audio- and video-recorded. Recordings were transcribed orthographically on the basis of the audio-recordings. Video-recordings were viewed when attempting to decipher unintelligible productions. For five (11%) of the language samples, all intelligible word tokens produced spontaneously and imitatively by the children were transcribed independently by a second transcriber, a qualified speech-language pathologist, who was provided with transcription guidelines and field notes taken during direct observation of the adult-child dyads. The mean percentage agreement between transcribed word tokens resulted in an inter-transcriber reliability value of 91.37%.

## 2.3. Caregiver report

Caregivers completed a Maltese-English adaptation of the vocabulary checklist of the first edition of the MacArthur Communicative Development Inventories: Words and Sentences (CDI: WS) (Fenson et al., 1993) for children exposed primarily to Maltese (Gatt, 2010). This consisted of 916 words, organized into 24 semantic categories. The inventory drew on the contents of the original U.S. English version (Fenson et al., 1993), Caselli and Casadio's (1995) Italian adaptation, as well as actual words reported on earlier checklist versions for 12-30-month-old children raised in Maltese-speaking families. Maltese lexical items made up 68.45% of the checklist entries while English words comprised 27.29%. The rest (4.26%) were words that were not clearly identifiable as Maltese or English, such as sound effects, homophones and cognate terms, hence referred to as 'non-specific language words'. The lexical items in semantic categories covering content words were presented in Maltese and/or English according to reported usage during piloting of the checklist adaptation. Here, English words were consistently fewer than Maltese words, with discrepancies varying in size depending on the semantic category. For example, the Animals section contained 28 Maltese words and 22 English words, whereas the Action words category listed 78 words, of which 70 were Maltese (see Gatt et al., 2011). In the function word categories, Maltese and English translation equivalents were available for most semantic concepts. Among the checklist entries, 215 pairs of Maltese and English words corresponded to the same meaning in adult language use. Each semantic category included a recall section in which caregivers could add words in their children's expressive repertoires not listed in the checklist. Gatt et al. (2014) reported vocabulary checklist scores to correlate positively and significantly with the number of word types produced spontaneously during play ($r = 0.635$, $p < 0.001$) and with the number of picture labels elicited through an informal naming task ($r = 0.556$, $p < 0.001$), providing evidence for the checklist adaptation's concurrent validity.

Primary caregivers also completed a language background questionnaire developed purposely for the study. Although the psychometric properties of the questionnaire were not evaluated, the instrument served to document each child's language exposure and confirm the predominant use of Maltese in the home, as claimed by children's caregivers at the recruitment stage. Questions addressed caregivers' Maltese and English proficiency and use, children's relative exposure to Maltese and English on a daily basis and language mixing patterns used with and around each child. Bilingual oral language and literacy skills were reported by 95.5% ($N = 42$) of the caregivers, with the remaining two reporting limited proficiency in English. Just over half of the respondents (52.3%, $N = 23$) reported speaking Maltese more confidently than English, with the rest feeling equally comfortable speaking both languages. Varying degrees of language mixing were reported in the children's language exposure, confirming the likelihood that none of the participants were exposed to monolingual Maltese and monolingual English input. Informal observation of adult language use patterns during play confirmed lexical mixing to be employed by all caregivers, including the two mothers having limited bilingual proficiency. All respondents reported that their children were addressed in Maltese for over 60% of the time on a daily basis.

## 2.4. Procedure

Data were collected in the children's homes over two sessions, one to two weeks apart. During the first session, details of the child's birth history, general health, physical and language development, parental education and occupation, as well as any family history of language impairment, were obtained during an interview with the main caregiver. The questionnaire and vocabulary checklist were then discussed and their completion solicited by the next visit, which took place one to two weeks later. In order to enhance accuracy in vocabulary reporting, each caregiver was briefed about the purpose of the checklist and taken through the bilingual guidelines attached to the tool. It was emphasized that only words used spontaneously by children were to be reported. During the second visit, each child's expressive language was sampled during 20 minutes of free play with the caregiver.

Preliminary transcription of children's vocalizations during free play was attempted as the caregiver-child dyads were observed. A full orthographic transcription of spontaneous and imitated utterances was subsequently carried out on the basis of the audio recordings. Video recordings helped decipher unintelligible productions captured on the audio recordings. The present study focused only on the spontaneous utterances, which amounted to 86% of all transcribed productions. Excerpts of nursery rhymes and songs were not transcribed.

## 2.5. Data coding

In this study, our primary focus was on words having a commonality of 50%+, that is, words produced by at least half the children in the whole cohort and in each age group. These 'more commonly produced words', as documented separately by checklist and sample data, were identified in relation to the age point/s at which they were produced. In each age group, the more common words were classified according to a commonality score ranging between 6 and 11. The score of 5 was only relevant to the 30-month-olds, since it represented 50% of the 10 children in this group. Since the other age groups were slightly larger, counting 12 (18-month-olds) and 11, their lowest commonality score was 6. Similarly, the highest score of 12 was only relevant to the 18-month-olds, with 11 being the maximum score for the 12-, 18-and 24-month-olds and 10 for the 30-month group. Words not reaching the designated commonality threshold were coded for the number of occurrences, but were not analyzed further.

### 2.5.1. Caregiver report measures

Reported words were entered as variables in an item-by-child database. Words ticked on the checklist, as well as recalled words, the additional items contributed by some caregivers in dedicated boxes attached to each semantic category, were all considered as variables. Recalled words were tagged accordingly, to distinguish them from recognized words. Including recalled words in checklist measures enhanced comparability with language sample data as it compensated somewhat for the predetermined number of word recognition opportunities provided by the checklist. All reported words were tallied individually and a *commonality score* was derived for each item. This represented usage across the cohort, in terms of the number of children reported to produce each word. For every child, a *Total Vocabulary (TV) score*, which summed all recognized and recalled words reported on the checklist, was computed. This represented the child's vocabulary size as indexed by the caregiver.

### 2.5.2. Sampled words

Spontaneous vocalizations conveying consistent meanings were identified as words. These comprised forms which approximated adult targets closely and others which showed reduced phonological complexity but still matched at least two consecutive phonemes of the adult target (see Huttenlocher et al., 1991). Meaningful use was established on the basis of children's preceding and subsequent utterances, focus of attention and accompanying gestures. Stability in meaning was determined if a sound-meaning pairing occurred more than once in the same sample and/or was recognized by the caregiver. Productions bearing no resemblance to an adult form, despite consistency in meaning, were assigned word status but were not analyzed in this study, in view of the decreased likelihood of them having counterparts in the checklist data. Productions were classified as unintelligible if three consecutive attempts at deciphering them were unsuccessful. Meaningful interjections (e.g., *ohoh*), sound effects (e.g., *brmmbrmm*) and routine words (e.g., *baħħ* [all gone]) were coded, but fillers (e.g., *emm*) and part-word repetitions were not. Incomplete lexical items were only inputted if the target word was unequivocally obvious. Instances of jargon were not coded. A Number of Different Words (NDW) score, tallying types (the sample-based counterpart to the TV score calculated for the checklist data), and the Total Number of Words (TNW, token count) were computed for every child. In the case of samples manifesting emergent grammar, word components having lexical-semantic meaning were counted as lexical items in their own right. Thus, Maltese enclitic pronouns attached to nouns, verbs and prepositions were coded as separate lexical items (e.g., *xagħri* [*xagħr* + −*i*] (hair + my) = 2 types), even when inflecting for number and/or gender, e.g., with *xagħru* [*xagħr* + −*u*] (hair + his) and *xagħrha* [*xagħr* + −*ha*] (hair + her), both instances were counted as two words, even if they occurred in the same sample. In contrast, word elements having solely grammatical meaning, such as gender and number inflections on verbs and adjectives, were tallied as tokens of the same type. In this vein, demonstrative pronouns inflecting for number and gender were not coded as unique words, e.g., *dan, din* (this m./f.) and *dawn* (these) were coded as *dan/din/dawn* (1 type, 3 tokens). Number words such as 'eight', 'five' and so on were coded collectively as one type, 'one, two, three…'. Word combinations employed invariably as a single lexical item to convey a specific meaning, e.g., *love you,* were coded as one word if the two components did not also appear individually in the same sample. Maltese enclitic pronouns suffixed to the same noun or verb and not appearing elsewhere in the sample, as in the case of the indirect object pronoun *-lu* (to him) and *biddel* (change) in *biddillu* (change (to) him), were also coded as one word. Some children embedded English words in Maltese grammatical constructions, in which case coding conventions described here were applied as necessary, e.g., *ball dak… dik pupa* (that (m.) (is) a ball… that (f.) (is) a doll) was coded as *ball*, *dak/dik*, *pupa* (NDW = 3; TNW = 4).

Sampled words that matched any of the 916 words on the vocabulary checklist were tagged according to the semantic categories on the latter. For example, identification of 'car' in a sample led to it being tagged as 'Vehicles – Real or Toy'. Sampled words that were not available for reporting on the checklist were assigned to one of the 24 semantic categories deemed to be most suited to the word's semantic function and tagged as <Specific category name>_Other'. Thus, spontaneous production of 'bicycle', a lexical item not on the vocabulary checklist, was identified as 'Vehicles – Real or Toy_Other'. For each child, identified types were assigned a score of 1 every time they were produced, enabling computation of the *frequency of occurrence* in the specific age group and across the composite sample. The *commonality score* of each word represented the number of children using it at least once.

### 2.5.3. Coding for grammatical categories

All inputted words were coded as either social words (sound effects, routine words), nouns, verbs, adjectives or function words, the latter including adverbs, pronouns, question words, prepositions, articles and quantifiers, negative markers and conjunctions. The auxiliary verbs *qed*, *qiegħed*, and *qiegħda*, together with the future particles *ħa* and *se*, were coded as function words. The relevant analytical framework was adopted from Caselli et al.'s (1999) study of grammatical categories in English and Italian and, accordingly, considered as a simplified rendition of the linguistic input received by children.

### 2.5.4. Coding for language

All items were also coded as either Maltese, English or non-specific language words, drawing on the language contact phenomena expected in Maltese children's input (see Section 2.1). Items tagged as Maltese were native Maltese words, as well as established English borrowings that had no Maltese equivalent, e.g., *kompjuter* (computer), 'hello'. The remaining English words were coded as English items since they were preferred to the Maltese equivalent. Examples included 'colours' instead of *kuluri* and 'thank you' for *grazzi*. Examples of non-specific language words included BANANA, BLU and BASKET.

## 3. Results

Aggregate vocabulary scores derived from checklist and sample data were examined in a preliminary analysis. Word-level data from each source were then examined separately, followed by comparative analyses. Throughout, the focus was on the words employed by 50% or more of the participants.

Table 1 shows descriptive information for Total Vocabulary (TV), Number of Different Words (NDW) and Total Number of Words (TNW) scores, each of which measured the full range of words produced, regardless of commonality. As expected, all sum and mean scores increased with age. TV values were consistently larger than sampled vocabulary scores, with Wilcoxon signed-rank tests showing mean differences to be significant (TV and NDW: $Z = -5.78$, $p < 0.001$, TV and TNW: $Z = -3.93$, $p = < 0.001$). On average, the TNW score, measuring all spontaneous occurrences of sampled words, was also significantly larger than NDW, which tallied the unique words ($Z = -5.30$, $p < 0.001$). Yet, partial correlations between TV and NDW, the aggregate scores quantifying reported and sampled vocabulary size respectively, yielded moderately positive and significant relationships ($r = 0.626$, $p < 0.001$) when age was partialled out. A close correspondence between checklist and sample measures was therefore present, with children's scores ranking similarly across both methods despite the numerical differences. The coefficient between TV and TNW counts was also significant but lower than for NDW ($r = 0.454$, $p < 0.05$). This could be because TNW scores were not a direct counterpart to TV since they incorporated all occurrences of words produced. These introductory results set the scene for the subsequent item-based analyses, establishing that the starting point for these was a statistical correspondence between vocabulary size as gauged by both methods.

## 3.1. Word usage as reported by caregivers

Table 2 lists the 43 words produced by at least 50% of the 44 participants, according to their caregivers. Commonality, expressed in terms of absolute and proportion scores, had an upper ceiling of 43 (97.7%). Words were spread across 12 semantic categories, with the most commonly employed being Daily experiences ($N = 15$), People ($N = 8$) and Sounds ($N = 6$). The five words reported for at least 75% of the cohort represented People and Daily experiences. The words MAMÀ/MUMMY, PAPÀ/DADDY and *nanna* (grandma), all belonging to the People category, were the most commonly used overall. Additional words recalled by caregivers were generally reported for only small numbers of children, explaining why none appeared in the list of most commonly reported words for all the participants. The most common recalled items were 'cereal' and 'medicine', each produced by 11.4% ($N = 5$) of the sample, while 'bread', 'eyes', 'good night', 'sorry' and *tersaq* (move) were reported for four children (9.1%). A total of 64 words, spanning 10 semantic categories, were never reported. Of these, 14 (21.9%) were Maltese and 50 (78.1%) were English, with three words in each language being equivalents. All the English items were function words. The latter totaled 60 (93.8%) of the unreported words and were accompanied by three nouns and one adjective.

In the analysis by age group, words that reached a commonality of 50% and over totaled 617. Supplementary Table S1A lists the words more commonly reported by caregivers when a lower commonality threshold of 50% was applied. Word usage increased with age, so that

**TABLE 1** Mean, standard deviation (SD), and sum of scores for Total Vocabulary (TV), Number of Different Words (NDW), and Total Number of Words (TNW) in relation to age.

| Age (months) | *N* | TV | | | NDW | | | TNW | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Sum | Mean | SD | Sum | Mean | SD | Sum | Mean | SD |
| 12 | 11 | 229.00 | 20.82 | 23.66 | 31 | 2.82 | 4.29 | 89 | 8.09 | 13.10 |
| 18 | 12 | 816.00 | 68.00 | 73.52 | 163 | 13.58 | 11.22 | 585 | 48.75 | 46.19 |
| 24 | 11 | 2473.00 | 224.82 | 161.74 | 315 | 28.64 | 22.65 | 1,038 | 94.36 | 79.42 |
| 30 | 10 | 4330.00 | 433.00 | 163.87 | 943 | 94.30 | 23.85 | 3,359 | 335.90 | 110.39 |

TABLE 2 The more commonly reported words across participants ($N = 44$, age range = 12–30 months) with respective checklist semantic category and raw (%) commonality score.

| Reported word | Semantic category | Commonality (%) | Reported word | Semantic category | Commonality (%) |
|---|---|---|---|---|---|
| mamà, mummy | People | 43 (97.7) | shoes | Clothing | 25 (56.8) |
| papà, daddy | People | 41 (93.2) | amm amm | Sounds | 25 (56.8) |
| *nanna* (grandma) | People | 39 (88.6) | *taqa'* (fall) | Action words | 24 (54.5) |
| bye, ciao, tatà | Daily experiences | 35 (79.5) | *qalbi* ('my heart') | Daily experiences | 24 (54.5) |
| *aħħ* (ouch) | Daily experiences | 33 (75.0) | bird | Animals | 24 (54.5) |
| one, two, three… | Daily experiences | 32 (72.7) | fish | Animals | 24 (54.5) |
| *bumm* <loud sound> | Sounds | 31 (70.5) | *mimmi* ('pain') | Daily experiences | 24 (54.5) |
| thank you | Daily experiences | 30 (68.2) | *baħħ* (all gone) | Daily experiences | 24 (54.5) |
| book | Everyday objects | 30 (68.2) | boy | People | 23 (52.3) |
| ball | Toys and games | 30 (68.2) | please | Daily experiences | 23 (52.3) |
| wuw wuw <dog sound> | Sounds | 30 (68.2) | *le* (no) | Daily experiences | 23 (52.3) |
| no | Daily experiences | 29 (65.9) | muu <cow sound> | Sounds | 23 (52.3) |
| kokò | Daily experiences | 28 (63.6) | mmm | Sounds | 23 (52.3) |
| *nannu* (grandpa) | People | 28 (63.6) | *bumma* | Daily experiences | 23 (52.3) |
| *(piff) jaqq* <expression indicating disgust> | Daily experiences | 28 (63.6) | *xita* (rain) | Outside | 22 (50.0) |
| pipì | Daily experiences | 27 (61.4) | fish | Food and drink | 22 (50.0) |
| car | Vehicles | 27 (61.4) | banana | Food and drink | 22 (50.0) |
| miao | Sounds | 27 (61.4) | child's name | People | 22 (50.0) |
| dog | Animals | 26 (59.1) | *bravu, brava* (good m., f.) | Descriptive words | 22 (50.0) |
| baby | People | 26 (59.1) | *dudu* (worm) | Animals | 22 (50.0) |
| hello | Daily experiences | 26 (59.1) | kiss | Action words | 22 (50.0) |
| pet's name | People | 26 (59.1) | | | |

Commonality threshold = 22 (50.0%). Italics denote *Maltese* words, small capitals denote non-specific language words.

older children added more items to their frequently-used repertoire. As expected, more words were present at the lower end of the commonality range (50%+), with numbers tapering off at higher commonalities. Relatively few words were produced by all children in each age group.

For every age group, the more commonly reported words were analyzed in relation to the grammatical categories and languages they represented. Figure 1 shows the distribution of social words, nouns, verbs, adjectives and function words across the more commonly reported checklist words, while Figure 2 inspects the grammatical category trends more closely, zooming in on the different levels of commonality embraced within every category for each age point. Across all commonality levels, social words increased considerably between 12 and 18 months but only increased marginally at subsequent age points. Nouns increased exponentially throughout each phase between 12 and 30 months. In contrast, verbs, adjectives and function words were absent at 12 and 18 months but increased exponentially from 24 to 30 months. The distribution of languages across the more commonly reported words for each age point is shown in Figure 3, with Figure 4 showing the Maltese, English and non-specific language words for each commonality level. Commonalities of Maltese and English words were comparable at 18 and 24 months. At 30 months, the number of Maltese words among those most commonly reported increased drastically.

## 3.2. Sampled word usage

Our next analysis addressed the commonality and frequency of words sampled during free play. At 12 months, no words reached the commonality threshold of 50%+, reflecting the limited production of words among the youngest participants. Words produced by >50% of the participant group ($N = 22$) were limited to two (see Table 3). The words *dan, din* [this (m., f.)] and *dawn* (these), coded as a single type, were produced by 63.6% ($N = 28$) of the participants. This item also placed highest in terms of frequency of production across all word tokens by all children (8.5%), which is understandable given that, in effect, the frequencies of three tokens were collapsed into a single frequency score. Mamà, mummy appeared at least once in the language samples of 21 children, ranking in fifth place overall in terms of frequency (2.4%). Other frequently produced words were *hawn, hawnhekk* (here) (3.8%), *il-, l-* and other definite articles (3.2%), together with *iva, eħe* (yes) (2.6%). Table 4 lists the 10 most frequently occurring words across all samples, in order of descending frequency. Generally, the more frequently produced words were also the more

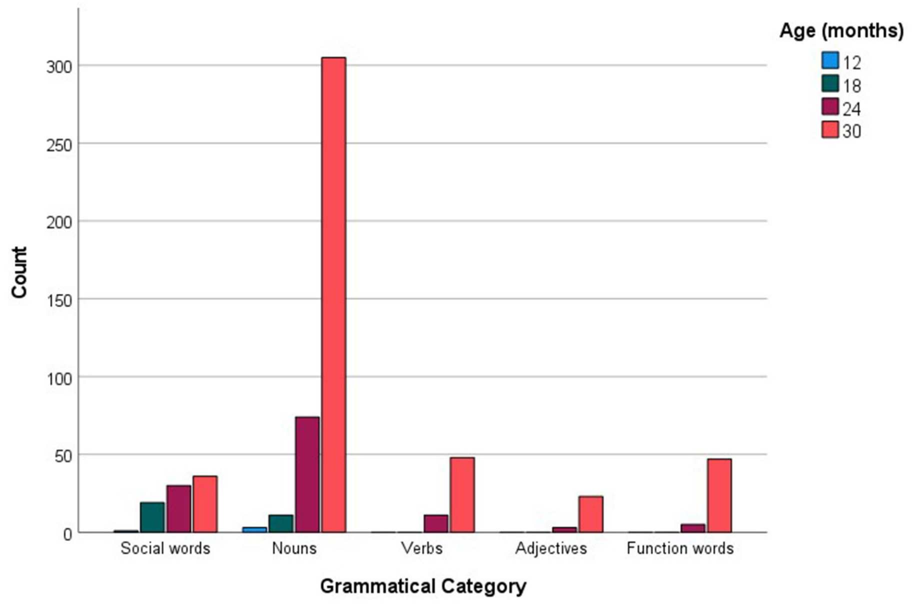


FIGURE 1
Grammatical category counts (Social words, Nouns, Verbs, Adjectives, Function words) as a function of more commonly reported words at 12, 18, 24, and 30 months.

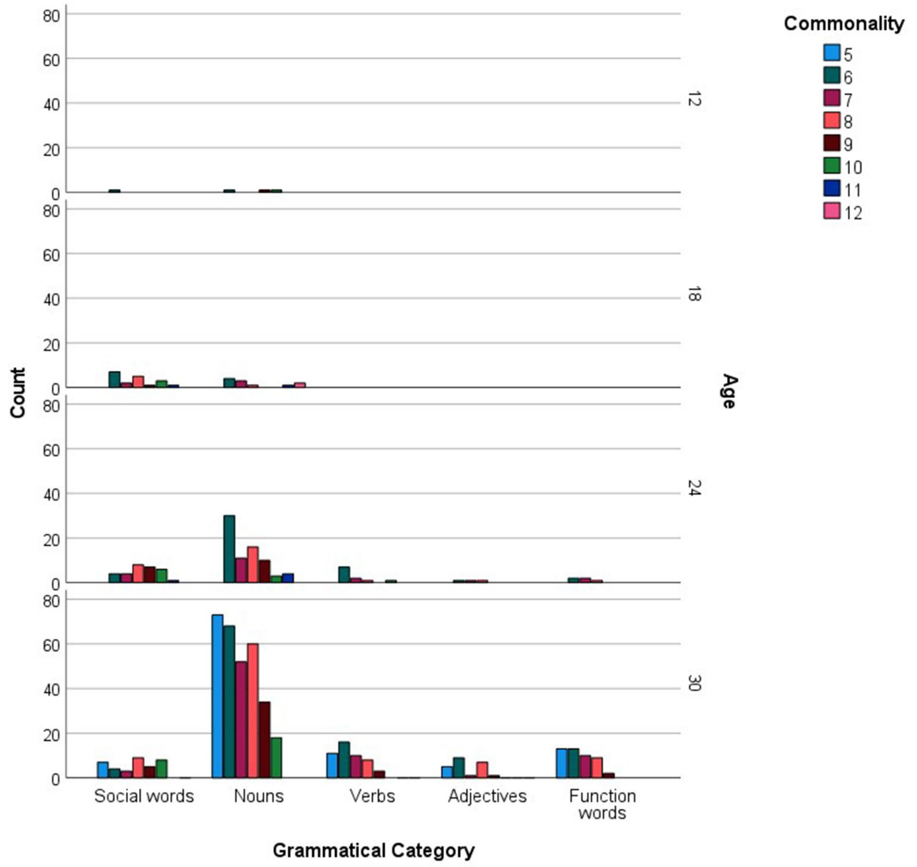


FIGURE 2
Grammatical category counts (Social words, Nouns, Verbs, Adjectives, Function words) as a function of more commonly reported words, including commonality levels (i.e., numbers of children using each word), at 12, 18, 24, and 30 months.

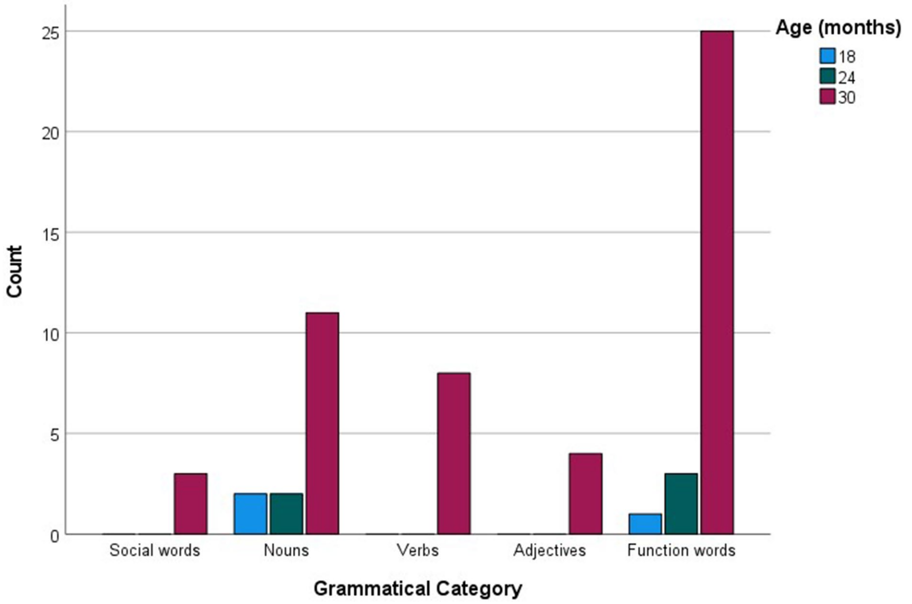


FIGURE 3
Grammatical category counts (Social words, Nouns, Verbs, Adjectives, Function words) as a function of more commonly sampled words at 12, 18, 24, and 30 months.

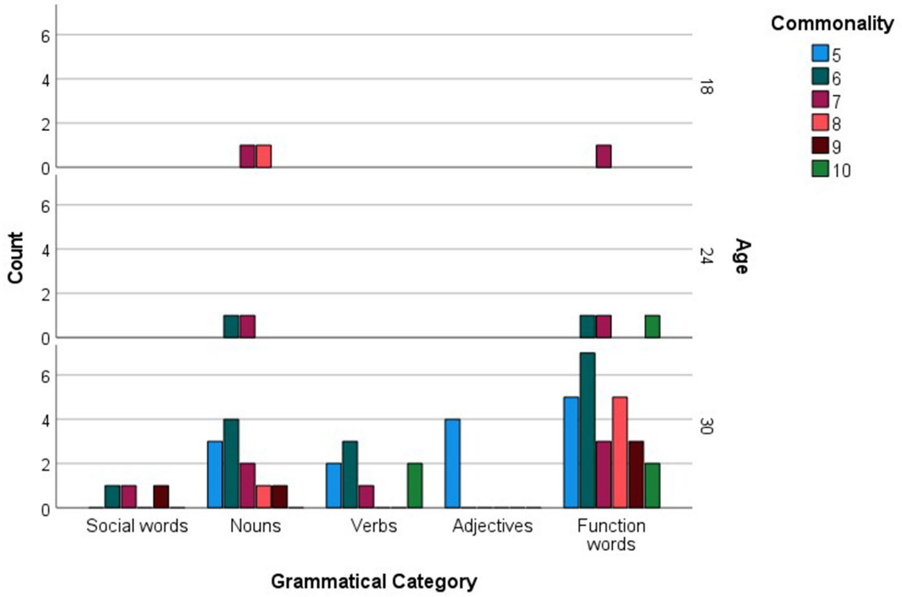


FIGURE 4
Grammatical category counts (Social words, Nouns, Verbs, Adjectives, Function words) as a function of more commonly sampled words, including commonality levels (i.e., numbers of children using each word), at 12, 18, 24, and 30 months.

common, although only *dan, din, dawn* surpassed our commonality threshold. The word *fejn* (where) appeared to be an outlier, ranking 7th in composite frequency with a commonality of 8.

When analyzed in relation to age, commonly produced words were much fewer in number compared to checklist data, amounting to 58 (see Supplementary Table S1B). For each age group, the more common words were then examined as a function of their grammatical categories (see Figure 5). Social words, verbs and adjectives were among the more commonly sampled words only at 30 months, while nouns were among the more commonly produced words at 18 and 24 months, reaching maximum commonality at 30 months. The most prominent trend, however, is a sheer increase in commonly produced function words at 30 months. Figure 6 presents the breakdown of commonality scores constituting the trends illustrated in Figure 5. When analyzed in relation to language, the more commonly sampled words were somewhat balanced across Maltese and English for the

TABLE 3 The more commonly sampled words across participants ($N = 44$, age range = 12–30 months) with assigned semantic category, raw (%) commonality score and composite frequency, expressed as a proportion of the composite Total Number of Words (i.e., 5,071) across all age groups.

| Sampled word | Semantic category | Commonality (%) | % Composite frequency |
|---|---|---|---|
| *dan, din, dawn* (this (m., f.), these) | Pronouns | 28 (63.6) | 8.5 |
| MAMÀ, MUMMY | People | 21 (47.7) | 2.4 |

Commonality threshold = 22 (50.0%). Italics denote *Maltese* words, small capitals denote NON-SPECIFIC LANGUAGE words.

TABLE 4 The more frequently sampled words across participants ($N = 44$, age range = 12–30 months) with assigned semantic category, composite frequency, expressed as a proportion of the composite Total Number of Words across all age groups (i.e., 5,071), and raw (%) commonality score.

| Sampled word | Semantic category | % Composite frequency | Commonality (%) |
|---|---|---|---|
| *dan, din, dawn* (this (m., f.), these) | Pronouns | 8.5 | 28 (63.6) |
| *hawn, hawnhekk* (here) | Prepositions and locations | 3.8 | 20 (45.5) |
| *il-, l-…* (def. Art.) | Quantifiers | 3.2 | 11 (25.0) |
| *iva, ehe* (yes) | Daily experiences | 2.6 | 16 (36.4) |
| MAMÀ, MUMMY | People | 2.4 | 21 (47.7) |
| *dak, dik, dawk* (that (m., f.), those) | Pronouns | 1.9 | 17 (38.6) |
| *fejn* (where) | Question words | 1.8 | 8 (18.2) |
| one, two, three… | Daily experiences | 1.8 | 14 (31.8) |
| *tara* (see) | Action words | 1.8 | 15 (34.1) |
| hello | Daily experiences | 1.7 | 13 (29.5) |

Italics denote *Maltese* words, small capitals denote NON-SPECIFIC LANGUAGE words.

18-and 24-month-olds (see Figures 7, 8). At 30 months, Maltese words were clearly predominant, while the relative increment in English words was more protracted.

## 3.3. Comparison of item-level results for checklist and sample data

Our final analysis compared main trends emerging in the checklist and sample datasets. Of the more commonly sampled words, 28 items had a matching counterpart in the checklist dataset. These items, together with age and commonality of occurrence, are listed in Table 5. While the number of item-level matches is extremely limited, particularly in comparison to the expansive checklist figures, correspondences are remarkably close for 30-month words, even in terms of commonality score. In terms of grammatical composition of the more commonly produced words, Figures 1, 3 revealed elements of a similar shift in emphasis across methods, although the captured stages are different. Notably, the sample data portray the full wave of reorganization from social words to nouns, verbs and adjectives and subsequently function words in the 30-month profiles (see Figure 3). In the checklist data, on the other hand, 24-and 30-month more commonly produced words seem to be at the point of shifting to a predominance of function words (Figure 1). When partialling out the effects of age, a significant albeit low correlation resulted ($r = 0.293$, $p < 0.001$), indicating similar trends overall, despite the differences in numerical values. In comparison, caregiver report and language sampling were more compliant in profiling the distribution of languages among the more commonly produced words. Both methods identified comparable numbers of English and Maltese words among the lexical items more commonly produced by 18-and 24-month-olds, as well as a predominance of Maltese words at 30 months (Figures 5, 7). A partial correlational analysis that controlled for age effects revealed a remarkably high and significant coefficient ($r = 0.935$, $p < 0.001$).

# 4. Discussion

The current study set out to identify word usage trends in a cohort of 12-30-month-olds who were predominantly exposed to Maltese in their homes, within a broader context of societal bilingualism. Its purpose was to compare word-level data identified through caregiver report and language sampling employed with the same children, in order to derive methodological implications that could guide theoretical understanding, as well as reporting instrument revision. To our knowledge, the comparison of word usage as determined by different methods for the same children is as yet unprecedented. The present investigation therefore attempted to add fresh insight to the long-debated issue of methodological bias in the measurement of young children's early vocabularies. While word-level analyses have long contributed to and complemented the substantial research base relating to vocabulary acquisition, their relevance might at times be overlooked because of the scrutiny of individual items required for this purpose. Moreover, a direct comparison of individual items sampled directly and reported by parents contributes important theoretical and methodological insights. Specifically, by focusing on more commonly produced words in Maltese children, this study not only documented trends and patterns in word usage in an under-researched language pair, but also attempted to add depth to current views on the effectiveness of the parental report method.

The first set of preliminary findings showed significant differences between mean scores tallying reported and sampled vocabulary size, but also positive and significant correlations between them. Total Vocabulary (TV) and Number of Different Words (NDW) scores both addressed the range of words used, albeit through different sources. Thus, it was relatively unsurprising that the more commonly reported and sampled words, different in nature from TV and NDW scores but directly related, correlated positively and significantly too, despite their different numerical bases. While the nature of the data collection method inevitably impacts the numbers and range of words picked up, it is encouraging that, in terms of more common words across participants, similarities were documented. The correlational analyses in particular indicate that in general, the caregivers were able to report

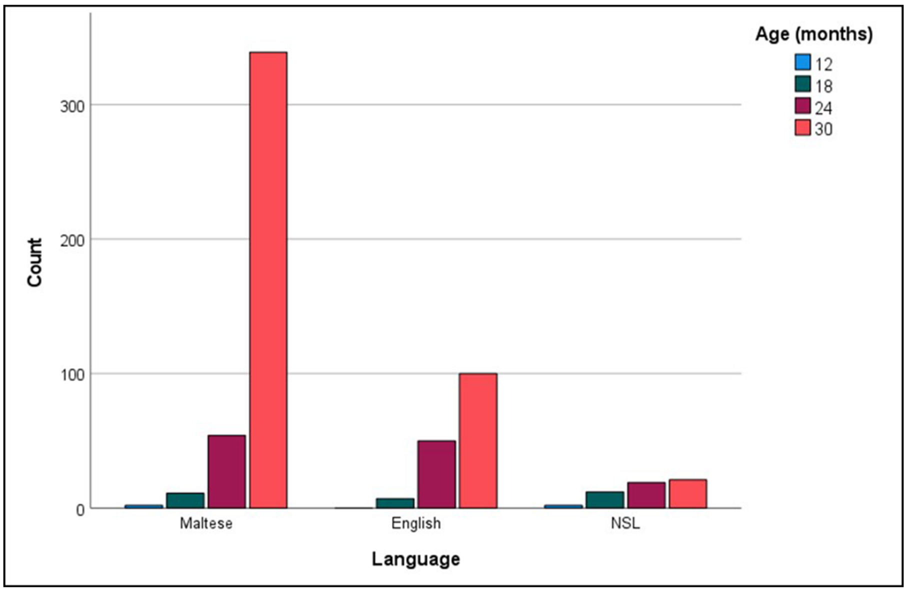


FIGURE 5
Language counts (Maltese, English and non-specific language (NSL) words) as a function of more commonly reported words at 12, 18, 24, and 30 months.

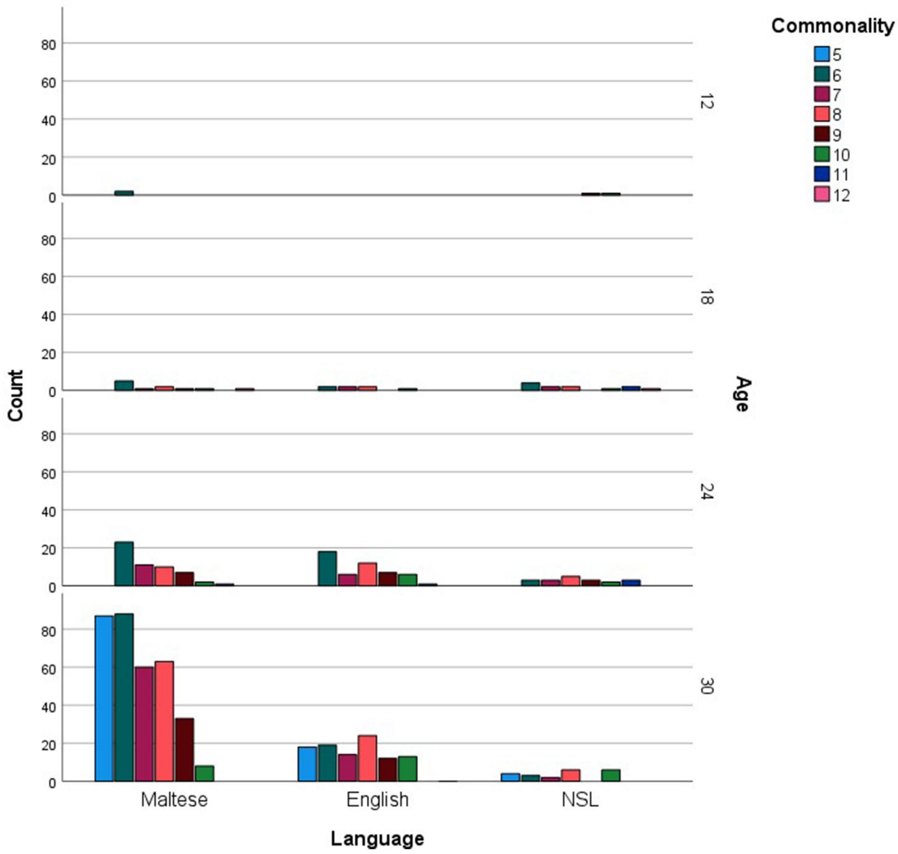


FIGURE 6
Language counts (Maltese, English and non-specific language (NSL) words) as a function of more commonly reported words, including commonality levels (i.e., numbers of children using each word), at 12, 18, 24, and 30 months.

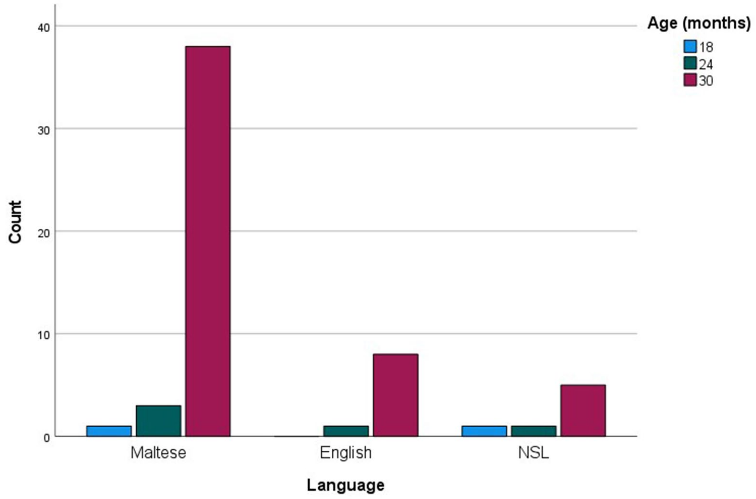


FIGURE 7
Language counts (Maltese, English and non-specific language (NSL) words) as a function of more commonly sampled words at 12, 18, 24, and 30 months.

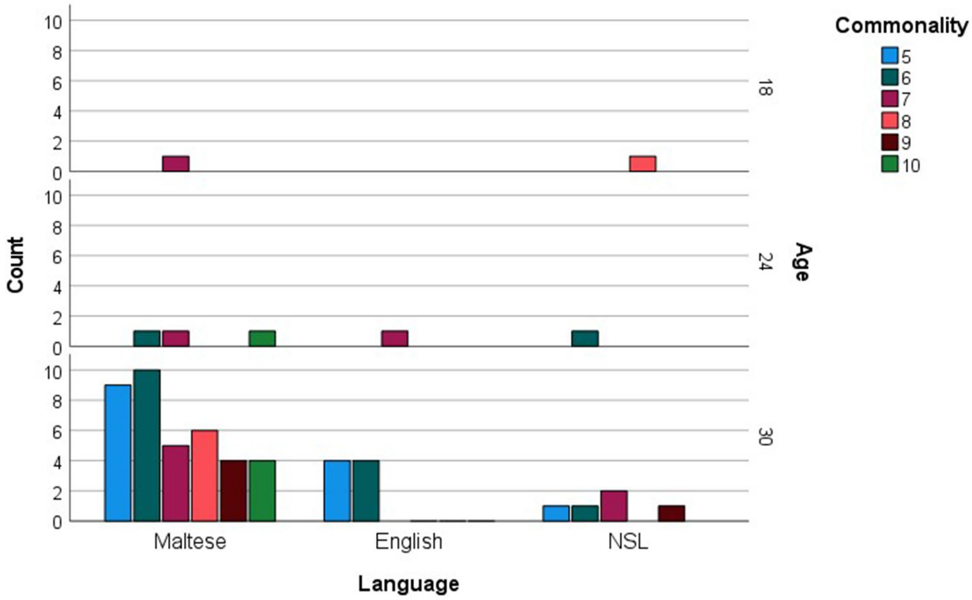


FIGURE 8
Language counts (Maltese, English and non-specific language (NSL) words) as a function of more commonly sampled words, including commonality levels (i.e., numbers of children using each word), at 12, 18, 24, and 30 months.

on usage of grammatical categories and languages in ways that ranked similarly to sampled trends.

The clear predominance of nouns in the checklist profiles reported here was somewhat predicted. Published core vocabulary lists have been compared to the CDI vocabulary checklist words, the latter taken as a representation of the words young typically-developing language learners are expected to use. Parallel studies by Frick Semmler et al. (2023) and Laubscher and Light (2020) both flagged a mismatch between the contents of core word lists and vocabulary checklists. The relative predominance of function words and scarcity of content words typical of core word lists conflicted with the distribution of grammatical categories expected in early vocabularies. The results we obtained for commonly reported words also suggest a noun bias, although this may be more related to caregiver reporting style than to the reporting tool itself. This is hypothesized on the basis of the bilingual vocabulary checklist employed in this study. The instrument employed had quantitatively similar noun and function word proportions since the function word categories generally presented items in both languages, effectively doubling the number of reporting opportunities, whereas noun categories contained much fewer equivalent terms. Therefore, while a noun bias is clearly evident in the reported data, a component of it

**TABLE 5** Matching words across checklist and sample datasets, with age of each occurrence and commonality score for each.

| Word | Sample occurrence (age, commonality) | Checklist occurrence (age, commonality) |
|---|---|---|
| *dan, din, dawn* (this (m., f.), these) | 24, 10<br>30, 10<br>18, 7 | 24, 8<br>30, 8 |
| *hawn, hawnhekk* (here) | 30, 10<br>24, 7 | 30, 7<br>24, 7 |
| *taqa'* (fall) | 30, 10 | 24, 10 |
| *tara* (see) | 30, 10 | 30, 5 |
| *iva, eħe* (yes) | 30, 9 | 24, 7<br>30, 8 |
| *dak, dik, dawk* (that (m., f.), those) | 30, 9<br>24, 6 | 30, 10 |
| *telefon* | 30, 8 | 30, 9 |
| *ħa, se* (fut. particle) | 30, 8 | 30, 6 |
| *jiena* (I) | 30, 8 | 30, 10 |
| *hekk* (so) | 30, 8 | 24, 5 |
| MAMÀ, MUMMY | 18, 8<br>30, 7 | 12, 11<br>18, 12<br>24, 11<br>30, 10 |
| *le* (no) | 30, 7 | 24, 9<br>30, 9 |
| PAPÀ, DADDY | 30, 7 | 12, 9<br>18, 11<br>24, 11<br>30, 10 |
| *tiġi* (come) | 30, 6 | 30, 9 |
| *one, two…* | 30, 6 | 18, 10<br>24, 10 |
| car | 30, 6 | 24, 11<br>30, 9 |
| toy | 30, 6 | 24, 9<br>30, 8 |
| *fejn* (where) | 30, 6 | 30, 10 |
| *iċċempel* | 30, 6 | 30, 7 |
| elephant | 30, 6 | 30, 7 |
| *x', xiex* (what) | 30, 6 | 30, 5 |
| *tajjeb, tajba* (good) | 30, 5 | 30, 8 |
| *għax* (because)_ | 30, 5 | 30, 6 |
| *mela* (so) | 30, 5 | 30, 5 |
| green | 30, 5 | 30, 5 |
| *għandek* (have) | 30, 5 | 30, 6 |
| BLU | 30, 5 | 30, 6 |
| te' | 30, 5 | 30, 9 |

Italics denote *Maltese* words, small capitals denote NON-SPECIFIC LANGUAGE words.

could have stemmed directly from the caregivers' filtering of reported information (Stiles, 1994). The contrasting predominance of function words manifested in the common words sampled adds further weight to this possibility.

The predominance of Maltese words in both checklist and sample datasets is perhaps one of the more interesting findings of this study. When transposed onto the grammatical category analysis, the inclination to produce more commonly occurring words in Maltese indicates that this is the language in which most of the nouns favored in the checklist dataset and the function words dominating the sample scores were produced. This implies that, for these children, Maltese was the more consistently employed language, regardless of the grammatical categories employed more. Based on the premise that Maltese children's word production reflects the linguistic input received, including the contact phenomena and language choices made by native Maltese-English bilinguals, documentation of word usage also has implications for young Maltese children's language milieu. For instance, this study's finding of a consistent English language presence among children's more commonly used words, accompanied by a substantial Maltese component, potentially reflects the relative salience of both languages in children's input, while possibly also mirroring societal trends in the language choices incorporated in child-directed language use.

The commonality analyses reported here emerged as a valid and efficient means of uncovering trends in vocabulary acquisition. The more commonly produced words appeared to condense trends and trajectories in word production typically identified across children's full range of expressive vocabularies. Specifically, the developmental reorganization of grammatical categories in the participants' more commonly used words tended to replicate findings reported in larger studies of vocabulary acquisition. The more commonly produced words across children might be seen as a method-specific 'core' central to the acquisition of a particular language or language pair.

The merits of parental report are widely recognized and often seen to exceed its pitfalls. Parents' observations of their children's emergent language, across daily settings and with various interlocutors, enable them to report comprehensively on their children's expressive vocabulary skills (Frank et al., 2021). When parents complete vocabulary checklists on the basis of their children's daily word production, they not only provide researchers and clinicians with a reliable and valid estimate of their children's language skills but they also contribute to a wider corpus that might be used for reference or norming purposes. Consideration of word usage adds depth to parent-based vocabulary measures. Examining how many children use specific checklist words sheds light on the relevance of these items in sensitively gauging various levels of vocabulary ability. In the study reported here, direct comparison between the more commonly reported words and those sampled naturalistically yielded objective indications as to which checklist words were more likely to resonate with reporting caregivers, compared to sampled commonality, and which items were well beyond the upper developmental limit of the target age range, as in the items that were never reported or sampled.

## 4.1. Limitations and recommendations

There are various limitations in this study that need to be acknowledged. Although typical development was a criterion for participant selection, it cannot be excluded that some participants may have been presenting with subtle language difficulties that were unidentified at the time of data collection. The small sample size, largely determined by the labor-intensive methodology, inevitably limited statistical power in analyses. Although additional dual-method data, collected in a longitudinal arm of this study, were available for few other 24-and 30-month-olds (two separate longitudinal cohorts, $N=9$ and 7 respectively), their addition to the present cohort would have unbalanced the close similarity across age groups in numbers of participants and gender distribution. In retrospect, however, it would have been useful to top up the 30-month-group by one or even two participants, partly due to it being the smallest age group and also because this age point revealed an intense word usage dynamic that would have benefited from more extensive investigation. In addition, the computation of proportion scores for grammatical and language components would have enabled a more equitable comparison across methods than raw scores. Moreover, choice of the 50% + usage threshold was somewhat arbitrary, although partly influenced by Fenson et al.'s (1994) consideration of this level in their individual item analyses of the CDI norming dataset. Our intention was to pitch word-level analysis at a level that would not favor the more linguistically advanced participants. On the other hand, data at lower levels of usage, such as 25–49%, would have enabled more fine-grained insight on the levels of difficulty and psychometric sensitivity of a broader range of words. This is particularly relevant since the study of early language acquisition in Maltese children is still in its infancy, with no previous evidence documenting the specific words appearing earlier and later in typical development. Analysis of more commonly used words as a function of gender and vocabulary size level would have also added depth to the current findings but could be considered as a possible avenue for further research. For instance, Weber et al. (2018) employed item response models to investigate the probability of items on the newly-adapted Wolof version of the CDI vocabulary checklist eliciting responses in relation to their difficulty level, as well as child characteristics such as gender and level of vocabulary ability.

## 4.2. Conclusion

The present study hopes to contribute toward bridging a conspicuous gap in the research literature. It compares word-level measures obtained in parallel for the same children using two methods, parent report and language sampling, with a focus on an under-researched language pair. When analyzed as a function of age, the more commonly reported words were noticeably more numerous than those sampled, with a relatively limited number of item-level matches. Nonetheless, when the more commonly produced words identified through both methods were analyzed in terms of the grammatical categories and languages they represented, positive and significant correlations resulted. The shifting distributions of grammatical categories in the words more commonly sampled and reported were similar. Even more striking was the resemblance in language profiles documented at 18, 24 and 30 months by both methods. In spite of its unprecedented examination of word usage trends documented by caregiver report and language sampling employed in parallel, the present study has only scratched the veritable tip of the iceberg – it draws on a modest sample of children, with a focus restricted to the words used by 50% and over of participants at four age levels. While breaking into previously uncharted territory, this study clearly flags a need for research that investigates larger samples of children with a denser distribution of age points. The resource demands of transcription and analysis of language sample data inevitably limit the numbers of children from whom naturalistic data are collected. Nevertheless, it is still recommended that a sub-sample of children in any parent-report based study are assessed directly, not only to increase the robustness of the methodological design but also to add important insights based on researchers' 'lived experience' in working with authentic and naturally-occurring data. Moreover, the present investigation showed sampled word usage to be well-honed in capturing shifting distributions of grammatical categories and languages despite its characteristically small numbers, further highlighting its methodological relevance.

An additional purpose of this study was to derive guidelines for objectively revising the Maltese-English CDI adaptation. The word commonality findings presented here can assist in prioritizing items to retain and eliminate from the current version, while prompting reflection on caregivers' predisposition toward reporting nouns. Moreover, the current instrument's language bias, deriving from it being originally adapted for children raised in Maltese-dominant homes, does not make it sufficiently comprehensive for English-dominant children and relatively balanced bilinguals in the Maltese childhood population. Striking a balance between Maltese and English components in the revised vocabulary checklist is called for.

Finally, the present findings can also inform functional vocabulary targets for young children with language difficulties. Although available research evidence related to the developmental trajectory of grammatical categories is relevant to speech and language intervention, word usage data add insight on the words that are likely to be employed in child-directed language and those likely to be picked up and internalized by young Maltese children. These words might well be relevant targets for children who are struggling with their acquisition of Maltese and English.

## Data availability statement

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

## Ethics statement

This study involved human participants and was approved and reviewed by the University of Malta Research Ethics Committee. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

## Author contributions

DG conceptualized and designed the study, wrote the first draft of the manuscript, and revised the manuscript. CG organized the database. LC performed the statistical analysis. LC and CG each wrote a section of the manuscript. All authors read and approved the submitted version.

## Funding

This work was partly supported by University of Malta Research Seed Grant 2023 (Ref. CMTRP01-23) awarded to DG.

## Acknowledgments

We are indebted to the children and their caregivers who generously offered their time to participate. We are grateful to May Agius for interesting discussions on the notion of core vocabulary in the field of Augmentative and Alternative Communication (AAC). Sincere appreciation goes to Philip Dale for his insightful comments and encouraging feedback on an earlier conceptualisation of this work.

## Conflict of interest

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

## Publisher's note



## Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2023.1214518/full#supplementary-material

## References

Alcock, K. J., Rimba, K., Holding, P., Kitsao-Wekulo, P., Abubakar, A., and Newton, C. R. (2015). Developmental inventories using illiterate parents as informants: communicative development inventory (CDI) adaptation for two Kenyan languages. *J. Child Lang.* 42, 763–785. doi: 10.1017/S0305000914000403

Banajee, M., Dicarlo, C., and Buras Stricklin, S. (2003). Core vocabulary determination for toddlers. *Core Vocab. Determin. Toddlers* 19, 67–73. doi: 10.1080/0743461031000112034

Bassano, D. (2000). Early development of nouns and verbs in French: exploring the interface between the lexicon and grammar. *J. Child Lang.* 27, 521–559. doi: 10.1017/S0305000900004396

Bates, E., Bretherton, I., and Snyder, L. (1988) *From first words to grammar: Individual differences and dissociable mechanisms*. Cambridge: Cambridge University Press.

Bates, E., Marchman, V., Thal, D., Fenson, L., Dale, P., Reznick, J. S., et al. (1994). Developmental and stylistic variation in the composition of early vocabulary. *J. Child Lang.* 21, 85–123. doi: 10.1017/S0305000900008680

Beukelman, D., Jones, R., and Rowan, M. (1989). Frequency of word usage by nondisabled peers in integrated preschool classrooms. *Augment. Altern. Commun.* 5, 243–248. doi: 10.1080/07434618912331275296

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., and Greenhalgh, T.the CATALISE consortium (2016). CATALISE: a multinational and multidisciplinary Delphi consensus study: identifying language impairments in children. *PLoS One* 11:e0158753. doi: 10.1371/journal.pone.0158753

Bleses, D., Vach, W., Slott, M., Wehberg, S., Thomsen, P., Madsen, T. O., et al. (2008). Early vocabulary development in Danish and other languages: a CDI-based comparison. *J. Child Lang.* 35, 619–650. doi: 10.1017/S0305000908008714

Borg, A. (1988) *Ilsienna: Studju Grammatikali*. Malta: Penprint Ltd.

Borg, A., and Azzopardi-Alexander, M. (1997) *Maltese*. London: Routledge.

Brincat, J. M. (2011). *Maltese and other languages: A linguistic history of Malta*. Venera: Midsea Books.

Camilleri, A. (1995) *Bilingualism in education: the Maltese experience*. Heidelberg: Julius Groos Verlag.

Camilleri Grima, A. (2013). A select review of bilingualism in education in Malta. *Int. J. Biling. Educ. Biling.* 16, 553–569. doi: 10.1080/13670050.2012.716813

Caselli, M. C., and Casadio, P. (1995). *Il Primo Vocabolario del Bambino: Scheda "Parole e Frasi"* Consiglio Nazionale delle Ricerche Istituto di Psicologia e Fondazione MacArthur.

Caselli, M., Casadio, P., and Bates, E. (1999). A comparison of the transition from first words to grammar in English and Italian. *J. Child. Lang.* 26, 69–111. doi: 10.1017/S0305000998003687

Dale, P. S. (1991). The validity of a parent report measure of vocabulary and syntax at 24months. *J. Speech Hear. Res.* 34, 565–571. doi: 10.1044/jshr.3403.565

Dale, P. S. (1996). "Parent report assessment of language and communication" in *Assessment of communication and language*. eds. K. N. Cole, P. S. Dale and D. J. Thal, vol. *6* (Baltimore, MD: Paul H. Brookes Publishing Co.), 161–182.

Dale, P. S., and Penfold, M. (2011). Adaptations of the MacArthur–Bates CDI into non-U.S. English languages. Available at: http://mb-cdi.stanford.edu/documents/AdaptationsSurvey7-5-11Web.pdf (Accessed February 20, 2023).

De Anda, S., Cycyk, L. M., Moore, H., Huerta, L., Larson, A. L., and King, M. (2022). Psychometric properties of the English-Spanish vocabulary inventory in toddlers with and without early language delay. *J. Speech Lang. Hear. Res.* 65, 672–691. doi: 10.1044/2021_JSLHR-21-00240

Ellis, E., and Thal, D. (2008). Early language delay and risk for language impairment. *Persp. Lang. Learn. Educ.* 15, 93–100. doi: 10.1044/lle15.3.93

Feldman, H. M., Dollaghan, C. A., Campbell, T. F., Kurs-Lasky, M., Janosky, J. E., and Paradise, J. L. (2000). Measurement properties of the MacArthur communicative development inventories at ages one and two years. *Child Dev.* 71, 310–322. doi: 10.1111/1467-8624.00146

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., and Pethick, S. J. (1994). Variability in early communicative development. *Monogr. Soc. Res. Child Dev.* 59:6093. doi: 10.2307/1166093

Fenson, L., Dale, P. S., Reznick, J. S., Thal, D., Bates, E., Hartung, J. P., et al. (1993) *Technical manual for the MacArthur communicative development inventories*. San Diego, CA: San Diego State University.

Fenson, L., Marchman, V., Thal, D. J., Dale, P. S., Reznick, J. S., and Bates, E. (2007) *MacArthur-Bates communicative development inventories: User's guide and technical manual. 2nd Edn*. Oxford: Brookes.

Fisher, E. L. (2017). A systematic review and meta-analysis of predictors of expressive-language outcomes among late talkers. *J. Speech Lang. Hear. Res.* 60, 2935–2948. doi: 10.1044/2017_JSLHR-L-16-0310

Frank, M., Braginsky, M., Yurovsky, D., and Marchman, V. (2017). Wordbank: an open repository for developmental vocabulary data. *J. Child Lang.* 44, 677–694. doi: 10.1017/S0305000916000209

Frank, M. C., Braginsky, M., Yurovsky, D., and Marchman, V. A. (2021). *Variability and consistency in early language learning: The Wordbank project*. Cambridge, MA: The Mit Press.

Frick Semmler, B. J., Bean, A., and Wagner, L. (2023). Examining core vocabulary with language development for early symbolic communicators. *Int. J. Speech Lang. Pathol.*, Early Online, 1–10. doi: 10.1080/17549507.2022.2162126

Galea, L., and Ussishkin, A. (2018). "Onset clusters, syllable structure and syllabification in Maltese" in *The languages of Malta*. eds. P. Paggio and A. Gatt (Berlin: Language Sciences Press), 55–79.

Galeote, M., Checa, E., Sánchez-Palacios, C., Sebastián, E., and Soto, P. (2016). Adaptation of the MacArthur-Bates communicative development inventories for Spanish children with down syndrome: validity and reliability data for vocabulary. *Am. J. Speech Lang. Pathol.* 25, 371–380. doi: 10.1044/2015_AJSLP-15-0007

Gatt, D. (2010). *Early expressive lexical development: Evidence from children brought up in Maltese-speaking families. Unpublished PhD thesis*. Msida: University of Malta.

Gatt, D., and Dodd, B. (2019). Preschoolers' lexical skills in two majority languages: is there evidence for the onset of sequential bilingualism? *Int. J. Biling.* 24, 222–245. doi: 10.1177/1367006919826408

Gatt, D., Grech, H., and Dodd, B. (2014). Early expressive vocabulary skills: a multi-method approach to measurement. *First Lang.* 34, 136–154. doi: 10.1177/0142723714521830

Gatt, D., O'Toole, C., and Haman, E. (2011) Questionnaire for parents of bilingual children: Infants and toddlers version (PaBiQIT). Available at: http://www.bi-sli.org (Accessed December 19, 2022).

Grech, H., and Dodd, B. (2008). Phonological acquisition in Malta: a bilingual language learning context. *Int. J. Biling.* 12, 155–171. doi: 10.1177/1367006908098564

Hoberman, R. D. (2007). "Maltese morphology" in *Morphologies of Asia and Africa*. ed. A. S. Kaye (University Park, PA: Eisenbrauns), 257–281.

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., and Lyons, T. (1991). Early vocabulary growth: relation to language input and gender. *Dev. Psychol.* 27, 236–248. doi: 10.1037/0012-1649.27.2.236

Jackson-Maldonado, D., Thal, D., Marchman, V., Bates, E., and Gutiérrez-Clellen, V. (1993). Early lexical development in Spanish-speaking infants and toddlers. *J. Child Lang.* 20, 523–549. doi: 10.1017/S0305000900008461

Kauschke, C., and Hofmeister, C. (2002). Early lexical development in German: a study on vocabulary growth and vocabulary composition during the second and third year of life. *J. Child Lang.* 29, 735–757. doi: 10.1017/S0305000902005330

Kidd, E., and Garcia, R. (2022). How diverse is child language acquisition research? First Language. *Adv. Online Publicat.* 42:6405. doi: 10.1177/01427237211066405

Krug, M., and Sönning, L. (2018). "Language change in Maltese English: the influence of age and parental languages" in *The languages of Malta*. eds. P. Paggio and A. Gatt (Berlin: Language Science Press), 247–270.

Laubscher, E., and Light, J. (2020). Core vocabulary lists for young children and considerations for early language development: a narrative review. *Augment. Altern. Commun.* 36, 43–53. doi: 10.1080/07434618.2020.1737964

Loeb, D. F., Imgrund, C. M., Lee, J., and Barlow, S. M. (2020). Language, cognitive and motor outcomes of toddlers who were born preterm. *Am. J. Speech Lang. Pathol.* 29, 625–637. doi: 10.1044/2019_AJSLP-19-00049

Montanari, S., and Nicoladis, E. (2016). "Concluding remarks and future directions" in *Bilingualism across the lifespan: Factors moderating language proficiency*. eds. E. Nicoladis and S. Montanari (Washington, DC: American Psychological Association), 325–330.

Myers-Scotton, C. (2002). *Contact linguistics: Bilingual encounters and grammatical outcomes*. Oxford: Oxford University Press.

Myers-Scotton, C. (2006). *Multiple voices: An introduction to bilingualism*. Hoboken, NJ: Blackwell.

National Council for the Maltese Language (2021). The Department of Maltese, University of Malta, in collaboration with the National Statistics Office, Malta. The State of the Maltese Language: National Survey 2021. Available at http://www.kunsilltalmalti.gov.mt/file.aspx?f=343

National Statistics Office (2007) *Census of population and housing 2005 Population. Vol. 1*, Washington, DC: National Statistics Office.

O'Toole, C., and Fletcher, P. (2010). Validity of a parent report instrument for Irish-speaking toddlers. *FirstLanguage* 30, 199–217. doi: 10.1177/0142723709359237

Rescorla, L., and Dale, P. (2013). "Where do we stand now? Conclusions & future directions" in *Late talkers: From research to practice*. eds. L. Rescorla and P. Dale (Oxford: Brookes), 377–388.

Ring, E. D., and Fenson, L. (2000). The correspondence between parent report and child performance for receptive and expressive vocabulary beyond infancy. *First Lang.* 20, 141–159. doi: 10.1177/014272370002005902

Roberts, J. E., Burchinal, M., and Durham, M. (1999). Parents' report of vocabulary and grammatical development of African American preschoolers: child and environmental associations. *Child Dev.* 70, 92–106. doi: 10.1111/1467-8624.00008

Roberts, J. E., Rosenfeld, R. M., and Zeisel, S. A. (2004). Otitis media and speech and language: a meta-analysis of prospective studies. *Pediatrics* 113, E238–E248. doi: 10.1542/peds.113.3.e238

Robillard, M., Mayer-Crittenden, C., Minor-Corriveau, M., and Bélanger, R. (2014). Monolingual and bilingual children with and without primary language impairment: Core vocabulary comparison. *Augment. Altern. Commun.* 30, 267–278. doi: 10.3109/07434618.2014.921240

Salerni, N., Assanelli, A., D'Odorico, L., and Rossi, G. (2007). Qualitative aspects of productive vocabulary at the 200-and 500-word stages: a comparison between spontaneous speech and parental report data. *First Lang.* 27, 75–87. doi: 10.1177/0142723707067545

Slobin, D. I., and Bowerman, M. (2007). Interfaces between linguistic typology and child language research. *Linguistic Typology* 11, 213–226. doi: 10.1515/LINGTY.2007.015

Stiles, J. (1994). "On the nature of informant judgements in inventory measures: … and so what is it you want to know? Commentary" in *Monogr. Soc. Res. Child Dev.* 59, 180–185.

Tardif, T., Gelman, S. A., and Xu, F. (1999). Putting the "noun bias" in context: a comparison of English and mandarin. *Child Dev.* 70, 620–635. doi: 10.1111/1467-8624.00045

Trembath, D., Balandin, S., and Togher, L. (2007). Vocabulary selection for Australian children who use augmentative and alternative communication. *J. Intellect. Develop. Disabil.* 32, 291–301. doi: 10.1080/13668250701689298

Vella, A. (2013). Languages and language varieties in Malta. *Int. J. Biling. Educ. Biling.* 16, 532–552. doi: 10.1080/13670050.2012.716812

Vihman, M. M., and McCune, L. (1994). When is a word a word? *J. Child Lang.* 21, 517–542. doi: 10.1017/S0305000900009442

Weber, A. M., Marchman, V. A., Diop, Y., and Fernald, A. (2018). Validity of caregiver-report measures of language skill for Wolof-learning infants and toddlers living in rural African villages. *J. Child Lang.* 45, 939–958. doi: 10.1017/S0305000917000605

Weisleder, A., Friend, M., Tsui, A., and Marchman, V. A. (2022) Using parent report to measure vocabulary in young bilingual children: A scoping review [online]. Available at: psyarxiv.com/un6yr

Zambrana, I. M., Pons, F., Eadie, P., and Ystrom, E. (2014). Trajectories of language delay from age 3 to 5: persistence, recovery and late onset. *Int. J. Lang. Commun. Disord.* 49, 304–316. doi: 10.1111/1460-6984.12073

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 03 August 2023
DOI 10.3389/fpsyg.2023.1123983



OPEN ACCESS

EDITED BY
Maria-José Ezeizabarrena,
University of the Basque Country UPV/EHU, Spain

REVIEWED BY
Jan De Jong,
University of Bergen, Norway
Mari Aguilera,
University of Barcelona, Spain

*CORRESPONDENCE
Odelya Ohana
✉ odelya20@gmail.com

RECEIVED 14 December 2022
ACCEPTED 28 June 2023
PUBLISHED 03 August 2023

CITATION
Ohana O and Armon-Lotem S (2023) Parental reports on the lexicon of children from diverse bilingual populations.
*Front. Psychol.* 14:1123983.
doi: 10.3389/fpsyg.2023.1123983



# Parental reports on the lexicon of children from diverse bilingual populations

Odelya Ohana[1]* and Sharon Armon-Lotem[1,2]

[1]Department of English Literature and Linguistics, Bar Ilan University, Ramat Gan, Israel, [2]The Gonda Multidisciplinary Brain Research Center, Bar Ilan University, Ramat Gan, Israel

Parental questionnaires have been widely used to assess children's vocabularies. The MacArthur-Bates Communicative Developmental Inventories (MB-CDI) have been adapted into over 100 languages, providing researchers with access to various languages. As the vocabularies of bilingual children are distributed across their two languages, language knowledge must be assessed in both languages. While this can be done with two questionnaires, one for each language, the present study makes use of a multicultural adaptation of the MB-CDI, within a single questionnaire, that was geared specifically for bilingual context. In order to explore the developmental trajectories of the vocabularies of 90 bilingual children from diverse linguistic populations (English-Hebrew ($n$=30), French-Hebrew ($n$=30), and Russian-Hebrew ($n$=30) speaking families) parents reported on both the Home Language (HL) and the Societal Language-Hebrew (SL-Hebrew). Parents also provided background information about the child, the child's family, and exposure to each language. Our findings show no significant difference between vocabulary size of children from diverse bilingual populations in the HL and the SL, for both production and comprehension. Moreover, children from all three groups demonstrate balanced bilingualism at the group level. Correlations were found between both exposure to and use of each language by children, and various vocabulary measures across the three groups. The similar vocabulary levels demonstrated by the three groups as well as the balanced bilingualism can be explained by the relatively high prestige of all languages tested. Exposure to each language shows support in that language and a negative effect on the other language, demonstrating the crucial role exposure plays in bilingual children's language performance.

KEYWORDS

**cross-linguistic comparison, bilingualism, language exposure, parental reports, multiculturalism**

## 1. Introduction

Bilingual children face a significant challenge when it comes to vocabulary. Multiple studies have shown that the distribution of words between the two languages is often unbalanced. This uneven distribution sometimes leads to gaps from monolingual children in at least one of the languages of bilingual children (Thordardottir et al., 2006; Miękisz et al., 2017). The challenge posed by vocabulary could be traced to its sensitivity to language contact (De Houwer, 2015), exposure variables (Armon-Lotem and Meir, 2019; Armon-Lotem et al., 2021), and literacy (Bialystok, 2002). Parental questionnaires, such as the MB-CDI (MacArthur-Bates Communicative Developmental Inventories) (Fenson et al., 1991) are often used in order to assess the lexicon of young monolingual children. For bilingual children, it has been suggested to use a questionnaire for each language in order to meet the need to assess both languages [American Speech-Language-Hearing Association (ASHA), 2004], but there are not enough assessment tools geared specifically for the bilingual population (Thordardottir et al., 2006; Boerma and Blom, 2017). Moreover, there are not many Speech and Language Pathologists who can assess bilingual children in both languages and no norms are available for bilingual children in most cases (Bedore et al., 2005; Abutbul-Oz and Armon-Lotem, 2022). In Cyprus, for example, a program for Speech and Language Therapy has only been introduced in recent years (Kambanaros and Grohmann, 2013). The multicultural questionnaire (Ohana and Armon-Lotem, 2023) has been developed to address these difficulties, in order to enable parents and Speech and Language Pathologists to report on the receptive and expressive vocabulary of bilingual children using a single questionnaire in the societal language. In the present study this questionnaire will be used to examine the effect of exposure variables such as Age of onset of Bilingualism (AOB), reported exposure to each language and its use by the child, and the effect of language of books and screens in both languages on the vocabulary of bilingual children of the three populations. Such close examination of bilingual vocabulary, comparing different populations with a single tool is expected to highlight the unique features of each population and at the same time identify commonalities in their vocabulary.

An accurate account of the vocabulary of a bilingual child should take into account both languages. This is especially important because many bilingual children are dominant in one of their languages. Evaluating bilinguals' vocabulary in their weaker language only may result in lower vocabulary levels in comparison to monolingual children (Thordardottir et al., 2006; Miękisz et al., 2017). Moreover, bilinguals show great variability in terms of their vocabulary in each language and in both languages together, since exposure to each language and Age of Onset of Bilingualism (AOB) is different for each individual (e.g., Schwartz et al., 2009; Armon-Lotem and Meir, 2019; Thordardottir, 2019).

Due to the distribution of vocabulary across two languages, conceptual vocabulary has been proposed as a way to capture the sum of concepts known by a bilingual child. Conceptual vocabulary takes into account concepts from both languages but credits children only once for each concept they know in either one or both languages (Pearson et al., 1993). Conceptual vocabulary was found to be a good indicator of the vocabulary of bilingual children in both languages together (Junker and Stockman, 2002; O'Toole et al., 2017). A cross-linguistic study by O'Toole et al. (2017) compared vocabulary levels of bilingual children ages 24–36 months speaking a variety of languages. In their study, O'Toole et al. measured both total vocabulary and total conceptual vocabulary. They found total conceptual vocabulary a better measurement than total vocabulary when comparing different bilingual population since conceptual vocabulary reflects smaller gaps between the different versions of the CDI. Moreover, several studies show that conceptual vocabulary obtained by two independent questionnaires in the two languages of bilinguals is comparable to conceptual vocabulary obtained from a single questionnaire (e.g., Ohana and Armon-Lotem, 2023; O'Toole and Fletcher, 2010). O'Toole and Fletcher (2010) adapted the English CDI to Irish and used a single questionnaire to report on both Irish and English. They compared vocabulary levels of children on this questionnaire with direct observations of children's vocabulary and found significant correlations between the two. Ohana and Armon-Lotem (2023) compared conceptual vocabulary that was obtained from a single multicultural questionnaire and conceptual vocabulary that was obtained from two independent questionnaires and found significant correlations between the two.

The vocabulary distribution and balance across the two languages can be influenced by the status of the acquired languages determining whether children are balanced bilinguals or dominant in one language. Languages that are spoken by a migrant minority compete with the language of the host society and their maintenance is affected by their status as well as support and acceptance by the society. In Israel, the majority of people are bilingual or multilingual, and there is a variety of languages spoken at different homes (The Central Bureau of Statistics, 2021). French and Russian are spoken in Israel by large communities, supporting their use. Both French and Russian have a relatively high status and speakers of these languages feel obliged to maintain them, and to pass them on to the next generations (Schwartz et al., 2009; Armon-Lotem and Meir, 2019). English enjoys an even higher status, being a *lingua franca* spoken by a large number of speakers, and supported by the education system in Israel (Armon-Lotem and Meir, 2019). Hebrew, as the societal language (SL), has naturally a high status and is by-and-large spoken by both monolingual and bilingual children. The high status of English and Hebrew was proposed as a possible explanation for the balanced lexicon observed among English-Hebrew bilingual children (Armon-Lotem and Ohana, 2017; Ohana and Armon-Lotem, 2023). Studies of Russian-Hebrew speakers show that despite their emphasis on HL-Russian maintenance, there is a growing trend toward SL dominance (Remennick, 2003; Altman et al., 2021). Russian in Israel has a high vitality with a large Russian speaking community that preserves Russian among the young generation. However, despite the high status of Russian for its speakers, they regard Hebrew proficiency as the key to academic success. By contrast, while English-Hebrew speakers regard Hebrew highly because of its religious and Zionist aspect, they view English as the key to academic success because of its being *lingua franca*. The different status of English and Russian, alongside the findings of the above studies, could lead to the hypothesis that the Russian-Hebrew speakers would outperform the English-Hebrew speakers in SL-Hebrew tests. A comparison of these two bilingual populations (Armon-Lotem et al., 2014), focusing on English-Hebrew and Russian-Hebrew bilingual children ages 4;4–6;1 showed, however, that the English-Hebrew speakers' vocabulary in Hebrew was relatively the same as Hebrew vocabulary of the Russian-Hebrew group, with no significant differences between the two. To the best of our knowledge there are no recent studies investigating the vocabularies of French-Hebrew speakers in Israel.

The length of exposure to a language and the age in which language exposure begins are inherently different for monolinguals and bilinguals. For monolingual children length of exposure to their only language is identical to their chronological age since it normally begins at birth. In contrast, bilingual language exposure can begin at birth for both languages, which is the case with simultaneous bilinguals, or with exposure to one language from birth, and the other later, e.g., after the age of three for sequential bilinguals (Paradis, 2010). Some researchers distinguish bilinguals who acquire both languages from birth, from bilinguals who acquire their second language only at their second year of life, defining the latter early sequential bilinguals (Armon-Lotem et al., 2011). In both cases, language exposure is always distributed across the two languages. Therefore, aside from the length of exposure and AOB, the amount of exposure to each language must be considered when examining the vocabulary of bilingual children. Moreover, studies have shown that the amount of exposure to each language is a stronger indicator of vocabulary size than AOB and length of exposure to each language (Thordardottir, 2019).

While monolingual children receive exposure to one language only at any time and context, the language exposure of bilingual children is distributed across two languages, receiving less input than monolinguals in each (Hoff, 2018; Hoff et al., 2018). This may explain the lower vocabulary levels observed when comparing monolingual and bilingual children's vocabulary in a single language (Thordardottir et al., 2006; Miękisz et al., 2017). Moreover, while home related concepts are often acquired in the HL, concepts related to school and to outside things may be first acquired in the SL, and only later in the HL.

Literacy and book reading at home are another important source of language exposure that must be taken into consideration when exploring the vocabulary of young children. Research has shown that exposure to literacy in a language affects vocabulary in this language positively, for both monolingual and bilingual children (Jiménez et al., 2006; Quiroz et al., 2010). Jiménez et al. (2006) investigated a sample of 16 Spanish speakers, ages 7–8 years old, exposed to English outside of their homes. Parents reported on the frequency of book reading at home, and were videotaped while engaged in book reading with their children. Their findings showed that book reading at home enhanced vocabulary in the language in which book reading was done. Another study by Quiroz et al. (2010) investigated the effect of literacy on the vocabulary of Spanish-English bilingual children ages 4–5. Quiroz et al. (2010) found that home literacy activities in one language correlated positively with vocabulary in that language and negatively with the other language. Research on the effect of literacy and book related activities at home on the vocabulary of bilingual children in other populations is rather limited, calling for further research.

It has long been suggested that parents are the best observers and reporters of the language of their children. While lab testing is limited to certain contexts and time, parents observe their children in a variety of contexts and for a lengthier period of time. Moreover, children might feel uncomfortable when tested by an experimenter, while they behave and speak freely in their natural environment (DeMayo et al., 2021). Multiple studies have used parent questionnaires that report on the vocabulary of their children, and found these questionnaires to be reliable for assessing children's knowledge (Marchman and Martínez-Sussmann, 2002; Fenson et al., 2007; O'Toole and Fletcher, 2010), and collecting background information related to the child, his family, and language exposure patterns at home and outside of it (Schwartz et al., 2009; Armon-Lotem et al., 2014; Abutbul-Oz and Armon-Lotem, 2022). Moreover, research has shown that parents are also able to report reliably on their children's language skills and assist in diagnosis of developmental language disorder (Auza et al., 2023).

The MacArthur-Bates Communicative Development Inventories (MB-CDI, Fenson et al., 1991) are a set of parent questionnaires allowing parents to report on the vocabulary and grammar of their children. It has been adapted to over 100 languages with several bilingual adaptations. Comparison of parental reports of their children's vocabulary with direct measures showed that parents were able to report on their children's knowledge accurately for both monolingual and bilingual children. Heilmann et al. (2005) tested the vocabulary of a hundred monolingual English speakers aged 30-months, with direct measures and the MB-CDI. They found significant correlations between the two, demonstrating the validity of parental reports in assessing their children's vocabulary.

Moreover, several studies have already validated the use of a single questionnaire to assess vocabularies in both languages of bilingual children (e.g., Gatt, 2007; O'Toole and Fletcher, 2010; Dale and Penfold, 2011). For example, O'Toole and Fletcher (2010) examined the vocabulary of 21 Irish-English bilinguals aged 1;4–3;4, using a new bilingual adaptation of the CDI. They compared parent reports on vocabulary with spontaneous language samples and found significant correlations between the two. These findings validate the ability of parents to report accurately on the vocabulary of their bilingual children in both languages and to distinguish between the two languages (Marchman and Martínez-Sussmann, 2002).

With this in mind, the multicultural questionnaire used in the present study has been developed and validated (Ohana and Armon-Lotem, 2023). The multicultural questionnaire, delivered in Hebrew, the SL, includes concepts that are shared by monolingual CDI questionnaires of the SL and the HLs of the tested populations, as well as a selection of culturally specific items which are unique to the different HLs. Thus, parents report on both languages within a single questionnaire. The multicultural questionnaire was validated by comparing vocabulary levels of 38 English-Hebrew bilinguals as reported on this questionnaire with vocabulary levels as were reported for 38 English-Hebrew bilinguals on two separate questionnaires-the English CDI (Fenson et al., 1991) and the Hebrew CDI (Maital et al., 2000). Children from both groups were matched on age (24–48 months), socio economic status (mid-high SES), and age of onset of bilingualism (Mean = 4 and Mean = 4.42 for the group using the monolingual questionnaires and using the multicultural questionnaire, respectively). The study showed no effect for using two different questionnaires or a single multicultural one, no effect for language (performance on Hebrew and English were similar), with a highly significant effect for modality with comprehension higher than production across the different questionnaires. That is, parents reported similar vocabulary levels in each of the languages, independently of the questionnaires that were used for these reports. The similar responses of parents using the multicultural questionnaire to those using two separate questionnaires, support the use of a single multicultural questionnaire to report on two different languages. More details of the validation of the multicultural questionnaire are provided in Ohana and Armon-Lotem (2023).

Other studies have tested the ability of parents to report the relative exposure to each language, and other background variables that might affect children's language performance (e.g., Hoff et al., 2012). Research has shown that parental estimation of the amount of exposure of their bilingual children to each language were accurate (Hoff et al., 2018). In a study by Place and Hoff (2011), parents were asked to report on relative exposure to each language for their bilingual English-Spanish speaking children (mean age: 25.66 months). They found that the relative exposure to each language was a significant predictor of vocabulary in that language, arguing that this demonstrated parents' ability to report accurately on exposure to each language.

The present study uses a multicultural questionnaire (Ohana and Armon-Lotem, 2023) for evaluating the lexicon of three bilingual populations speaking Societal Language-Hebrew (SL-Hebrew) with Home Language (HL) English, French or Russian, in order to explore the differential effect of HL, language exposure and literacy exposure on the vocabulary of bilingual children in both languages.

In light of the above research, three questions will be examined next:

1. Do bilingual children, exposed to the same SL with different HLs, demonstrate different developmental trajectories of their vocabularies in each language separately and in both languages together?
2. Do reported exposure patterns (such as, reported languages spoken with the child) and reported language use (such as, reported languages used by the child) coincide with the vocabulary levels of children in each language? Is there a difference between the different HL populations?
3. How does exposure to books and screens affect vocabulary in each language? Is this effect similar across the HL groups?

The following hypotheses are tested:

1. The developmental trajectories in each language separately and in both languages together are hypothesized to reflect their status and vitality within each community. It is predicted that bilingual children exposed to English, Russian, or French at home, with SL-Hebrew are expected to demonstrate balanced bilingualism as a group. This expectation for balanced bilingualism is due to the intense exposure children receive to both Hebrew and the HLs. While Hebrew is the SL, supported by the educational system, has a religious prestige and often viewed as key to integration in society, the three HLs enjoy a high status, dense communities, and high maintenance and support within the home and community. These large communities view their HLs as means for communicating with transnational family and preserving their homeland culture. English speakers are expected to present an advantage in their HL vocabulary over French and Russian speakers, since English is also a *lingua franca* with an academic value supported by the education system in Israel.
2. It is hypothesized that the amount of language exposure impact vocabulary size (Hoff et al., 2012). That is, the more children are exposed to one language, the higher their vocabulary should be in that language. Thus, we predict that reported exposure to SL-Hebrew is expected to have a positive effect on reported vocabulary in Hebrew and a negative effect on vocabulary in the HL. That, is, parent reports of languages spoken by the child are expected to be consistent with reports on vocabulary in each language.
3. Exposure to books and screens in one language is expected to correlate positively with vocabulary in that language and negatively with vocabulary in the other language (e.g., Quiroz et al., 2010).

# 2. Materials and methods

## 2.1. Participants

Data were collected from parents of 90 bilingual children, aged 24–48 months: 30 English-Hebrew speakers (15 girls) (M = 37.63, SD = 8.87), 30 French-Hebrew speakers (15 girls) (M = 37.60, SD = 8.02), and 30 Russian-Hebrew speakers (14 girls) (M = 37.57, SD = 9.20). All children were either simultaneous bilinguals from one-parent-one-language homes or early sequential bilinguals who were exposed to their second language before the age of two, acquiring the HL at home and the SL-Hebrew, outside of their homes. Most children ($n$ = 80) were attending a day care where the SL-Hebrew was used. Children had at least 6 months of exposure to the SL-Hebrew, similarly to the threshold determined in previous studies (e.g., O'Toole et al., 2017). Most children come from mid-high SES with parents who have an academic degree or at least a professional certification. Aside from one family from the French-Hebrew speaking group where both parents are unemployed, and seven families out of the three groups, where one parent is reported to be unemployed, all other parents are employed and several others are enrolled in academic studies. Table 1 presents background information for the entire sample ($N$ = 90).

No significant between-group differences were observed for the chronological age of the children, and the AOB and onset of exposure to the HL. In terms of family size, the majority of Russian-Hebrew speakers come from small families with one or two children, whereas in the English-Hebrew and the French-Hebrew speakers about a half of the group reports on three or more children. A chi-square test of independence showed there was a significant association between group and family size, due to significant difference between the English and the Russian cohort ($p$ = 0.006), but no significant association between group and birth order. For family income, parents reported whether the family income is average, below, or more than the average. Family income was found to have a highly significant association with group, showing the following hierarchy: Russian>English>French.

## 2.2. Instruments and procedures

The vocabularies of bilingual children were reported using a multicultural questionnaire (see Figure 1) that enabled parents to report on the vocabularies of bilingual children in both the HL and the SL-Hebrew with a single questionnaire in the SL-Hebrew (Ohana and Armon-Lotem, 2023). The multicultural questionnaire is an adaptation of the Hebrew CDI-Words and Sentences (HCDI: WS) (Maital et al., 2000), which originally consisted of a list of 602 items. From this list, three irrelevant items were removed (e.g., tape/cassette). To these, 34 culturally specific words selected from the English, Russian, and French CDI questionnaires were added, and three were removed, resulting in a list of 633 items (Fenson et al., 1991; Kern, 2007; Vershinina et al., 2011). These items were added mainly, but no only, to the category of Foods and Drinks (e.g., peanut butter, cabbage and baguette, from the English, Russian, and French CDIs, respectively). The selected items were added to the questionnaire in consultation with groups of parents from each bilingual population. Each group of parents was presented individually with the Hebrew adaptation of the CDI, along with the CDI version of their HL. The parents explored both questionnaires and pointed to relevant concepts that were found on the CDI in their HL, and were missing on the Hebrew CDI. These items were added in order to capture concepts in use by children from different homes and cultures, making the multilingual questionnaire a valid tool for the assessment of bilingual children. This resulted in a list of 633 concepts in the SL-Hebrew,

TABLE 1 Background information ($N$=90).

| | | English-Hebrew | French-Hebrew | Russian-Hebrew | Statistics |
|---|---|---|---|---|---|
| Number (females) | | 30 (15) | 30 (15) | 30 (14) | |
| Age in Months | Mean | 37.63 | 37.60 | 37.57 | $F$ (2,87) = 0.000 |
| | SD | 8.87 | 8.02 | 9.20 | |
| | Range | 24–48 | 24–48 | 24–48 | $p$ = 0.816 |
| AOB (in months) | Mean | 4.62 | 7.88 | 9.41 | $F$ (2,85) = 1.77 |
| | SD | 10.01 | 8.74 | 11.37 | |
| | Range | 0–40 | 0–33 | 0–36 | $p$ = 0.176 |
| Family size* (no. of children) | 1–2 | 13 | 18 | 25 | $X^2$ (8, $N$ = 90) = 15.73 |
| | 3–4 | 9 | 6 | 2 | |
| | 5+ | 5 | 5 | 3 | $p$ = 0.046 |
| Birth order | First born | 13 | 13 | 20 | $X^2$ (6, $N$ = 90) = 8.05 |
| | Second born | 6 | 4 | 5 | |
| | Later born | 11 | 12 | 5 | $p$ = 0.235 |
| Family Income** | > average | 15 | 12 | 19 | $X^2$ (6, $N$ = 90) = 21.41 |
| | = average | 7 | 6 | 11 | |
| | < average | 5 | 10 | 0 | $p$ = 0.002 |
| Mother Education*** | Academic/professional | 25 | 25 | 29 | |
| | High school graduate | 2 | 3 | 1 | |
| | Elementary/none | 0 | 1 | 0 | |
| Father Education*** | Academic/professional | 23 | 20 | 23 | |
| | High school graduate | 4 | 4 | 4 | |
| | Elementary/none | 0 | 3 | 0 | |

*Data about family size is missing for three participants in the English-Hebrew group and for one child in the French-Hebrew group. **Data about family income is missing for three participants in the English-Hebrew group and for two participants in the French-Hebrew group. ***Data about parents' education is missing for six participants in the English-Hebrew group, for four participant in the French-Hebrew group, and for three participants in the Russian-Hebrew group. AOB-Age of onset of bilingualism – the age in which exposure to SL-Hebrew has started.

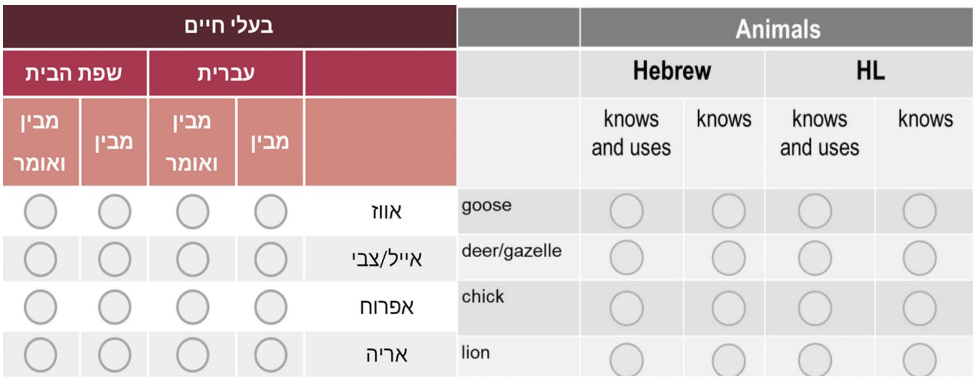


FIGURE 1
A part of the multicultural questionnaire with its translation* (four concepts from the category of animals'). *The questionnaire was in Hebrew only. The English is a translation for the benefit of the reader.

divided into categories (such as, animals, people etc.). For each concept, parents indicate whether their child knows this concept in the HL and/or the SL-Hebrew, addressing both comprehension and production. Parents also completed a background information form (Gendler-Shalev and Dromi, 2021) which included general information regarding language exposure patterns, child's developmental milestones, as well as information about the parents and the family. Participant recruitment was done through the social media, through groups of speakers of languages other than Hebrew in Israel, and by word-of-mouth. Parents used a link to the home page of the study to complete the questionnaire at their own convenience.[1] Once parents completed a short registration form, and gave their consent to participating in the study, they were transferred directly to the questionnaire. A full account of the procedures of creating the multicultural questionnaire is provided in Ohana and Armon-Lotem (2023).

## 2.3. Data analysis

The number of words comprehended and produced in each language on its own, and the conceptual vocabulary from both languages were calculated for each child automatically. Data analysis was performed using the IBM SPSS Statistics. Analyses included a three way – ANCOVA for repeated measures with Group (English, French, Russian), Language (HL, SL-Hebrew), and Modality (production, comprehension) as the independent variables, vocabulary levels as the dependent variable, and Age and Age of Onset of Bilingualism (AOB) as the covariates. A separate two-way ANOVA was computed for the conceptual vocabulary, with Group and Modality as the independent variables.

Further analyses were conducted to explore the effect of exposure variables on the individual vocabulary levels and to determine whether there is a match between reports of the two. Exposure to the SL-Hebrew and the use of SL by the child were reported on a 1–7 Likert scale (1. Only HL, 2. 2 h of Hebrew every day, 3. 4 h of Hebrew, 4. 6 h of Hebrew, 5. 8 h of Hebrew, 6. 10 h of Hebrew, 7. Only Hebrew). Under the assumption that children aged 2–4 years old have around 12 waking hours, children with reported 6 h of Hebrew per day were defined as children with equal exposure to both languages (HL = SL), whereas less than six hours of exposure to Hebrew is defined as dominant exposure in the HL (HL > SL) and more than six hours of Hebrew is defined as dominant exposure to SL (SL > HL).

Correlational analyses were computed to test exposure variables such as, chronological age, and AOB, on exposure and use of each language by the child, and their relation to vocabulary reports. In addition, for each child the gap between both languages was calculated (i.e., the number of words in HL minus the number of words in SL) and its relation to exposure variables was examined, in order to explore the effect of exposure on language dominance. A positive score indicates a larger vocabulary in HL in comparison to SL and vice versa.

Finally, exposure to books and screens was tested in order to evaluate their relative contribution to vocabulary levels in each language. Exposure to books was reported on a 1–4 Likert scale (1. rarely; 2. 1–2 times a week; 3. 3–5 times a week; 4. At least 1 book every day). Exposure to screens was reported on a 1–5 Likert scale, indicating the relative exposure to screens every day (1. no exposure, 2. rarely, 3. 1 h per day, 4. 2 h per day, 5. 3 h or more every day). Language of books/screens was reported on a Likert scale of 1–3 indicating the language in which books are read (1. Mainly in the HL, 2. Equally in both languages, 3. Mainly in the SL-Hebrew). Correlational analyses followed by hierarchical regression analyses were performed on the data to explore the effect of frequency of exposure as well as the language in which children were exposed to books and screens. Language of screens was entered into the regression as a variable determining the amount of Hebrew exposure through books and screens. Low Hebrew exposure means higher HL exposure through books and screens since exposure was reported on a scale between reading/watching only in Hebrew, in both language or only in HL without taking frequency into consideration. Separate hierarchical regressions were performed for each group, for both vocabulary production and vocabulary comprehension as the dependent variables, and age and exposure to books and screens as the predictors. Age was entered into each regression in the first step, and exposure to books and screens were entered in the second step to explore their effect on vocabulary size beyond children's age.

# 3. Results

To address the above questions, we start by comparing vocabulary size for the three groups and commence with an exploration of the relation between exposure and background factors and vocabulary size.

## 3.1. Vocabulary size: by group

Descriptive statistics for the vocabulary of the entire sample are presented in Table 2, for both languages of English-Hebrew, French-Hebrew, and Russian-Hebrew speakers, for both production and comprehension ($N = 90$).

A Three-Way Mixed ANCOVA with Group, Language, and Modality (production/comprehension) as independent variables, vocabulary levels as the dependent variable, and age as a covariate, shows that there is no main effect of Group, with all groups performing similarly overall. Moreover, independently of the group tested, there is a highly significant main effect for Modality, with comprehension rates higher than production rates. Additionally, no effect for Language was found, showing that children demonstrated similar vocabulary levels in both the HL and SL-Hebrew, with no significant differences between the two across the entire sample.

For the conceptual vocabulary, ANOVA performed on the data revealed no main effect of Group, with all three groups showing similar conceptual vocabulary levels for both production and comprehension. Similarly to results in each language separately, conceptual vocabulary demonstrates a significant main effect for Modality, with comprehension rates significantly higher than production rates. These findings remained consistent when controlling for age.

## 3.2. Vocabulary size: individual scores

In order to further investigate the effect of Group on the developmental trajectories of each language, comparisons of the individual vocabulary production scores in the two languages are presented in Figures 2–4 for each group separately (for the English-Hebrew, French-Hebrew, and Russian-Hebrew groups, respectively). Each figure presents the number of concepts each child produces in both the HL and the SL- Hebrew, for the 30 participants in the group. Since the multicultural questionnaire is in the SL-Hebrew each

1 www.bilingual-kids-israel.com

TABLE 2 Vocabulary levels of English-Hebrew speakers, French-Hebrew speakers, and Russian-Hebrew speakers in each language and in both languages together (conceptual vocabulary).

| | | English-Hebrew ($n$ =30) | French-Hebrew ($n$ =30) | Russian-Hebrew ($n$ =30) | Statistics |
|---|---|---|---|---|---|
| Home language mean (SD) | Production | 443.37 (199.68) | 415.47 (198.90) | 443.00 (216.60) | Group: $F$ (1, 86) = 0.079, $p$ = 0.924, $\eta_p^2$ = 0.002 |
| | Comprehension | 556.03 (115.15) | 541.07 (140.53) | 529.60 (148.34) | Modality: $F$ (1,87) = 101.316, $p$ < 0.001, $\eta_p^2$ = 0.538 |
| SL-Hebrew mean (SD) | Production | 415.43 (200.63) | 451.97 (184.95) | 436.87 (197.41) | Language: $F$ (1,87) = 0.004, $p$ = 0.950, $\eta_p^2$ <0.001 |
| | Comprehension | 541.30 (123.00) | 538.53 (167.19) | 527.07 (158.73) | |
| Conceptual vocabulary mean (SD) | Production | 517.77 (154.23) | 533.43 (120.78) | 515.63 (175.83) | Group: $F$ (1,87) = 0.099, $p$ = 0.906, $\eta_p^2$ = 0.002 |
| | Comprehension | 578.37 (84.31) | 589.37 (87.14) | 563.20 (120.14) | Modality: $F$ (1,87) = 41.584, $p$ < 0.001, $\eta_p^2$ = 0.323 |

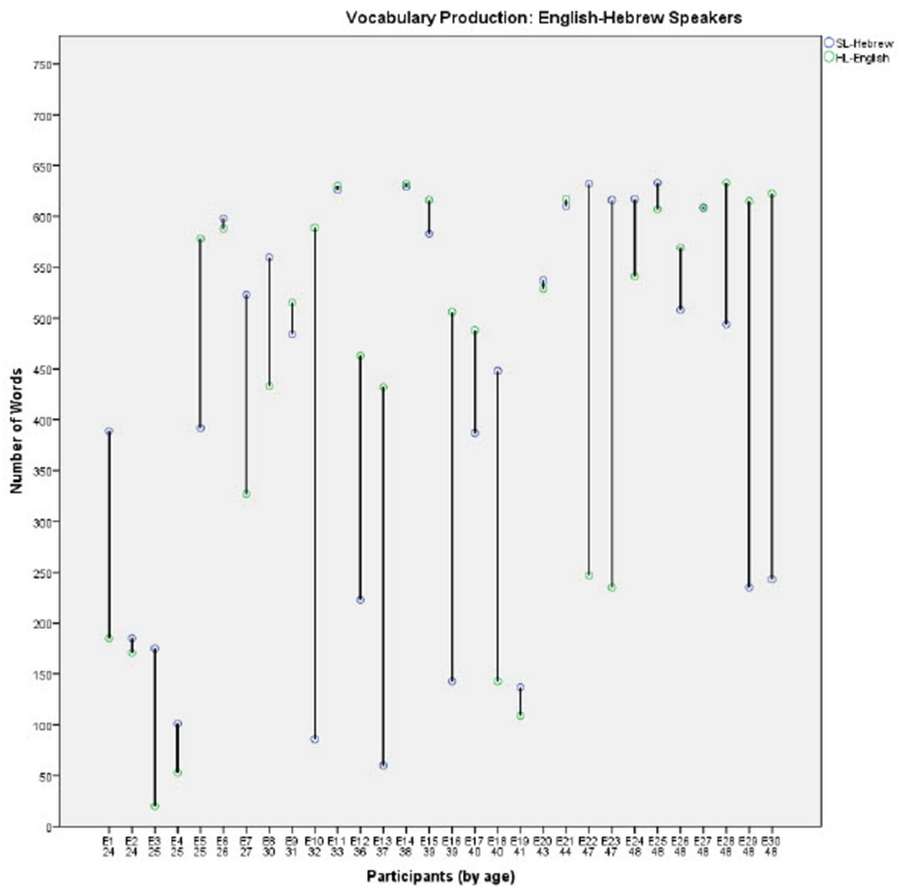


FIGURE 2
Vocabulary production levels for the English-Hebrew speakers.

concept represents two words, one in SL-Hebrew and the other one in the HL. The participants are presented in ordinal numbers with a capital letter representing their HL (for example, E1 represents participant 1 in the English-speaking group, E2-English participant 2, F1-French speaking participant 1, R1-Russian speaking participant 1 etc.). Under each participant's number, the age of the child is provided. For each participant, two data points are presented, for vocabulary in the HL and the SL-Hebrew.

Figures 2–4 illustrate the great variability between individuals, within the three groups. Some children demonstrate similar vocabulary levels in both the HL and the SL-Hebrew, and many others are highly dominant in one of their languages. A close inspection of

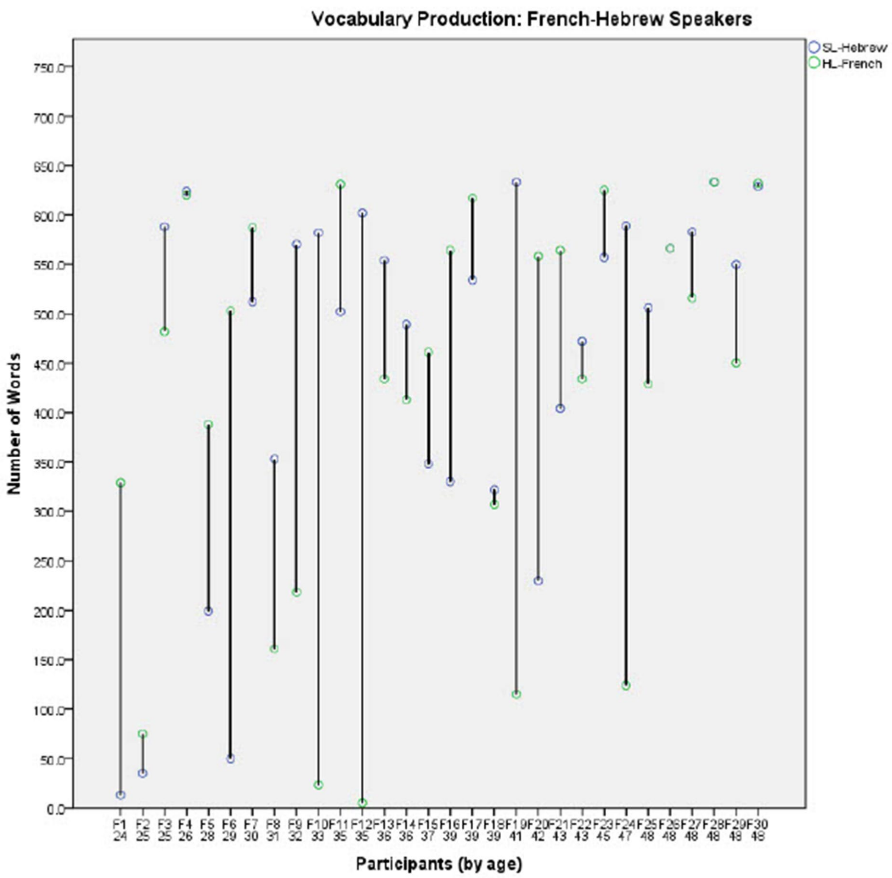


FIGURE 3
Vocabulary production levels for the French-Hebrew speakers.

individual children suggests that the amount of exposure to each language reported for individual children could be a possible source of the large gaps some children have between their vocabularies in the HL and SL-Hebrew. For example, E30 in the English-Hebrew sample (Figure 2) shows a great dominance in English over Hebrew, and this report is consistent with reported exposure patterns as parents reported that their child hardly speaks Hebrew. Another example is shown by F24 in the French-Hebrew sample (Figure 3) and R30 in the Russian-Hebrew sample (Figure 4) who both show a great advantage of vocabulary in their SL-Hebrew over their HL, and for both parents reported that they speak only in Hebrew and not in the HL. Thus, we next turn to the relation between exposure patterns and vocabulary size.

## 3.3. Language exposure and use

Table 3 presents AOB, language exposure by others, and language use by the child, for all three groups, providing the number of participants for the different patterns of exposure and use.

Table 3 shows that across the three groups, all the participants were exposed to the SL-Hebrew in the first year of life. Moreover, about half of the participants in each group were exposed to and used SL-Hebrew more than their HL.

Table 4 demonstrates that SL-exposure presents limited correlation with both HL or SL-vocabulary production (apart from a negative correlation with HL-vocabulary production among English-Hebrew speakers), while significant correlations were observed between HL- and SL-vocabulary production and language use by the child. Since the language directed at the child did not correlate with language outcomes, we next turn to investigate the relation between the language used by the child and vocabulary measures. SL use by the child correlated negatively with HL-vocabulary production scores for the English-Hebrew and Russian-Hebrew speaking populations but not for the French-Hebrew speaking population, while SL use by the child showed positive correlations with SL-vocabulary production across the three groups. The gap between the HL and the SL (HL minus SL), for both production and comprehension, strongly correlated with both language exposure and language use by the child (apart from the correlation between exposure and production gap for the Russian-Hebrew group). As a negative gap score indicates Hebrew dominance, the negative correlations indicate that with more exposure to SL-Hebrew children become more dominant in Hebrew, and being more dominant in Hebrew they use it more and understand it better.

## 3.4. Exposure to books and screens

We next turn to the effect of books and screens on the vocabulary of children in each language. Table 5 presents reports on the language exposure by books and screens, for each group, by the number of

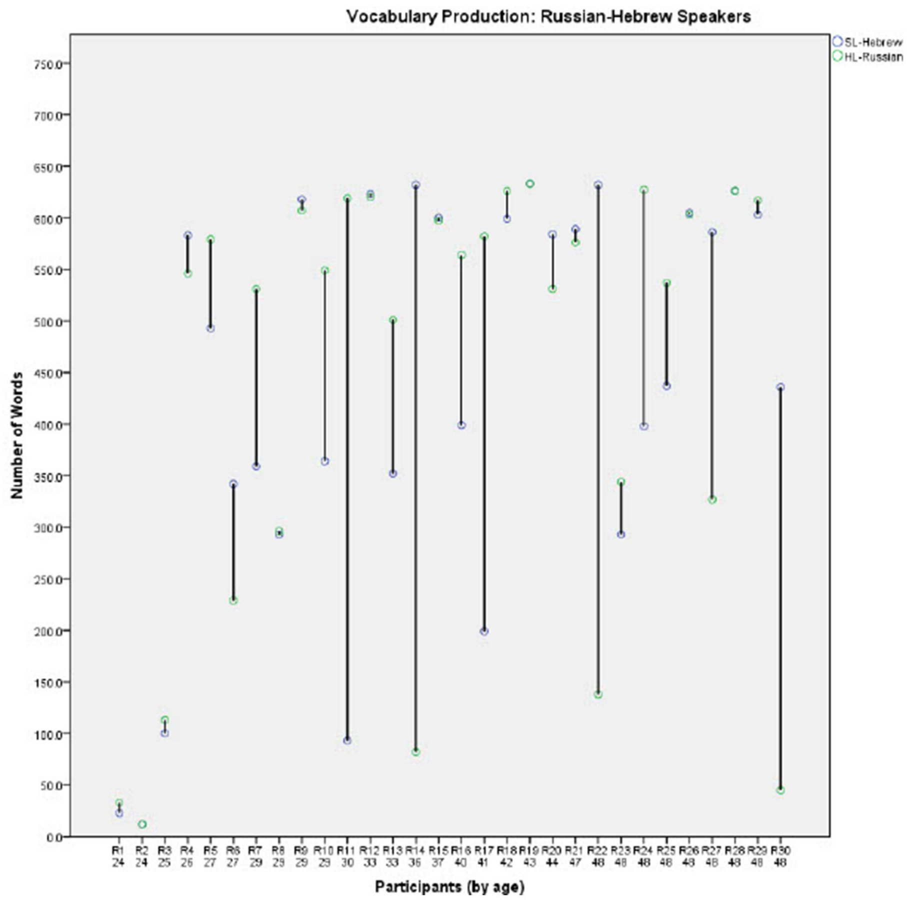


FIGURE 4
Vocabulary production levels for the Russian-Hebrew speakers.

participants. Each row indicates the number of participants with each exposure pattern.

A chi-square test of independence showed there was no association between group and both language of books ($X^2$ (4, $N=89$) = 6.26, $p=0.181$) and language of screens ($X^2$ (4, $N=77$) = 7.87, $p=0.096$). Further correlational analyses were performed to examine the effect of language of books and screens on vocabulary levels. Table 6 presents correlations between amount of Hebrew in books and screens and vocabulary in each language for both production and comprehension.

Exposure to the SL-Hebrew in books and on screens was not consistent across the three groups. For the English-Hebrew speaking group, SL-Hebrew exposure in both books and screens correlated negatively with both vocabulary production and comprehension in the HL-English. All these correlations were significant aside from the negative correlation between SL-Hebrew exposure to screens and vocabulary production in the HL-English which was nearly significant. For the French-Hebrew speaking group SL-Hebrew exposure in books correlated positively with both vocabulary production and comprehension in the SL-Hebrew. In addition, SL-Hebrew exposure in screens correlated positively with vocabulary production in this language, but not with vocabulary comprehension. For the Russian-Hebrew speaking group no correlations were found between SL-Hebrew exposure in books and screens and any of the vocabulary measures.

Hierarchical regression analyses were performed to further explore whether the effect of language exposure in books and screens on vocabulary size in each language goes beyond the age effect. Four separate Hierarchical regression models were conducted for each group, in each language, for both production and comprehension (for example, for the English-Hebrew speaking group there were separate models for English production, English comprehension, Hebrew production, Hebrew comprehension). In each model, age was introduced in the first step. Both language of books and language of screens were added, as two separate variables, in the second step. Vocabulary was the dependent variable. Results are presented in Table 7. Across the three groups only five models were significant and are presented in the table.

In the English-Hebrew group, both models which predict vocabulary production in English were significant. The model with exposure to books and screens together with age explained 36% of the variance in vocabulary size. While age made a significant contribution to the model, and exposure to books made a marginal contribution to the model, exposure to screens did not make a statistically significant contribution to the model. Likewise for comprehension, the second model, where age is combined with exposure to books and screens explained 23% of variance, beyond the 13% of the variance which was explained by age only in the first step. Overall, the regression explained 36% of the variance. Only the contribution of exposure to books contributed to the explained variance of English comprehension beyond age and the exposure to screens was not a significant predictor. Similar results were found for the Russian-Hebrew speaking group. Both models predicting Hebrew vocabulary production were significant. The first model with age as the only predictor explains 24% of the variance while the second model explains 33%. For comprehension both models were significant but account for relatively the same variance in vocabulary size (32% and 33% for the first and the second models, respectively). Interestingly enough, in the French-Hebrew speaking group a similar picture was revealed only for the models predicting Hebrew vocabulary comprehension which were both significant. While the model with age as the only predictor accounts for 19% of the variance, the model with age combined with exposure variables explains 33% of the variance.

TABLE 3 Exposure variables by number of participants for each population.

| | | English-Hebrew | French-Hebrew | Russian-Hebrew |
|---|---|---|---|---|
| AOB (in months) | Mean | 4.62 | 7.88 | 9.41 |
| | SD | 10.01 | 8.74 | 11.37 |
| | Range | 0–40 | 0–33 | 0–36 |
| Language exposure (by others) | HL > SL | 4 | 5 | 9 |
| | HL = SL | 8 | 10 | 2 |
| | HL < SL | 18 | 15 | 19 |
| Language use (by the child) | HL > SL | 9 | 8 | 7 |
| | HL = SL | 3 | 3 | 4 |
| | HL < SL | 16 | 15 | 19 |

AOB, Age of Onset of Bilingualism refers to the age in which exposure to SL-Hebrew has started, HL > SL – up to 4 h of Hebrew, HL = SL – 6 h of Hebrew, HL < SL – 8 or more hours of Hebrew.

# 4. Discussion

The present study aims at identifying the developmental trajectories of the vocabularies of children from three bilingual populations, English-Hebrew, French-Hebrew, and Russian-Hebrew speakers. Our finding for each research question will be addressed separately.

## 4.1. Vocabulary level in both HL and SL-Hebrew

Our findings show that, bilingual children speaking English, French, and Russian as the HL and Hebrew as the SL have similar vocabulary levels in each language on its own and in both languages together (Tables 2, 3). The similar vocabulary levels can be attributed to the characteristics of these three bilingual populations. All three populations are exposed to SL-Hebrew in daycare centers and preschools, and very often at home as well. On the other hand, aside from the fact that English is a *lingua franca*, all three languages are widely spoken and by large communities which support and strengthen the use of the HLs. As expected, (Ring and Fenson, 2000; Abdelwahab et al., 2021) parents report significantly higher comprehension rates than production rates across the three groups, independently of the language examined. This is also true for the conceptual vocabulary which represents vocabulary from both the HL and SL-Hebrew. The ability of parents to distinguish comprehension from production is documented in the literature

TABLE 4 Correlation between exposure to and use of Hebrew by the child and the gap between the two languages.

| | SL-Hebrew exposure | | | SL-Hebrew used by the child | | |
|---|---|---|---|---|---|---|
| | En-Heb | Fre-Heb | Rus-Heb | En-Heb | Fre-Heb | Rus-Heb |
| HL-production | −0.433* | ns. | ns. | −0.554** | ns. | −0.421* |
| SL-production | ns. | ns. | ns. | 0.397* | 0.522** | 0.593** |
| ProdGap | −0.703** | −0.540** | ns. | −0.823** | −0.689** | −0.754** |
| CompGap | −0.566** | −0.368* | −0.453* | −0.598** | −0.429* | −0.689** |

*$p < 0.05$, **$p < 0.01$. ProdGap and CompGap refer to the gaps between words in the two languages (HL-SL) for both production and comprehension. The larger the gap between the languages is, the better children are on the HL in comparison to the SL-Hebrew.

TABLE 5 Reported language of books and screens for each group.

| | | English-Hebrew (number of participants) | French-Hebrew | Russian-Hebrew | Statistics |
|---|---|---|---|---|---|
| Language of Books | HL > SL | 13 | 8 | 10 | $X^2$ (4, $N = 89$) = 6.26 |
| | HL = SL | 11 | 7 | 12 | |
| | HL < SL | 6 | 14 | 8 | $p = 0.181$ |
| Language of screens | HL > SL | 19 | 10 | 13 | $X^2$ (4, $N = 77$) = 7.87 |
| | HL = SL | 6 | 7 | 10 | |
| | HL < SL | 1 | 6 | 5 | $p = 0.096$ |

HL > SL – up to 4 h of Hebrew, HL = SL – 6 h of Hebrew, HL < SL – 8 or more hours of Hebrew.

(Ring and Fenson, 2000). Moreover, though children demonstrate balanced bilingualism at the group level, there is a great variability within the group. Individual results show that some children demonstrate balanced bilingualism with similar vocabulary levels in both languages, but many others demonstrate dominant bilingualism with large gaps between their reported vocabulary in the two languages, demonstrating dominance in either the HL or SL-Hebrew. This variability is shown by the large SDs presented for each bilingual group and is in line with reports from previous research (Fenson et al., 2000; Armon-Lotem and Meir, 2019; Frank et al., 2021).

## 4.2. Language exposure

To further understand the great variability within the groups and the factors affecting language dominance, exposure patterns to each language were investigated. Across the three groups, children showed very similar AOB ranges, similar patterns of exposure to each language by others, and similar language use by the child. AOB did not correlate significantly with vocabulary measures in all three groups apart from a correlation with the gap between the HL and SL vocabulary within the English-speaking group. This correlation can be explained by the characteristics of English-speaking homes and the prestigious status of English which enable parents to maintain exposure to HL-English until children are officially exposed to the SL-Hebrew. It is important to note, that for all three populations the majority of children were attending a Hebrew speaking day-care that usually starts at the age of 3.5–6 months, and therefore they showed balanced bilingualism as a group. Moreover, most of the children were exposed to the SL-Hebrew before the age of 6 months, and many were exposed to the SL-Hebrew from birth. This could explain the lack of correlation between AOB and vocabulary measures. Furthermore, previous studies found AOB is not a strong enough predictor of vocabulary since it provides information about the starting point, and the length of exposure to the SL-Hebrew but not the amount of exposure. The amount of exposure to each language was found to be a better predictor of vocabulary size than the length of exposure to each language (Thordardottir, 2019). Children with the same AOB and length of exposure can still vary on the actual exposure they get to the SL-Hebrew (Armon-Lotem and Meir, 2019).

In line with findings from the literature, there is a relation between parent reports on vocabulary size and reports on both exposure by others and use of SL-Hebrew by the child. SL-Hebrew use by the child correlated significantly with vocabulary production in Hebrew across the three bilingual populations and negatively with vocabulary production in the HL for the English and Russian speaking groups. The lack of correlation between child Hebrew use and the HL vocabulary production score of the French speakers could reflect the recency of this migration and the enclaved neighborhoods of French speakers, where French is supported outside the homes and not just in the home. Exposure to SL-Hebrew mostly does not correlate with the production of either HL or SL-Hebrew, but rather with the gap between the two languages. This shows that the more exposure a bilingual child receives to one language, the more he/she uses that language, and achieves higher vocabulary levels. Exposure, use and higher vocabulary levels in one language, inevitably reduce vocabulary levels in the other language.

TABLE 6 Correlation between exposure to Hebrew in books and screens and vocabulary levels in both Hebrew and the HL.

| | SL-Hebrew exposure in books | | | SL-Hebrew exposure in screens | | |
|---|---|---|---|---|---|---|
| | En-Heb | Fr-Heb | Rus-Heb | En-Heb | Fr-Heb | Rus-Heb |
| HL-Prod | −0.453** | ns. | ns. | −0.326 | ns. | ns. |
| HL-Comp | −0.461** | ns. | ns. | −0.380* | ns. | ns. |
| SL-Prod | ns. | 0.397* | ns. | ns. | $r = 0.337$ | ns. |
| SL-Comp | ns. | $r = 0.399$* | ns. | ns. | ns. | ns. |

*$p < 0.05$, **$p < 0.01$.

TABLE 7 Statistical reporting (including $\Delta R^2$ and $\Delta F$) of Hierarchical regressions with age and exposure to books and screens predicting vocabulary size across the three groups.

| | English-Hebrew speakers | | | | French-Hebrew speakers | | Russian-Hebrew speakers | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | English Production | | English Comprehension | | Hebrew Comprehension | | Hebrew Production | | Hebrew Comprehension | |
| | Estimate | $t$ | Estimate | $t$ | Estimate | $t$ | Estimate | $t$ | Estimate | $t$ |
| Step 1: age | | | | | | | | | | |
| Age | 4.51 | 2.146* | 2.54 | 1.89 | 4.37 | 2.22* | 3.67 | 2.885** | 2.62 | 3.47** |
| $R^2$ | 0.161 | | 0.13 | | 0.19 | | 0.24 | | 0.32 | |
| $F$ | 4.60* | | 3.57 | | 4.94* | | 8.32** | | 12.03** | |
| Step 2: exposure | | | | | | | | | | |
| Age | 4.12 | 2.13* | 2.29 | 1.86 | 4.23 | 2.19* | 3.59 | 2.95** | 2.72 | 3.37** |
| Language of screens | 69.78 | −0.77 | 38.76 | −1.11 | 48.39 | 0.54 | 47.23 | 1.11 | 35.68 | 0.55 |
| Language of books | 47.79 | −1.97 | 26.55 | −1.89 | 45.66 | 1.20 | 46.83 | 0.79 | 35.38 | 0.08 |
| $\Delta R^2$ | 0.36 | | 0.36 | | 0.331 | | 0.334 | | 0.33 | |
| $\Delta F$ | 4.13* | | 4.08* | | 3.13* | | 4.01* | | 3.92* | |

*$p < 0.05$, **$p < 0.01$. Only significant models are presented.

### 4.3. Exposure to literacy and screens

A differential effect was observed for exposure to books and screens in Hebrew. For English speakers, a negative relation was observed with English production and comprehension, while being not significant for Hebrew. Among the French-Hebrew speakers, exposure to Hebrew in books and screens is related to better Hebrew production and comprehension with no impact on HL-French. The different patterns for English speakers and French speakers might reflect the observed difference in the preference of reading, with English speakers reading more in HL than SL, and French speakers reading more in SL than HL, as well as the value attributed by the two populations for integration within the host society and academic system. With English being *lingua franca* supported in schools and academic studies, its speakers support the literacy in this language (including pro HL reading practices), while for French speakers, SL-Hebrew literacy is a key to academic integration. These observations require further research to test this hypothesis. Finally, for the Russian-Hebrew speaking group no effect was found for books and screens, perhaps due to the similar exposure received through these means for both the HL-Russian and the SL-Hebrew. These findings are in line with results from previous studies demonstrating the positive effect of book reading on the language in which reading is done, and the negative effect on the other language (Jiménez et al., 2006; Quiroz et al., 2010). Interestingly enough, regression analyses showed that while age explains relatively small portion of the variance in vocabulary size, across the three groups, the combination of age and exposure to stories and screens is a better predictor of vocabulary size and explains a large portion of its variance. These findings stress out the strong effect of exposure variables on vocabulary size.

## 5. Conclusion

The purpose of the present study was to investigate the developmental trajectories of the vocabularies of bilingual children from diverse bilingual populations. This study has shown that English-Hebrew, French-Hebrew, and Russian-Hebrew speakers demonstrate similar vocabulary levels as well as balanced bilingualism at the group level. This study further validates the use of the multicultural questionnaire (Ohana and Armon-Lotem, 2023) with various bilingual populations and sets the ground for future research with larger samples. Future research might want to address some of the limitations of this study. First, the sample size is relatively small and so future studies should aim at collecting data from a larger sample. Second, in terms of language exposure of children to their two languages, two important notes should be considered. Information was obtained from parents in relation to the quantity of exposure as an estimated time period with no measure of the frequency of exposure. In addition, no information was received about the quality of exposure to each language. These variables might account for the individual variability observed in each group, and should be addressed in future studies. Despite these limitations, using the multicultural questionnaire is likely to enable researchers as well as health professionals to better assess the vocabularies of bilingual children from different linguistic background, as well as children who are exposed to each of their languages to a different extent.

## Data availability statement

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

## Ethics statement

The studies involving human participants were reviewed and approved by Bar Ilan University. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

## Author contributions

OO and SA-L contributed to the design of the study and performed the statistical analysis. OO collected data and organized it and wrote the first draft of the manuscript. All authors contributed to the article and approved the submitted version.

## Funding

This research was supported by Bar-Ilan President's Doctoral Fellowship of Excellence, Bar-Ilan Impact Center for Multilingualism and Multiculturalism across the Life Span (MultiLinC), and the Israel Science Foundation (Grant No. 454/18).

## Acknowledgments

We would like to thank the Bilingualism Matters Israel team for their help with participant recruitment and data collection. Special thanks to Michael Isjaki for designing the online questionnaire and the ongoing support with the data. And finally, we would like to thank all the parents who participated in the study, allowing us to learn about their children's language.

## Conflict of interest

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

## Editor's note

Maria-José Ezeizabarrena edited the article in collaboration with Melita Kovacevic, University of Zagreb, Zagreb, Croatia.

## Publisher's note

## References


Abdelwahab, A. G. S., Forbes, S., Cattani, A., Goslin, J., and Floccia, C. (2021). An adaptation of the MacArthur-Bates CDI in 17 Arabic dialects for children aged 8 to 30 months. Language Learning and Development. 17, 425–446.

Abutbul-Oz, H., and Armon-Lotem, S. (2022). Parent questionnaires in screening for developmental language disorder among bilingual children in speech and language clinics. *Front. Educ.* 7:846111. doi: 10.3389/feduc.2022.846111

Altman, C., Burstein-Feldman, Z., Fichman, S., Armon-Lotem, S., Joffe, S., and Walters, J. (2021). Perceptions of identity, language abilities and language preferences among Russian-Hebrew and English-Hebrew bilingual children and their parents. *J. Multiling. Multicult. Dev.*, 42, 1–16. doi: 10.1080/01434632.2021.1974462

American Speech-Language-Hearing Association (ASHA). (2004). Preferred practice patterns for the profession of speech-language pathology. Available at: http://www.asha.org/policy/PP2004-00191

Armon-Lotem, S., Joffe, S., Abutbul-Oz, H., Altman, C., and Walters, J. (2014). Language exposure, ethnolinguistic identity and attitudes in the acquisition of Hebrew as a second language among bilingual preschool children from Russian-and English-speaking backgrounds. *Input Exp. Biling. Dev.* 13, 77–98. doi: 10.1075/tilar.13.05arm

Armon-Lotem, S., and Meir, N. (2019). The nature of exposure and input in early bilingualism. In HouwerA. De and L. Ortega (Eds.), *The Cambridge handbook of bilingualism* (pp. 193–212). Cambridge, UK: Cambridge University Press.

Armon-Lotem, S., and Ohana, O. (2017). A CDI study of bilingual English-Hebrew children– frequency of exposure as a major source of variation. *Int. J. Biling. Educ. Biling.* 20, 201–217. doi: 10.1080/13670050.2016.1179257

Armon-Lotem, S., Rose, K., and Altman, C. (2021). The development of English as a heritage language: the role of chronological age and age of onset of bilingualism. *First Lang.* 41, 67–89. doi: 10.1177/0142723720929810

Armon-Lotem, S., Walters, J., and Gagarina, N. (2011). The impact of internal and external factors on linguistic performance in the home language and in L2 among Russian-Hebrew and Russian-German preschool children. *Linguist. Approaches Biling.* 1, 291–317. doi: 10.1075/lab.1.3.04arm

Auza, A., Murata, C., and Peñaloza, C. (2023). Predictive validity of a parental questionnaire for identifying children with developmental language disorders. *Front. Psychol.* 14:2370. doi: 10.3389/fpsyg.2023.1110449

Bedore, L. M., Peña, E. D., García, M., and Cortez, C. (2005). Conceptual versus monolingual scoring: when does it make a difference? *Lang. Speech Hear. Serv. Sch.* 36, 188–200. doi: 10.1044/0161-1461(2005/020)

Bialystok, E. (2002). Acquisition of literacy in bilingual children: A framework for research. *Lang. Learn.* 52, 159–199. doi: 10.1111/1467-9922.00180

Boerma, T., and Blom, E. (2017). Assessment of bilingual children: what if testing both languages is not possible? *J. Commun. Disord.* 66, 65–76. doi: 10.1016/j.jcomdis.2017.04.001

Central Bureau of Statistics. (2021). The social survey on languages. Available at: https://www.cbs.gov.il/he/Surveys/ (Accessed September 12, 2022).

Dale, P. S., and Penfold, M. (2011). Adaptations of the MacArthur-Bates CDI into non-US English languages. Available at: (last accessed 04.23)http://mb-cdi.stanford.edu/documents/AdaptationsSurvey.Web.pdf.

De Houwer, A. (2015). Harmonious bilingual development: young families' well-being in language contact situations. *Int. J. Biling.* 19, 169–184. doi: 10.1177/1367006913489202

DeMayo, B., Kellier, D., Braginsky, M., Bergmann, C., Hendriks, C., Rowland, C. F., et al. (2021). Web-CDI: A system for online administration of the MacArthurBates communicative development inventories. *Lang. Dev. Res.* 1, 55–98. doi: 10.34842/kr8e-w591

Fenson, L., Marchman, V.A., Thal, D.J., Dale, P.S., Reznick, J.S., and Bates, E. (2007). *MacArthur-Bates communicative development inventories*. Paul H. Brookes (2nd ed.) Baltimore: Publishing Company.

Fenson, L., Marchman, V.A., Thal, D.J., Dale, P.H., Reznick, J.S., and Bates, E. (1991). *MacArthur- Bates CDI words and sentences*. Baltimore: Brookes Publishing.

Fenson, L., Pethick, S., Renda, C., Cox, J. L., Dale, P. S., and Reznick, J. S. (2000). Short-form versions of the MacArthur communicative development inventories. *Appl. Psycholinguist.* 21, 95–116. doi: 10.1017/S0142716400001053

Frank, M. C., Braginsky, M., Yurovsky, D., and Marchman, V. A. (2021). *Variability and consistency in early language learning: the wordbank project*. Cambridge, MA: MIT Press.

Gatt, D. (2007). Establishing the concurrent validity of a vocabulary checklist for young Maltese children. *Folia Phoniatr. Logop.* 59, 297–305. doi: 10.1159/000108336

Gendler-Shalev, H. and, & Dromi, E. (2021). The Hebrew web communicative development inventory (MB-CDI): lexical development growth curves. *J. Child Lang.*, 49, 486–502. doi: 10.1017/S0305000921000179

Heilmann, J., Ellis Weismer, S., Evans, J., and Hollar, C. (2005). Utility of the Macarthur-Bates communicative development inventory in identifying language abilities of late-talking and typically developing toddlers. *Am. J. Speech Lang. Pathol.* 14, 40–51. doi: 10.1044/1058-0360(2005/006)

Hoff, E. (2018). Bilingual development in children of immigrant families. *Child Dev. Perspect.* 12, 80–86. doi: 10.1111/cdep.12262

Hoff, E., Core, C., Place, S., Rumiche, R., Señor, M., and Parra, M. (2012). Dual language exposure and early bilingual development. *J. Child Lang.* 39, 1–27. doi: 10.1017/S0305000910000759

Hoff, E., Giguere, D., Quinn, J., and Lauro, J. (2018). The development of English and Spanish among children in immigrant families in the United States. [Article]. *Pensam. Educ.* 55, 1–17. doi: 10.7764/PEL.55.2.2018.1

Jiménez, T. C., Filippini, A. L., and Gerber, M. M. (2006). Shared reading within Latino families: an analysis of reading interactions and language use. *Biling. Res. J.* 30, 431–452. doi: 10.1080/15235882.2006.10162884

Junker, D. A., and Stockman, I. J. (2002). Expressive vocabulary of German-English bilingual toddlers. *Am. J. Speech Lang. Pathol.* 11, 381–394. doi: 10.1044/1058-0360(2002/042)

Kambanaros, M., and Grohmann, K. K. (2013). Profiling (specific) language impairment in bilingual children: preliminary evidence from Cyprus. *Solutions Assessment Biling.*, 146–174. doi: 10.21832/9781783090150-010

Kern, S. (2007). Lexicon development in French-speaking infants. *First Lang.* 27, 227–250. doi: 10.1177/0142723706075789

Maital, S. L., Dromi, E., Sagi, A., and Bornstein, M. H. (2000). The Hebrew communicative development inventory: language specific properties and cross-linguistic generalizations. *J. Child Lang.* 27, 43–67. doi: 10.1017/S0305000999004006

Marchman, V. A., and Martínez-Sussmann, C. (2002). Concurrent validity of caregiver/parent report measures of language for children who are learning both English and Spanish. *J. Speech Lang. Hear. Res.* 45, 983–997. doi: 10.1044/1092-4388(2002/080)

Miękisz, A., Haman, E., Łuniewska, M., Kuś, K., O'Toole, C., and Katsos, N. (2017). The impact of a first-generation immigrant environment on the heritage language: productive vocabularies of polish toddlers living in the UK and Ireland. *Int. J. Biling. Educ. Biling.* 20, 183–200. doi: 10.1080/13670050.2016.1179259

O'Toole, C., and Fletcher, P. (2010). Validity of a parent report for Irish-speaking toddlers. *First Lang.* 30, 199–217. doi: 10.1177/0142723709359237

O'Toole, C., Gatt, D., Hickey, T. M., Miękisz, A., Haman, E., Armon-Lotem, S., et al. (2017). Parent report of early lexical production in bilingual children: a crosslinguistic CDI comparison. *Int. J. Biling. Educ. Biling.* 20, 124–145. doi: 10.1080/13670050.2016.1179258

Ohana, O., and Armon-Lotem, S. (2023). Using a multicultural monolingual questionnaire to assess both languages of bilingual Children. [Manuscript in preparation].

Paradis, J. (2010). The interface between bilingual development and specific language impairment. *Appl. Psycholinguist.* 31, 227–252. doi: 10.1017/S0142716409990373

Pearson, B. Z., Fernández, S., and Oller, D. K. (1993). Lexical development in bilingual infants and toddlers: comparison to monolingual norms. *Lang. Learn.* 43, 93–120. doi: 10.1111/j.1467-1770.1993.tb00174.x

Place, S., and Hoff, E. (2011). Properties of dual language exposure that influence 2-year-olds' bilingual proficiency. *Child Dev.* 82, 1834–1849. doi: 10.1111/j.1467-8624.2011.01660.x

Quiroz, B. G., Snow, C. E., and Zhao, J. (2010). Vocabulary skills of Spanish—English bilinguals: impact of mother—child language interactions and home language and literacy support. *Int. J. Biling.* 14, 379–399. doi: 10.1177/1367006910370919

Remennick, L. (2003). Language acquisition as the main vehicle of social integration: Russian immigrants of the 1990s in Israel. *Int. J. Sociol. Lang.* 2003, 83–105. doi: 10.1515/ijsl.2003.057

Ring, E. D., and Fenson, L. (2000). The correspondence between parent report and child performance for receptive and expressive vocabulary beyond infancy. *First Lang.* 20, 141–159. doi: 10.1177/014272370002005902

Schwartz, M., Kozminsky, E., and Leikin, M. (2009). Socio-linguistic factors in second language lexical knowledge: the case of second-generation children of Russian-Jewish immigrants in Israel. *Lang. Cult. Curric.* 22, 15–28. doi: 10.1080/07908310802504119

Thordardottir, E. (2019). Amount trumps timing in bilingual vocabulary acquisition: effects of input in simultaneous and sequential school-age bilinguals. *Int. J. Biling.* 23, 236–255. doi: 10.1177/1367006917722418

Thordardottir, E., Rothenberg, A., Rivard, M. E., and Naves, R. (2006). Bilingual assessment: can overall proficiency be estimated from separate measurement of two languages? *J. Multiling. Commun. Disord.* 4, 1–21. doi: 10.1080/14769670500215647

Vershinina, E. A., Eliseeva, M. B., Lavrova, T. S., Ryskina, V. L., and Cejtlin, S. N. (2011). "Nekotorye normativy rečevogo razvitia detej ot 8 do 18 mesjatsev" in *Special'noe obrazovanie: tradicii i inovacii: Sbornik naučno-metodičeskix trudov s meždunarodnym učastiem* (St-Peterburg: RGPU im. A.I. Gertsena).

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 10 April 2024
DOI 10.3389/fpsyg.2024.1211548



OPEN ACCESS

EDITED BY
Maria-José Ezeizabarrena,
University of the Basque Country, Spain

REVIEWED BY
Jelena Kuvac Kraljevic,
University of Zagreb, Croatia
Grzegorz Krajewski,
University of Warsaw, Poland

*CORRESPONDENCE
Marie Pourquié
✉ mariepourquieweb@gmail.com

RECEIVED 24 April 2023
ACCEPTED 13 March 2024
PUBLISHED 10 April 2024

CITATION
Pourquié M, Etxeberria U, Lacroix H and Prévost P (2024) "HEGA": the Basque version of the PaBiQ parental questionnaire, for clinicians and educators working in the Basque multilingual environment. *Front. Psychol.* 15:1211548. doi: 10.3389/fpsyg.2024.1211548



# "HEGA": the Basque version of the PaBiQ parental questionnaire, for clinicians and educators working in the Basque multilingual environment

Marie Pourquié[1*], Urtzi Etxeberria[1], Hugues Lacroix[1] and Philippe Prévost[2]

[1]Centre National de la Recherche Scientifique, IKER (UMR 5478), Bayonne, France, [2]Université de Tours, INSERM, Imaging Brain & Neuropsychiatry iBraiN U1253, Tours, France

In this paper, we investigate the relevance of using a parental questionnaire (HEGA) to gather information on children's language experience in Basque and early language development in order to better interpret language performance in that language. Both this questionnaire and use of language assessment in Basque are needed in the Basque Country, where multilingualism is well attested. The questionnaire was developed after the PaBiQ with additional questions meant to reflect the Basque context, notably its schooling linguistic model. The HEGA was administered to the parents of 186 bilingual children of the Northern Basque Country (age 4;2–9;1) whose language skills in Basque were assessed via a new test battery targeting different linguistic domains (HIGA). Several significant correlations were found between exposure to, and use of Basque and performance in lexical and morphosyntactic production and comprehension. Mixed-effect regression analyses revealed that language experience in Basque, and particularly the fact of being schooled entirely in Basque, were strong predictors of lexical and morphosyntactic outcomes. In contrast, phonological performance, as measured by nonword repetition, appeared to be less impacted by language experience in Basque. Finally, two children were identified as being at risk of language impairment, due to low language performance in Basque despite extended language experience. These results have important implications for clinicians and educators, in particular for detecting language difficulties in Basque-speaking bilingual children. They also show the need for assessing language abilities in Basque for children growing up in a solid Basque-speaking environment.

KEYWORDS

**multilingualism, speech and language therapy, parental questionnaire, language exposure, Basque**

## 1 Introduction

Over the last 10 years or so, tremendous attention has been drawn to how best identify language impairment in bilingual children. This includes diagnosis of Developmental Language Disorder (DLD), a neurodevelopmental disorder involving persistent deficits in language that 'are not explained by another neurodevelopmental disorder or a sensory impairment or neurological condition' (ICD-11, World Health Organization, 2018). In many cases, language assessment tools are lacking in one or all languages of the child, and when language is assessed, the challenge is being able to disentangle low performance due to potential DLD or poor language experience, namely insufficiently long or rich exposure or use (Armon-Lotem et al., 2015).

The Basque Country is no exception to this situation. The 21st Century Basque society is typically multilingual, as Basque speakers also naturally acquire French and/or Spanish that are majority languages in the countries where Basque is spoken, i.e., France and Spain, in the Northern and Southern Basque Country, respectively.[1] According to the Basque Government (2016) carried out by the Basque Government (Basque Autonomous Community, BAC), the Government of Navarre, and the Office Public de la Langue Basque, the number of Basque speakers has been increasing since the 1990s, which in 2016 amounted to 751,527 (28.4%) people aged over 16. This reflects an increase of 223,000 individuals since 1991, notably among young people (16–24), of whom 55% speak Basque.[2] This trend is observed across the whole Basque Country, including the Northern Basque Country (NBC) in France, which has seen the largest increase of young Basque speakers compared to other age groups (19% in 2016 *vs* 11% in 1996). This increase results from the recovered prestige of the Basque language all over the territory and from language teaching policies, specifically the opening (officially in the 80s) of primary schools across the Basque Country, where teaching is carried out in Basque either partly (so-called bilingual schools) or entirely (immersion schools). In 2018–19, such schools received 40% of primary school enrolment in the NBC. The number of children enrolled in Basque schools has constantly been increasing since their creation. Importantly, while all Basque speakers are bilinguals (mainly with French or Spanish), there is a wide range of bilingual profiles across the Basque Country, with some speakers being more dominant in Basque than French or Spanish, others displaying the reverse pattern, and others having no language dominance.

Despite the evident bilingual nature of the Basque Country, language assessment of Basque-speaking children does not include Basque, due to a lack of (standardized) evaluation tools and defined developmental milestones in this language. Moreover, investigation of bilingual language development in Basque-speaking children has mainly focused on longitudinal case studies of Basque-Spanish pre-school age children (Elosegi, 1998; Larrañaga, 2000; Barreña, 2003) and on bigger data set collected from parental questionnaires – the Basque MB-CDI parental questionnaire[3] (García et al., 2008, 2011, 2014; Barreña et al., 2008a,b; Ezeizabarrena et al., 2013). It has rarely addressed children growing up in the context of a neurodevelopmental disorder, including DLD. There is thus a lack of knowledge about language development in Basque-speaking children and about specific difficulties that Basque children with DLD may have (Pourquié, 2017).

When Basque is assessed in clinical contexts, evaluation is largely qualitative, based on spontaneous verbal interaction with the child and not in a normalized manner, i.e., by referring to norms on typically developing (bilingual) children. Instead of Basque, language assessment usually targets the other language of the child, typically French or Spanish, using (standardized) evaluation tools available in that language and the norms associated to those tools. Two main problems arise from this situation: first, since Basque speakers are for the vast majority – if not all – bilingual, using language assessment tools with monolingual norms is inadequate, with high risks of over- and underdiagnosis of language impairment (Thordardottir, 2015a). Second, using language evaluation tools created in French or Spanish to assess language in Basque-speaking children does not allow for assessment of Basque specific grammatical features. There are indeed major differences between Basque, which remains a language isolate with no known relatives and uncertain origins, and French and Spanish, which are both Romance languages. For instance, in what concerns grammatical features, Basque is a SOV language but French and Spanish are SVO languages; French and Spanish use prepositions that are free morphemes while Basque uses case marking corresponding to bound morphemes; Basque verbs agree with both subjects and objects while French and Spanish verbs only agree with subjects; relative clauses precede the noun in Basque while they follow it in French and Spanish; French and Spanish use clitics while Basque does not; etc. Further language evaluation in Basque would thus provide a more accurate picture of the children's language abilities, which would lead to more appropriate language support. Testing children in Basque would also be feasible since many Speech-Language Therapists (SLTs) and educators in the Basque Country are bilingual.

While bilingual norms exist for some assessment tools (e.g., in German, Schulz and Tracy, 2011, and Lebanese Arabic, Zebib et al., 2017), this is the exception rather than the rule, including for French and Spanish. Many SLTs, especially those providing services to multilingual populations, call for norms on bilingual language development (Volpin et al., 2020). When such norms are not available, obtaining information on the language experience of the child, as well as his/her early language development is crucial (see Kašćelan et al., 2022). At minimum, this should provide SLTs and educators with information as to whether low language performance is obtained despite long and sustained experience in the language in question (which may be indicative of language difficulties or DLD), or whether it is accompanied with low language experience (which may not be indicative of language difficulties or DLD).

Parent questionnaires have been extensively used in research to gather background information, including evaluation of children's language skills, and their adaptation to professional settings has been shown to be particularly relevant. Among the best-known parental questionnaires is the MacArthur-Bates

1 Basque is not only spoken in the Basque Country, but also in various Basque communities around the world, in the so-called Basque diaspora, i.e., people of Basque origin living outside the borders of the Basque Country. We will not consider these communities here.

2 A new survey containing data collected in 2021 was published in 2023. For the moment, only the data from the BAC is available, i.e., not counting the data from the NBC and Navarre, and they continue showing an increase of Basque speakers in people aged over 16years. According to this new survey (2021eko Inkesta Soziolinguistikoa) 36.2% of the population of the BAC speaks Basque, which corresponds to 680.629 people, i.e., about 50.000 more than in 2016.

3 Notice that there are five CDI instruments adapted to Basque: long CDI-1 & CDI-2 (Barreña et al., 2008a); Short CDI-1 & CDI-2 (García et al., 2008, 2011) and CDI-3 (Garcia et al., 2014).

Communicative Developmental Inventories (MB-CDI, Fenson et al., 1994), for which parents are asked to document their children's lexical and grammatical abilities and gesture production. Significant correlations have been reported between the MB-CDI and direct language measures, showing the reliability of parents' ratings (Feldman et al., 2005; Heilmann et al., 2005). Parent evaluation of their children's language abilities has also been shown to be a strong predictor of DLD or language difficulties in children growing up in a monolingual setting (Callu et al., 2003; Surakka et al., 2023) or in a bilingual environment (Restrepo, 1998; Paradis et al., 2010). Parent assessment of their children's language skills is particularly useful when one of the languages cannot be assessed directly.

Using a Basque version of the MB-CDI (Barreña et al., 2008a), studies on language development in Basque-speaking children have reported a significant impact of exposure to Basque on the development of lexical and morphosyntactic abilities. Barreña et al. (2008b) investigated 947 children aged 16 to 30 months. The sample was divided into three groups according to the percentage of Basque present in the children's immediate environment (> 90%, 60–90%, and < 60%). In general, for both lexical knowledge and mean length of utterance, the group with the least exposure to Basque performed lower than the two other groups, especially as of age 27–28 months.

Previous research on language development in bilingual children has shown that different language domains may be impacted differently by language experience (see Unsworth, 2016 and Paradis, 2023 for overviews). For the lexicon, amount of exposure and socio-economic status (e.g., as measured by the mother's education level) have been found to be particularly predictive of performance on both lexical production and comprehension (Cobo-Lewis et al., 2002; Golberg et al., 2008; Scheele et al., 2010). Likewise, quantity and quality of input can significantly influence performance and outcomes in morphosyntax, in production and comprehension, although the extent of this impact may differ across grammatical phenomena, owing, e.g., to their morphological or syntactic complexity and to the tasks being used (Paradis, 2010; Thomas et al., 2014; Thordardottir, 2015b).

In some studies, SES has been found to be a predictive factor of morphosyntactic outcomes as well (De Cat, 2021). Another aspect of quality of exposure that has drawn the attention of researchers concerns the proficiency level of the parents. In particular lower language performance (in lexicon and morphosyntax) by children has been found to correlate with lower degrees of nativeness of their parents in the language (Paradis and Jia, 2017; Unsworth et al., 2019). In Barreña et al.'s (2008b) study on Basque-speaking children, parents' knowledge of Basque was also reported to affect the results: children with both parents speaking Basque were found to outperform those with only one parent speaking Basque for both lexicon and morphosyntax. In contrast, development of phonological skills in bilingual children seem to be less affected by language experience, especially when phonology is assessed via tools that control for lexical knowledge, such as nonword repetition tasks (Thordardottir, 2014; Dos Santos and Ferré, 2018).

One parental questionnaire now used in a variety of bilingual contexts is the Parents of Bilingual Children Questionnaire (PaBiQ, Tuller, 2015) developed during COST Action IS0804 (*Language Impairment in a Multilingual Society: Linguistic Patterns and the Road to Assessment,* 2009–2013).[4] This questionnaire, available in 20 languages, documents variables known to impact bilingual language development, such as age of onset, quantity and quality of exposure, as well as early exposure (before the age of four). It also asks parents to evaluate the language skills of their children in all of his/her languages. Furthermore, a section is devoted to the child's early history, such as the age of first word and age of first sentence, since delay in language emergence is observed in children with DLD (Rice et al., 2008; see also ICD-11, 2018), and whether the parents were concerned about language development in their children.[5] The PaBiQ allows for the calculation of several composite scores and indexes about the risk of language impairment, early language exposure (before age 4), current language skills, and quantity and quality of current exposure and use, which can be used to better interpret language performance by the child.

In particular, studies using the PaBiQ have shown the relevance of the No risk index (and its component, the Positive early development index), which has been found to be a significant predictor of language performance across different language domains and in different bilingual settings. Based on stepwise multiple regression analyses on results from the PaBiQ and sentence and nonword repetition tasks administered to Bi-TD and Bi-DLD children in France and Germany, Tuller et al. (2018) found that the No risk index – and not the measures of language experience – was the main predictor, and often the only predictor of language performance, in both countries (see also Boerma and Blom, 2017).

Studies integrating bilingual children with DLD have found a differential impact of language exposure on morphosyntactic abilities compared to children with typical development (TD). De Almeida et al. (2017) investigated language skills in French in bilingual children (ages 5–8) with different first languages (Arabic, Portuguese and Turkish). After being tested in both of their languages via standardized tests, the bilingual children were divided into two groups, depending on whether they were deemed to be at risk of DLD (the Bi-DLD group) or not, i.e., showing Typical Development (the Bi-TD group). Using information collected from the PABIQ, significant correlations were found in the Bi-TD group between performance on a sentence repetition task and two composite scores of the PABIQ: use of French at home and language richness (in French). These correlations did not arise in the Bi-DLD group, suggesting that Bi-DLD children's morphosyntactic skills did not improve as language experience increased (see also Armon-Lotem and Meir, 2016). Interestingly, no bilingualism variables, included the two that were reported to impact morphosyntactic performance, were found to significantly correlate with performance in nonword repetition.

To our knowledge no parental questionnaire is commonly used in the Basque Country for collecting information on the multilingual experience of Basque-speaking children. Some clinical, educational, and research centers use their own questionnaires (Anderson et al.,

4 http://www.bi-sli.org/

5 The PaBiQ was originally inspired by two parental questionnaires, the Alberta Language and Development Questionnaire (ALDeQ; Paradis et al., 2010), focusing on variables related to bilingualism, and the Alberta Language Environment Questionnaire (ALEQ; Paradis, 2011), which documents L1 development and risk factors of language impairment.

2019) and some questionnaires seem to be restricted to specific studies (e.g., Barreña et al., 2008b). Therefore, there is a need for the development of an easy-to-use parental questionnaire to be shared among the Basque community, in order to improve research, education and clinical practices adapted to the Basque multilingual environment.

In order to address the issue of the identification of atypical language development in Basque and provide adequate clinical services to Basque-speaking children with DLD, a parental questionnaire and a language assessment tool in Basque were developed by an interdisciplinary group of SLTs and researchers in psycholinguistics within the scope of the Nouveaux Commanditaires Sciences (NCS) program,[6] which encourages a dialog between researchers and citizens, from a participative research perspective.

The aim of this paper is to present HEGA (*Haur Elebidunen Gurasoentzako Galdetegia* 'Parental Questionnaire for Bilingual Children'), the Basque adaptation of the PABIQ questionnaire and its specificities, and to show its usability by clinicians, educators or researchers as a complementary tool to language assessment in Basque. In particular, this study sought to establish which measures of language experience correlate with, and predict, language skills in Basque.

We first hypothesized that the language skills (in Basque and French) estimated by the parents would be significantly correlated with the results on the different factors of language experience obtained throughout the questionnaire (early experience, length of exposure, language use, language richness, the parents' proficiency in their languages, SES, and schooling model). The No risk index was also expected to impact language proficiency, as estimated by the parents.

We also hypothesized that language outcomes in Basque should be predicted by language experience in Basque (early exposure, language use at home, language richness and schooling model), with lesser impact on performance on phonology than on lexicon and morphosyntax. As to which measures from HEGA best predicted language skills and outcomes (in different language domains), this remained an open question, different predictors having been found in the literature, based on different methodological designs.

# 2 Materials and methods

## 2.1 Participants

HEGA questionnaires were completed by 186 parents of children enrolled in two types of schools in the NBC: Basque immersive schools where teaching is all in Basque ($n = 136$) and bilingual schools ($n = 50$) where half the teaching is in Basque and half in French. The children (88 boys and 98 girls) were aged 4;2 to 9;1 ($M = 6;10$, $SD = 1;4$) and had all received exposure to Basque and French. Fifteen of them had been exposed to another language, mainly Spanish (Spanish $n = 8$, English $n = 5$, Portuguese $n = 1$, and Wolof $n = 1$). Regarding age of exposure, 93/186 children (50%) were simultaneous Basque/French bilinguals and 77 (41.4%) were sequential bilinguals, including 40 who were exposed to the other language after age three. Among the 77 sequential bilinguals, 57 were first exposed to French (and in 33 cases exposure to Basque started after age three) and 20 were first exposed to Basque (and in seven cases, exposure to French started after age three). In the remaining 16 cases of our sample (8.6%), information about the age of first contact to Basque and/or French was missing. Regarding SES, all but eight children came from families where both parents had received post-secondary or university education, and only one child came from a family where both parents had received secondary education. Further information on the participants is presented in Section 3.1.

## 2.2 Materials

In order to address the issue of the identification of atypical language development in Basque and provide adequate clinical services to Basque-speaking children with DLD, as already mentioned above, the HEGA questionnaire and a language assessment tool in Basque named HIGA (*HIzkuntza Garapenaren Azterketa* 'Language Development Assessment') were developed by an interdisciplinary group of SLTs and researchers in psycholinguistics.

### 2.2.1 HIGA: an oral language assessment tool in Basque

The HIGA assessment tool targets children aged 4–8 years. It has been normed on data collected from 254 children enrolled in immersive schools and percentile standards have been defined using the following scale: 95, 75, 50, 25 and 5. It contains 13 oral production and comprehension tasks targeting phonology, lexicon and morphosyntax (see Supplementary material 1). Five of them were selected for the present study in order to assess children's phonological, lexical, and morphosyntactic abilities in production and comprehension: Non-word repetition, Object naming, Lexical recognition, Sentence production and Sentence comprehension. For homogeneity's sake the other tasks were not analyzed. Following Tomblin et al.'s (1996) recommendations, children were considered to be at risk of having DLD when their language performance was low (below −1.25 SD) in at least two different domains.

#### 2.2.1.1 Object naming task

Object naming aims at assessing semantic knowledge and lexical access in production. Semantic knowledge corresponds to words' meaning and lexical access to words' phonological form retrieval. The task includes 32 items selected on the basis of their phonological structure (with or without coda) and their frequency. Five word types were established (see Supplementary material 2). The selected words had to have limited dialectal variability. For instance, the word *sagua* 'mouse' was included because it shows little variability in Basque; by comparison, the word *xinaurria* 'ant' was not selected as it can be said in different manners (*xinaurria, inurria, txindurria*, etc.). Color pictures depicting various objects, such as fruits, vegetables and animals are presented to the child (one at a time) who is instructed to say what the picture represents. If after 10 s the child has not produced the word, the examiner gives him/her a phonological cue in the form of the first sound of the target word. No other cue is allowed.

6 https://www.joursavenir.org/activities/ncs/en

#### 2.2.1.2 Lexical recognition task

The lexical recognition task aims at assessing semantic knowledge and lexical access in comprehension. The task includes 16 nouns: 8 nouns taken from the Object naming task in order to assess whether some items that are not produced may nonetheless be understood, and 8 nouns related to various semantic categories and involving small dialectal variability. However, for four items two dialectal variants were considered as targets (*gauainara/xaguxarra* 'bat'; *saskia/otarra* 'basket'; *eskorga/karretila* 'wheelbarrow'; *ganita/labana* 'knife'). Each word is presented orally to the child, along with four pictures. The child is instructed to point to the picture corresponding to the oral stimulus. Among the four pictures, one is the target picture, one is the picture of a semantically related item (semantic distractor) and two depict unrelated items.

#### 2.2.1.3 Non-word repetition task

Non-word repetition (NWR) is generally used to assess phonological perception and production skills, and has been shown to be sensitive to DLD (Conti-Ramsden et al., 2001). The method followed to create the task items followed Ferré and dos Santos (2015). All segments used in the task are so-called language independent sounds, meaning that they are present in the majority of the languages of the world. Repetition of these segments should therefore be little impacted by language experience. The length of the items does not exceed three syllables, so as to minimize memory effects. Syllables are either simple (CV) or complex (i.e., involving a branching onset – CCV – or a coda – CVC). Moreover, in order to control lexical knowledge, it was made sure that no item resembled a real word in Basque, neither in standard Basque nor in any dialect. A total of 12 test items (see Supplementary material 3) are included, preceded by two training items. The task is based on color pictures depicting monsters, whose names correspond to the nonwords children are asked to repeat. All oral stimuli are pre-recorded.

#### 2.2.1.4 Sentence production and comprehension tasks

The HIGA morphosyntactic production and comprehension tasks focus on verb agreement with singular and plural subjects, direct objects and indirect objects. A total of seven inflected verb forms (verb auxiliaries) are tested twice for a total of 14 stimuli (see Supplementary material 4). The production task is a sentence completion task based on mini-scenes represented by two color pictures. The two pictures are quite similar but they differ in singular and plural agreement. The child is asked to describe all the pictures. If children have difficulties completing a sentence, examiners are allowed to provide the lexical verb, but not the auxiliary (in any form). The 14 test items of the task are preceded by one example.

The same 14 inflected verb forms are assessed in comprehension through a picture-selection task. This is a picture-sentence matching task in which children are asked to identify, from a group of four color pictures, the picture that best corresponds to a sentence presented orally. The four pictures are quite similar but differ in terms of verb agreement form or transitivity (see Supplementary material 5). Two training items are presented before the 14 test items. As in the NWR task, all oral stimuli are pre-recorded.

### 2.2.2 The HEGA parental questionnaire

The HEGA questionnaire is a Basque adaptation of the PaBiQ. It was specifically designed to be used in clinical and educational contexts in the Basque Country, aiming to gather information on the child's multilingual environment and his/her language developmental milestones.

The questionnaire is divided into nine sections for a total of 47 questions (see Table 1; Supplementary material 6 for the full list of questions). It exists in three versions (Basque, French, and Spanish) thus enabling a wide range of users to fill it in.

The first seven sections are similar to the PaBiQ (Tuller, 2015), while the eighth section was added to gather information on the child's education (e.g., the school linguistic model the child was enrolled in), as this is very relevant to the Basque Country. Finally, the ninth and last section was added to gather parents' free comments. All in all, a total of 11 questions were added to the original PaBiQ. Eight were taken from the Basque adaptation of the MacArthur-Bates CDI (García et al., 2014) and were added to the sections on general information ($n = 4$), language use in the family ($n = 1$), and information about the parents ($n = 3$). Moreover, following suggestions by the SLTs within the NCS action, one question was added to the section on language richness regarding the language in which the child is told stories (in addition to asking about the language in which the child reads), and one response choice was added to the question "Before age 4, did you worry about your child's language?" in the early language

TABLE 1 Organization of the HEGA.

| Section | Information collected | Number of questions |
|---|---|---|
| I. General information | Date and country of birth, country of residence, gender, languages currently spoken by the child, preferred language, number of siblings and position in siblings | 11 |
| II. Early language history (before age 4) | Age of the child's first word and 1st sentence, parental concerns about the child's language development, age of first contact with each language, exposure to each language before age 4 | 6 |
| III. Current skills | Child's proficiency in each language, as estimated by the parents | 5 |
| IV. Language use in the family | Languages used between the child and the parents, siblings, and other caretakers | 6 |
| V. Language richness | Languages used with friends and during specific activities, e.g., reading and watching TV | 3 |
| VI. Information about the parents | Birth country, language used at work, education and self-rated proficiency in each of their languages | 6 |
| VII. Language difficulties in the family | Difficulties concerning reading, spelling, speaking, and understanding | 3 |
| VIII. Educational information | Grade; schooling system; skip or repeat grade; | 6 |
| IX. Free comments | Any comments that the parents would like to share, either in general or on the questionnaire. | 1 |

history section. The answer "YES *after* age 3/4" was added to the original "YES/NO" answer. Finally, two questions about exposure to, and use of code-switching were added to the section on language use within the family, as this can provide relevant information regarding language experience in multilingual societies (Kašćelan et al., 2022).

As in the original PaBiQ questionnaire, different composite indexes and scores can be calculated: (1) a no risk index, (2) an early language exposure ratio (before age 4), (3) a parents' estimate of their child's current language skills score, (4) a score of language exposure and use at home, and (5) a score of language exposure and use with friends and during activities (also called language richness) (see Tuller, 2015). These indexes are explained below.

#### 2.2.2.1 No risk index

The no risk index brings together all the risk factors whose influence on the chances of a child having DLD is well-established: age of the early stages of acquisition (first words and first sentences), parental concern for the child's language and existence of language difficulties within the family, with points associated with each answer (see Tuller, 2015). Three age range options are proposed in the questionnaire for age of first word (≤ 15 months, 16–24 months, and ≥ 25 months) and for age of first sentence (≤ 24 months, 25–30 months, and ≥ 31 months). In both cases, emergence of first word and first sentence in typical development corresponds to the first two options. Six points are associated with the first option, four with the second option, and none for the third option. Regarding parental concern, if none is expressed, an extra two points is added. Otherwise, no additional point is awarded. Finally, absence of language difficulties in the family (with respect to reading, understanding others, and expressing oneself) corresponds to 9 points. The maximum number of points is 23. Although the no risk index has proved to be sensitive to DLD, as seen above, the score below which concern should be raised as to a potential risk of language impairment is yet to be established. In Tuller et al. (2015), which involved bilingual children with or without DLD, all children with DLD had a no risk index score of 18 and below.[7]

#### 2.2.2.2 Early language exposure ratio (before age 4)

In Section 3 of HEGA, parents are asked to select the contexts in which their children were exposed to each of their languages before age 4 (e.g., with the mother, the father, the grandparents, a nanny, etc.). An overall number of contexts of language exposure is thus obtained, combining all contexts of exposure to all of the child's languages. The early language exposure ratio is the percentage of contexts in which the child is exposed to a particular language with respect to the overall number of contexts of exposure.

#### 2.2.2.3 Parental estimation of current skills

This index combines, for each language, the scores of the five questions appearing in Section 4 of the questionnaire devoted to the children's current language skills, as estimated by their parents. The answers are presented in a four-point Likert scale, which are associated to 0 to 3 points, the score of 3 corresponding to the highest (estimated) skills. The maximum number of points is 15. A score of 10 points (with five answers corresponding to 'good' language skills – 2 points) and above may be considered to be indicative of typical development for the language concerned.

#### 2.2.2.4 Score of language exposure and use at home

For each language of the child, parents are asked to rate the frequency of exchanges between the child and the mother, the father, the siblings, and any other caregiver (Section 5 of HEGA). Possible answers appear on a five-point Likert scale, ranging from 0 (never) to 4 (very often/always). The maximum score is 16 points.

#### 2.2.2.5 Language richness score (=score of language exposure and use with friends and during activities)

Language richness combines two sets of questions appearing in Section 6 of the questionnaire: (1) two questions about frequency of exchanges, for each language, between the child and his/her friends, and friends of the family (with answers presented on a five-point Likert scale, as above), and (2) four questions about frequency of language use during particular activities, i.e., reading, being read to, watching TV, and storytelling [with answers presented on a five-point Likert scale, ranging from 0 (never) to 4 (very often/always)]. The maximum score is 24 points.

## 2.3 Procedures and scoring

### 2.3.1 General procedures

Eight schools in the NBC accepted to take part in the study: five Basque immersive schools and three French-Basque bilingual schools. Consent forms explaining the nature of each task and asking for permission to record data anonymously were obtained from all participating families. Testing always took place in a silent room at the child's school during school hours. The examiner sat in front of the child and used the HIGA stimulus book to show the pictures, a computer to display the auditory stimuli in comprehension and repetition tasks, and a recording device to record the whole session. A total of 13 examiners participated in the data collection process. They were all members of the NCS group and were all familiar with the testing material as they actively participated in its design. Procedures, unanimously approved by the NCS group, were enforced concerning the application of a stop criterion (following 5 non answers in a row)

7 Note that in contrast to the PaBiQ, an additional option ("I do not know") was inserted as an answer to the questions targeting age of first word and first sentence in the HEGA questionnaire, as some parents may find it difficult to answer. When the parents chose the "I do not know" option for both questions, the no risk index was not calculated (14 cases). When the "I do not know" option was selected for one of the two questions, it was decided that some points should be awarded if the answers to the other two questions (age of 1st sentence/word and parental concerns) were congruent. In particular, 6 or 4 points were assigned if the age of 1st word/sentence was deemed to be typical (i.e., not appearing after 25 and 31months respectively) AND parents expressed no concerns (22 cases). Zero points were assigned if the age of 1st word/sentence was above these cut-offs AND parents expressed concerns (2 cases). When incongruency was observed in the answers of the two questions (e.g., 1st word/sentence above 25 or 31months AND no parental concerns), no adjustment was made and the no risk index was not calculated (2 cases). Note also that when parents answered positively to the answer about parental concerns *after* age 3 or 4, which was added to the original PaBiQ, it was decided to attribute the same score as when they answered positively to the question about concerns *before* age 3 or 4 present in the original PaBiQ (i.e., 0 points).

and how many times an item could be presented in the NWR task (only once). Due to the COVID-19 sanitary crisis, the examiners and the children older than 6 years old were required to wear a mask. This did not, however, hamper the testing as the stimuli of the comprehension tasks were displayed from a computer. No unintelligible answers were reported due to the mask. The tasks were administered in a fixed order. At the end of the study, a present was sent to each school and addressed to the children and the staff that participated in the study.

The HEGA questionnaire was made available on-line to the parents through a Google form. In case some parents preferred to fill it in on paper, a printed version was made available. It was not possible to interview each family one by one due to the high number of participants and the COVID-19 crisis. This prevented us from checking that all the questions were answered and from providing help for questions that were felt to be unclear. Some space was left at the end of the questionnaire for parents to share comments or express what they had not understood. Eighteen parents left a comment. Only one complained that the survey was very long, and none reported any unclear question. In general, the parents used the comment section to share their multilingual experience, to explain the reason why they did not use Basque at home, to request Basque support for parents at school, and to explain the type of difficulty their child had (e.g., difficulty with pronunciation).

### 2.3.2 Data scoring and analysis

In all the HIGA tasks that were used in this study, a correct answer was coded as 1 and an incorrect answer as 0. In the NWR task, a score of 1 corresponded to an item that was repeated identically as the stimulus. Otherwise, a score of 0 was awarded. In the Object naming task, correct answers that were produced spontaneously were scored as 1 and those for which help was provided were first coded as h1 and then scored as 1 (see Supplementary material 7 for examples). In the Sentence production task, production of inflected verb forms other than the expected one was counted as correct (so, as 1) in some specific cases: e.g., when the forms did not clash with the targeted tense (e.g., using the present progressive form *erortzen ari da/dira* 'he/they are falling down' for the present tense *erortzen da/dira* 'he/they fall(s)'); when the forms corresponded to dialectal variants of the target forms too (e.g., *ematen dako* '(s)he gives it to him/her' in Low Navarrese Basque used for *ematen dio* in Standard Basque).

Correlation analyses (controlled for age) were performed to explore the link between measures of language experience and language skills and outcomes. To model the relationship between accuracy in HIGA linguistic tasks (as indexed by the response variables from each task) and the potential predictors from the HEGA parental questionnaire, generalized linear mixed-effects regression analyses were performed. This kind of model can account for various predictors at the same time, for variance with either continuous or categorical predictors, and for random variation, using random effects. Taking into account random variation allowed us to control for sampling effects in our population (due to unbalanced groups of participants: 136 in immersive schools versus 50 in bilingual schools) and in our items (due to specific properties of each item). Therefore, in the models used in our analyses, Participant and Item are always included as random effects alongside the fixed effects. In all analyses, we tested whether the following variables were significant predictors of language performance: Age, Gender, Schooling Model, Length of Exposure in Basque, Total Basque used at home, Richness in Basque, Early exposure to Basque ratio (before age 4), No risk index, Positive early development index, Parental estimation of current skills in Basque, Mother's education, and Father's education.

## 3 Results

### 3.1 General results from the HEGA

General findings on the bilingualism variables documented by HEGA, for each language, appear in Table 2.

As can be seen, pairwise comparisons yielded significant differences between the two languages, including age of contact (earlier for French), and length of exposure (LoE), language use at home, language richness, and proficiency levels of the parents (all larger in French). In contrast, no difference between the two languages were found on variables targeting exposure during the first four years (early exposure, early contact – total, and percentage of exposure to Basque or French). No difference was found either between language skills in Basque and French as estimated by the parents.

However, these general results masked important differences in our population based on schooling type. Significant differences for all

TABLE 2 General results (Mean and SD) from the main HEGA measures for Basque and French.

| | Basque | French | *t* | *df*[a] | *p* |
|---|---|---|---|---|---|
| Age of first contact (months) | 9.3 (15.6) | 2.6 (9.7) | 4.381 | 167 | < 0.001 |
| Frequency of early exposure (0–4 scale) | 3.0 (1.1) | 3.2 (1.1) | −1.248 | 183 | 0.214 |
| Early contacts total (max. 8 pts) | 4.6 (2.0) | 4.6 (1.9) | 0.145 | 175 | 0.885 |
| Early language exposure ratio (before age 4) | 48.9 (20.7) | 49.6 (19.1) | −0.657 | 178 | 0.512 |
| Length of exposure (months) | 72.9 (21.0) | 79.8 (19.4) | −4.381 | 167 | < 0.001 |
| Language used at home (max. 16 pts) | 8.8 (4.9) | 10.3 (5.0) | −2.124 | 183 | 0.035 |
| Language richness (max. 24 pts) | 10.6 (5.6) | 13.8 (7.4) | −3.496 | 183 | < 0.001 |
| Proficiency level (mother) (0–4 scale) | 2.2 (1.4) | 3.8 (0.4) | −14.343 | 181 | < 0.001 |
| Proficiency level (father) (0–4 scale) | 1.9 (1.5) | 3.8 (0.5) | −16.293 | 181 | < 0.001 |
| Current skills (max. 15 pts) | 9.3 (3.6) | 9.7 (3.6) | −1.039 | 183 | 0.300 |

[a]*df* differed according to the number of parents who provided the expected information.

TABLE 3 Partial correlation analyses with current skills in Basque or French estimated by the parents (controlled for age).

| | Basque | | French | |
|---|---|---|---|---|
| | *r* | *p* | *r* | *p* |
| Age | −0.115 | 0.118 | 0.232 | 0.001 |
| Age of first contact | −0.365 | <0.001 | −0.246 | < 0.001 |
| Frequency of early exposure (0–4 scale) | 0.476 | < 0.001 | 0.544 | < 0.001 |
| Early contacts total (max. 8 pts) | 0.472 | < 0.001 | 0.435 | < 0.001 |
| Early language exposure ratio (before age 4) | 0.472 | < 0.001 | 0.522 | < 0.001 |
| Length of exposure (months) | 0.365 | < 0.001 | 0.248 | < 0.001 |
| Language used at home (max. 16 pts) | 0.547 | < 0.001 | 0.626 | < 0.001 |
| Language richness (max. 16 pts) | 0.496 | < 0.001 | 0.586 | < 0.001 |
| Proficiency level (mother) (0–4 scale) | 0.528 | < 0.001 | 0.311 | < 0.001 |
| Proficiency level (father) (0–4 scale) | 0.445 | < 0.001 | 0.233 | 0.002 |
| Education level (mother) | 0.186 | 0.015 | 0.044 | 0.549 |
| Education level (father) | 0.061 | 0.412 | 0.027 | 0.715 |
| No risk index (max. 23 pts) | 0.189 | 0.013 | 0.353 | < 0.001 |

bilingualism variables were found between children enrolled in Basque immersion schools versus Basque/French bilingual schools, suggesting large variability in our sample (see Supplementary Table S1). Children enrolled in immersion schools had significantly wider language experience in Basque than children in bilingual schools. The reverse was found for French. In addition, parents of children in immersion schools tended to have significantly higher proficiency in Basque than parents of children in bilingual schools. Children did not significantly differ in terms of length of exposure to Basque, even though children in immersion schools had significantly earlier exposure to this language. Yet, children in bilingual schools tended to be significantly older ($M = 90.0$ months, $SD = 11.3$) than those in immersion schools ($M = 79.4$ months, $SD = 17.6$), accounting for similar LoE to Basque in each school system. It is important to note that no differences were found between the two groups regarding early developmental milestones and the no risk index (see Supplementary Table S2).

Finally, nineteen participants (17/136 in Basque immersive schools and 2/50 in French-Basque bilingual schools) had a low no risk index and could be considered to be at risk of DLD. Particular attention was paid to these children regarding the results on bilingualism variables and language performance presented below.

## 3.2 Analyses internal to the HEGA questionnaire

Given the wide age range of the child participants, partial correlation analyses (controlling for age) were performed between the estimated proficiency skills of the children, in both Basque and French, and measures of language experience and the No risk index.

As can be seen in Table 3, estimated proficiency measures significantly correlated with all measures of early exposure, LoE, Language used at home and Language richness, for both languages. For both Basque and French, Language used at home and Language richness yielded the strongest correlations (0.547 and 0.496, respectively, for Basque, and 0.626 and 0.586, respectively, for French). There were also significant correlations between the parents' (self-rated) proficiency levels in each language and the estimated language skills of the children, with higher correlation coefficients observed for Basque. Note that the parents' (self-rated) proficiency levels were also significantly correlated with Age of first contact, LoE, Language use at home, and Language richness, for Basque and French (see Supplementary material 8). For both the mother and the father, the highest correlation coefficients involved use of either language at home. In contrast, Table 3 shows that parent's education did not strongly correlate with the children's estimated language skills. One significant correlation was observed in Basque for Mother's education, but it was low (0.186). Finally, there was a significant correlation between the measures of proficiency estimated by the parents in both Basque and French and the No risk index, with a higher correlation coefficient for French. Note that the correlation coefficient increased to 0.326 ($p < 0.001$) for Basque when Basque used at home and Basque richness were controlled for. For French, the correlation coefficient climbed to 0.367 ($p < 0.001$) with these two variables controlled for.

Finally, we compared the children's language skills, as estimated by the parents, according to schooling model. For Basque, language skills were estimated to be significantly higher for children in immersion schools than in bilingual schools [$t(182) = 4.626$, $p < 0.001$, Cohen's $d = 0.777$]. The reverse obtained for language skills in French, with large effect sizes [$t(182) = -5.126$, $p < 0.001$, Cohen's $d = -0.861$].

## 3.3 Analyses involving measures of the HEGA questionnaire and performance in the HIGA language tasks

In this section, we cross the data from the HEGA questionnaire with the performance on the five language tasks in Basque. We first report on correlation analyses between the HEGA measures and the language measures (Table 4). We then compare language performance in children enrolled in bilingual vs. immersion schools. Finally, we present the results of multiple regression analyses to identify predictors of language performance.

As can be seen in Table 4, mild to strong correlations were found between several HEGA measures and performance in Object naming (lexical production), Lexical recognition (lexical comprehension), Morphosyntactic production, and Morphosyntactic comprehension (with lower correlation coefficients for comprehension). Across the four tasks, Basque used at home, Basque richness, and early contacts in Basque yielded the strongest correlations. Proficiency level in Basque (for both parents) were also significantly correlated to performance on the four tasks. In contrast, Education level was mildly correlated with language performance on these tasks (for the father only). For NWR, fewer significant correlations were observed, and the coefficient correlations were lower than what was found for the other tasks. In particular, there was no significant correlation between NWR performance and Basque richness ($p = 0.107$), and significance was barely reached with Basque used at home ($p = 0.038$). Age was the strongest variable with which NWR performance was significantly correlated ($r = 0.476$). The score for current language skills in Basque as estimated by the parents significantly correlated with all individual responses (moderately so with Morphosyntactic comprehension, $r = 0.256$), except for NWR ($p = 0.493$). As to the No risk index, the only significant correlation was found with the performance on NWR, but here again, the correlation coefficient was moderately high ($r < 0.300$).

Finally, the schooling model was found to affect language performance. Children in immersion schools performed significantly better than children in bilingual schools in all tasks [Object naming: $t(181) = 15.138$, $p < 0.001$; Lexical recognition: $t(181) = 7.023$, $p < 0.001$; NWR: $t(181) = -3.426$, $p < 0.001$; Morphosyntactic production: $t(179) = 9.249$, $p < 0.001$; Morphosyntactic comprehension: $t(180) = 5.718$, $p < 0.001$]. Scattered plots for all analyses detailed above can be found in the Supplementary material.

We now turn to multi-regression analyses. All the potential predictors (see Section 2.3.2) were included for all tasks using stepwise regression, but only the most relevant regression model is reported.

### 3.3.1 Object naming

The variables related to Basque exposure had a significant impact on the performance at Object naming. In particular, as can be seen in Table 5, the schooling model (with Basque only used as the baseline) negatively predicted accuracy, meaning that children in French-Basque bilingual schools scored significantly lower than the children in immersive Basque schools. In contrast, LoE to Basque and Use of Basque at home positively predicted accuracy. Finally, the level of education of both parents was also predictive of accuracy on the Object naming task, but only significantly so for the father. Looking

TABLE 4 Partial correlation analyses (controlled for age) between HEGA measures and performance in the five language tasks in Basque (Lexical production, Lexical comprehension, NWR, Morphosyntactic production, and Morphosyntactic comprehension).

| | Lexical prod. | | Lexical comp. | | NWR | | Morphosynt. prod. | | Morphosynt. comp. | |
|---|---|---|---|---|---|---|---|---|---|---|
| | *r* | *p* | *r* | *p* | *r* | *p* | *r* | *p* | *r* | *p* |
| Age | −0.083 | 0.265 | 0.089 | 0.227 | 0.476 | < 0.001 | 0.154 | 0.038 | 0.306 | < 0.001 |
| Age of first contact (Basque) | −0.426 | < 0.001 | −0.317 | < 0.001 | 0.190 | 0.013 | −0.394 | < 0.001 | −0.205 | 0.007 |
| Frequency of early exposure (0–4 scale) | 0.499 | < 0.001 | 0.409 | < 0.001 | −0.135 | 0.066 | 0.483 | < 0.001 | 0.267 | < 0.001 |
| Early contacts total (Basque) (max. 8 pts) | 0.638 | < 0.001 | 0.562 | < 0.001 | −0.088 | 0.239 | 0.480 | < 0.001 | 0.282 | < 0.001 |
| Early exposure to Basque ratio (before age 4) | 0.645 | < 0.001 | 0.528 | < 0.001 | −0.134 | 0.071 | 0.497 | < 0.001 | 0.279 | < 0.001 |
| Length of exposure (months) | 0.426 | < 0.001 | 0.317 | < 0.001 | −0.190 | 0.013 | 0.394 | < 0.001 | 0.205 | 0.007 |
| Basque used at home (max. 16 pts) | 0.707 | < 0.001 | 0.590 | < 0.001 | −0.154 | 0.038 | 0.547 | < 0.001 | 0.274 | < 0.001 |
| Basque Richness (max. 24 pts) | 0.717 | < 0.001 | 0.522 | < 0.001 | −0.119 | 0.107 | 0.577 | < 0.001 | 0.290 | < 0.001 |
| Proficiency level (mother) (0–4 scale) | 0.603 | < 0.001 | 0.520 | < 0.001 | −0.135 | 0.06 | 0.483 | < 0.001 | 0.283 | < 0.001 |
| Proficiency level (father) (0–4 scale) | 0.493 | < 0.001 | 0.479 | < 0.001 | −0.043 | 0.565 | 0.454 | < 0.001 | 0.252 | < 0.001 |
| Education level (mother) | −0.022 | 0.762 | −0.003 | 0.973 | −0.063 | 0.391 | −0.033 | 0.659 | −0.044 | 0.551 |
| Education level (father) | 0.158 | 0.033 | 0.108 | 0.147 | −0.108 | 0.144 | 0.214 | 0.004 | 0.223 | 0.003 |
| Total current skills (Basque) (max. 15 pts) | 0.442 | < 0.001 | 0.348 | < 0.001 | 0.051 | 0.493 | 0.409 | < 0.001 | 0.256 | < 0.001 |
| No risk index (max. 23 pts) | −0.029 | 0.705 | 0.040 | 0.605 | 0.254 | < 0.001 | −0.029 | 0.712 | 0.074 | 0.336 |

TABLE 5 Regression analysis on HEGA measures and performance in object naming (lexical production).

| Term | Estimate | SE | z value | *p*-value |
|---|---|---|---|---|
| (Intercept) | −3.82 | 1.357 | −2.814 | 0.005 |
| Schooling model (Bilingual Basque/French) | −2.309 | 0.276 | −8.358 | < 0.001 |
| Length of exposure (Basque) | 0.015 | 0.005 | 2.83 | 0.005 |
| Basque used at home (max. 16 points) | 0.182 | 0.029 | 6.383 | < 0.001 |
| Education level (mother) | 0.48 | 0.289 | 1.66 | 0.097 |
| Education level (father) | 0.379 | 0.168 | 2.251 | 0.024 |

further at the data, we found that there was more variation in the father's educational level, with the mother's level being generally higher.

### 3.3.2 Lexical recognition

As was the case for Object naming, the schooling model (with Basque only used as the baseline) negatively predicted accuracy on lexical recognition, while Use of Basque at home was a positive predictor of performance on this task (see Table 6). No other predictors were identified for the performance on lexical recognition.

### 3.3.3 Non-word repetition

For the NWR task, none of the tested variable, except for Basque use at home, significantly predicted performance (see Table 7). The schooling model and the No risk index had a significant impact on the Akaike information criterion (AIC) index of the model, but alone they were not significant.

### 3.3.4 Sentence production

The schooling model also had a significantly negative effect on performance in Sentence production (Table 8), with individual responses in the French-Basque bilingual school group being significantly lower than in the immersive Basque school group. The total amount of Basque used at home and the no risk index also predicted performance significantly. As to the father's educational level, although its impact was not statistically significant, it improved the fit of the model according to the AIC.

### 3.3.5 Sentence comprehension task

As seen for all the other language measures, children in Basque immersion schools had significantly higher performance on Sentence comprehension than children attending bilingual schools (Table 9). Age and the father's education level also had a significant (and positive) impact on Sentence comprehension.

We end this section with some findings on the 19 children with a low No risk index, ranging from 9 to 17 (out of 23). For 14 of these children, the parents estimated their language skills in Basque to be low (below 10 out of 15), including eight with low skills in the other language as well. Using Tomblin et al.'s (1996) recommendation for identifying DLD (see above), we found that five of these 19 children had language performance below −1.25 SD in at least two different domains. However, these children were among the

TABLE 6 Regression analysis on HEGA measures and performance in lexical recognition (lexical comprehension).

| Term | Estimate | SE | z value | P-value |
|---|---|---|---|---|
| (Intercept) | 1.205 | 0.367 | 3.279 | 0.001 |
| Schooling model (Bilingual Basque/French) | −0.711 | 0.222 | −3.198 | 0.001 |
| Basque used at home (max. 16 points) | 0.121 | 0.024 | 5.116 | < 0.001 |

TABLE 7 Regression analysis on HEGA measures and performance in NWR.

| Term | Estimate | SE | z value | P-value |
|---|---|---|---|---|
| (Intercept) | 1.481 | 0.974 | 1.52 | 0.13 |
| Schooling model (Bilingual Basque/French) | 0.364 | 0.318 | 1.144 | 0.25 |
| Basque used at home (max. 16 points) | −0.093 | 0.03 | −3.153 | 0.002 |
| No risk index (max. 23 points) | 0.077 | 0.042 | 1.823 | 0.068 |

TABLE 8 Regression analysis on HEGA measures and performance in sentence production.

| Term | Estimate | SE | z value | P-value |
|---|---|---|---|---|
| (Intercept) | −4.679 | 2.033 | −2.302 | 0.021 |
| Schooling model (Bilingual Basque/French) | −2.411 | 0.484 | −4.986 | < 0.001 |
| Basque used at home (max. 16 points) | 0.166 | 0.047 | 3.506 | < 0.001 |
| No risk index (max. 23 points) | 0.139 | 0.074 | 1.893 | 0.058 |
| Education level (father) | 0.565 | 0.309 | 1.826 | 0.068 |

TABLE 9 Regression analysis on HEGA measures and performance in sentence comprehension.

| Term | Estimate | SE | z value | P-value |
|---|---|---|---|---|
| (Intercept) | −2.632 | 0.934 | −2.818 | 0.005 |
| Age (in years) | 0.544 | 0.091 | 5.98 | < 0.001 |
| Schooling model (Bilingual Basque/French) | −1.831 | 0.262 | −6.984 | < 0.001 |
| Education level (father) | 0.433 | 0.177 | 2.446 | 0.014 |

 

youngest in our sample (younger than 4;6). In three cases, experience with Basque during the first four years was low (e.g., fewer than 50% Basque exposure). For the two other children, exposure to Basque was much higher, and scores for Basque use at home and Basque richness were at ceiling, which could be cause for concern. For the other children whose skills were rated low by their parents, language performance was above −1.25 SD in at least four of the tasks.

# 4 Discussion and conclusion

In this paper, we investigated the relevance of using a parental questionnaire (HEGA) to gather information on children's language experience in Basque and early language development in order to better interpret language performance in that language. Both this questionnaire and use of language assessment in Basque are needed in the Basque Country, where multilingualism is well attested. The questionnaire was developed after the PaBiQ (Tuller, 2015) with additional questions meant to reflect the Basque context, notably its schooling linguistic model. The language tasks came from a new language battery in Basque targeting different linguistic domains (HIGA). A total of 186 children of the NBC (age 4–9) and their parents participated in the study.

As hypothesized, significant correlations were found between several measures of bilingualism factors and the language skills of the children, as estimated by their parents or assessed via language tasks (except for NWR, see below). High correlations were particularly observed with language use at home and richness, which confirms what has been reported in the literature. Parent proficiency also correlated with estimated language skills and all individual responses, which is akin to recent findings pointing to the importance of the quality of language exposure for language development (see Paradis, 2023). These results also confirm what has been found for Basque on younger children regarding the impact of exposure to Basque and parent proficiency in Basque on language abilities (Barreña et al., 2008b). Language experience played a lesser role for NWR, with lower correlation coefficients than for all other language measures, which confirms what has been reported in the literature (Thordardottir, 2014; de Almeida et al., 2017). During the development of the NWR task of the HIGA, particular care had been paid to making the items as less word-like as possible, which included taking into account different Basque dialects. Lexical knowledge was thus well controlled for this task, which can explain the very low impact of language experience on the results. In contrast to measures of language experience, few significant correlations were found between the no risk index and language measures. The only ones involved parent ratings of language skills and NWR performance. Note, however, that some of the children with a low no risk index and low individual language performance were quite young (below 4;6 years) with little Basque experience, thus preventing any conclusion about a potential language disorder. It should also be noted that the studies that reported on a significant impact of the no risk index on language performance all involved a group of TD children and a group of children with DLD, in contrast to our study. Moreover, some parents may have had difficulties answering some of the questions directly impacting on the calculation of the no risk index, such as age of first words and age of first sentences. These questions have been identified as particularly complicated for some parents (i.e., what should be considered as a word or a combination of words may not appear to be very transparent for many), which is the reason why some authors advocate for a person-to-person administration of questionnaires such as the PaBiQ (Tuller, 2015). However, due to the pandemic, the parents filled in the questionnaire by themselves on-line, and did not benefit from any assistance. This notwithstanding, two of the young children with a low no risk index in our study had high ratings for Basque use at home and Basque richness, which may be cause for concern. A clear research perspective involves the recruitment of French/Basque children that receive speech-language therapy in order to investigate the effect of language experience in their language performance in Basque, as well as the effect of the no risk index.

As announced in the introduction, an open question remained as to which measures from HEGA could best predict language skills and outcomes in different language domains because different predictors were found in the literature based on different methodological designs. Of all the predictors investigated in this study, the schooling system in which the children were enrolled came out systematically, with children attending immersion schools outperforming children in bilingual schools in all language measures. Large and significant differences between the two school models were found with respect to many measures of language experience in our study, including use of Basque at home, richness in Basque, exposure to Basque until age four, and parent proficiency in Basque. We take the results of the regression analyses to reflect this difference. In short, it is the combination of the different measures of language experience, which comes out as the main predictor of language performance. Among the other potential predictors, SES, as measured by parent education, only played a significant role in the results for Object naming (lexical production), which has been widely reported in the literature. It was not identified as a major predictor for the other language tasks (including lexical comprehension) and for the language skills as estimated by the parents. This could be explained by the little variability in SES in our population sample, which mainly consisted of individuals with post-secondary and university education. Further studies on language development in Basque-speaking children should be more inclusive, expanding recruitment to more under-privileged communities.

Finally, we found that performance in production tended to be impacted by language experience to a larger extent than comprehension. Studies have reported that bilinguals, as compared to monolinguals, present a larger gap between production and comprehension abilities, in both their languages, with comprehension typically surpassing production. This asymmetry is commonly called the 'expressive-receptive gap' (Gibson et al., 2014) and has been attributed to a weakness in lexical-semantic links, which has a stronger impact on production than on comprehension and is highly influenced by language exposure (Gollan et al., 2008; Keller et al., 2015). This would explain why Object naming and Sentence production are particularly affected by language exposure in our study. Another factor that might contribute to the expressive-receptive gap is linguistic typology. Anderson et al. (2019) presented evidence for such a gap in the grammatical abilities of school-aged Basque-Spanish bilingual children and found it to be wider regarding grammatical structures that are not shared between the languages of the child. This proposal would merit to be investigated more thoroughly, for example by comparing bilinguals with a combination of languages that are either typologically related or unrelated, and with a comparable amount of language experience in the L2.

In short, the present study has shown that the parental questionnaire HEGA is a useful tool as a complement of language assessment in Basque, allowing better interpretation of children's linguistic abilities by taking into account their multilingual environment. More specifically, it was shown that language experience in Basque, and particularly the fact of being schooled entirely in Basque, was the best predictor of lexical and morphosyntactic outcomes. In contrast, phonological skills appeared to be less impacted by exposure to, and use of Basque. Finally, two children were identified as being at risk of language impairment, which further shows that crossing information from the HEGA questionnaire and the HIGA tools can be particularly useful for identifying potential language disorders that would be the manifestation of underlying developmental deficits.

The results of the present study have important implications for clinicians and educators. First, the information provided by parents is coherent, suggesting that they can be taken into account when a decision has to be made about whether a child is in need of specific support or not. The HEGA also appears to be user-friendly, as shown by the absence of negative comments regarding its length or the complexity of some of its questions. Second, the fact that measures of language experience in Basque may impact on language performance means that language scores should not be considered on their own; they should also be put in perspective with information related to the linguistic environment of the child, which goes in the same direction as findings on language assessment in bilingual children growing up in different multilingual contexts (Armon-Lotem et al., 2015). Including a subpart on the schooling system (immersive or bilingual) into the questionnaire proved to be crucial in this respect. Regarding children from immersive schools, whose number is steadily increasing these last years (from 2004 to 2021, there was an 86% increase in the number of children receiving immersive teaching in the immersive schools; OPLB 2021), the information obtained from HEGA reveals that they tend to be more exposed to Basque than to French. Therefore, it is absolutely necessary that SLTs be able to assess the language skills of these children with language tasks in Basque, which should lead to more reliable diagnosis. As SLTs opt for assessing children from bilingual schools in French rather than in Basque because French is mostly their dominant language, the reverse, i.e., assessing children from immersive schools in Basque rather than in French, only seems natural. Third, our results show that performance on NWR, being less impacted by language experience, is particularly useful for identifying language disorder in multilingual contexts. As to children's performance in vocabulary and morphosyntax, it should be interpreted along with the information collected from HEGA, especially Basque use at home and Basque richness. This is a very important message to convey to professionals in need of assessing language in bilingual children, given the challenge that they face. The next step is to conduct a study involving SLTs and educators to investigate the usefulness of HEGA in their daily practice for identifying children with DLD or in need of language support.

## Editor's note

Maria-José Ezeizabarrena edited the article in collaboration with Melita Kovacevic, University of Zagreb, Zagreb, Croatia.

## Data availability statement

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

## Author contributions

UE participated in the Basque adaptation of the PaBiQ questionnaire, wrote the introduction, and took part in discussions regarding the results of the study. HL participated in the Basque adaptation of the PaBiQ questionnaire (Spanish version in particular), and running all the analyses of the paper, creating graphs, and wrote the methods and results sections. PP took part in discussions regarding the results of the study, wrote all the parts of the manuscript (introduction, methods, results, and discussion), and shared his knowledge on the French PaBiq and his experience within the LITMUS community. MP took the initiative to collaborate with researchers and SLTs to adapt the PabiQ for Basque in order to gather information on the multilingual environment of the participants that would be tested on the HIGA oral language assessment tool, described the materials, she and her colleagues have developed within the scope of the program Nouveaux Commanditaires Sciences along with a dozen Speech and Language Therapists, and wrote all the parts of the manuscript (introduction, methods, results, and discussion). All authors contributed to the article and approved the submitted version.

## Funding

This research was supported by the following grants: EDELAB (Région Nouvelle-Aquitaine et Fondation GEROA); HEHGAN (GECT Eurorégion Aquitaine-Euskadi-Navarre); LABA ANR-22-CE28-0024-02 (ANR).

## Acknowledgments

We would like to express our gratitude to the editors for inviting us to write this paper as part of the Research Topic. A big thank you to the Association of Basque SLTs (Euskal Ortofonisten Elkartea) that helped us pass the tests. Thanks also to Nouveaux Commanditaires Sciences (https://www.joursavenir.org/activities/ncs) and Fondation de France for their help.

## Conflict of interest

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

## Publisher's note

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

## Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2024.1211548/full#supplementary-material

## References


Anderson, R. M., Giezen, M., and Pourquié, M. (2019). Basque-Spanish bilingual children's expressive and receptive grammatical abilities. *Linguis. Approach. Bilingualism* 9, 687–709. doi: 10.1075/lab.17034.and

Armon-Lotem, S., de Jong, J., and Meir, N. (Eds.). (2015). *Assessing multilingual children. Disentangling bilingualism from language impairment*. Bristol: Multilingual Matters.

Armon-Lotem, S., and Meir, N. (2016). 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.* 51, 715–731. doi: 10.1111/1460-6984.12242

Barreña, A. (2003). Euskararen sistema fonologikoaren jabekuntzari buruzko oharrak. *Pierre Lafitteren Sortzearen Mendemugako Biltzarra. Euskaltzaindiaren XV.* eds. J. M. Makazaga and B. Oyahrçabal Euskaltzaindiaren XV. Biltzarra. Euskaltzaindia. 161–179.

Barreña, A., Ezeizabarrena, M. J., and García, I. (2008b). "Influence of the linguistic environment on the development of the lexicon and grammar of Basque bilingual children" in *A portrait of the young in the new multilingual Spain*. eds. C. Pérez-Vidal, M. Juan-Garau and A. Bel (Clevedon, UK: Multilingual Matters), 86–110.

Barreña, A., Garcia, I., Ezeizabarrena, M. J., Almgren, M., Arratibel, N., Olano, I., et al. (2008a). "MacArthur-Bates" in *Komunikazio Garapena Neurtzeko Zerrenda. Euskarara egokitua. Erabiltzaileentzako gida eta eskuliburu teknikoa* (Bilbao: Udako Euskal Unibertsitatea)

Basque Government (Eusko Jaurlaritza), Government of Navarre (Nafarroako Gobernua), and Office Public de la Langue Basque (2016). VI. Inkesta Soziolinguistikoa: Euskararen Eremu Osoa. Available at: https://www.irekia.euskadi.eus/uploads/attachments/9954/VI_INK_SOZLG-EH_eus.pdf?1499236557

Boerma, T., and Blom, E. (2017). Assessment of bilingual children: what if testing both languages is not possible? *J. Commun. Disord.* 66, 65–76. doi: 10.1016/j.jcomdis.2017.04.001

Callu, D., Jacquier-Roux, M., Cusin, F., Giannopulu, I., and Dellatolas, G. (2003). Pertinence du repérage par les parents des retards de langage chez l'enfant entre quatre et six ans. *Arch. Pediatr.* 10, 1061–1067. doi: 10.1016/j.arcped.2003.09.034

Cobo-Lewis, A. B., Pearson, B., Eilers, R. E., and Umbel, V. C. (2002). "Effects of bilingualism and bilingual education on oral and written Spanish skills: a multifactor study of standardized test outcomes" in *Language and literacy in bilingual children*. eds. D. K. Oller and R. E. Eilers (Clevedon: Multilingual matters), 64–97.

Conti-Ramsden, G., Botting, N., and Faragher, B. (2001). Psycholinguistic markers for specific language impairment (SLI). *J. Child Psychol. Psychiatry* 42, 741–748. doi: 10.1111/1469-7610.00770

de Almeida, L., Ferré, S., Morin, E., Prévost, P., dos Santos, C., Tuller, L., et al. (2017). Identification of bilingual children with specific language impairment in France. *Linguistic Approaches Bilingualism* 7, 331–358. doi: 10.1075/lab.15019.alm

De Cat, C. (2021). Socio-economic status as a proxy for input quality in bilingual children? *Appl. Psycholinguist.* 42, 301–324. doi: 10.1017/S014271642000079X

Dos Santos, C., and Ferré, S. (2018). A nonword repetition task to assess bilingual Children's phonology. *Lang. Acquis.* 25, 58–71. doi: 10.1080/10489223.2016.1243692

Elosegi, K. (1998). *Kasu eta preposizioen jabekuntza-garapena haur elebidun batengan [Non published thesis]*. Leioa: University of the Basque country.

Ezeizabarrena, M. J., Barnes, J., García, I., Barreña, A., and Almgren, M. (2013). "Using parental report assessment for bilingual preschoolers: the Basque experience" in *Solutions for the assessment of bilinguals*. ed. V. C. Mueller Gathercole (Clevedon: Multilingual Matters), 57–80.

Feldman, H. M., Dale, P. S., Campbell, T. F., Colborn, D. K., Kurs-Lasky, M., Rockette, H. E., et al. (2005). Concurrent and predictive validity of parent reports of child language at ages 2 and 3 years. *Child Dev.* 76, 856–868. doi: 10.1111/j.1467-8624.2005.00882.x

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., et al. (1994). Variability in early communicative development. *Monogr. Soc. Res. Child Dev.* 59, i–185. doi: 10.2307/1166093

Ferré, S., and dos Santos, C. (2015). Comment évaluer la phonologie des enfants bilingues? *Lidil* 51, 11–34. doi: 10.4000/lidil.3678

García, I., Arratibel, N., Barreña, A., and Ezeizabarrena, M.-J. (2008). Adaptation of the MacArthur-Bates inventories to the Basque language: communicative development in children between 8 and 30 months of age. *Infancia y Aprendizaje* 31, 411–424. doi: 10.1174/021037008786140940

García, I., Barreña, A., Almgrem, M., Arratibel, N., Barnes, J., and Ezeizabarrena, M. J. (2011). MacArthur-Bates Communicatives development inventories tresnaren bertsio laburren egokitzapena euskarara: Komunikazio Garapena Neurtzeko Zerrendaren bertsio laburrak. *Uztaro* 79, 49–73.

García, I., Barreña, A., Ezeizabarrena, M. J., Almgren, M., Arratibel, N., and Barnes, J. (2014). Haur euskaldunen komunikazio-garapena neurtzen 30–50 hilabete bitartean: MacArthur–Bates CDI-III tresnaren euskal bertsioa. *Uztaro* 88, 33–72.

Gibson, T. A., Peña, E. D., and Bedore, L. M. (2014). The receptive–expressive gap in bilingual children with and without primary language impairment. *Am. J. Speech Lang. Pathol.* 23, 655–667. doi: 10.1044/2014_AJSLP-12-0119

Golberg, H., Paradis, J., and Crago, M. (2008). Lexical acquisition over time in minority first language children learning English as a second language. *Appl. Psycholinguist.* 29, 41–65. doi: 10.1017/S014271640808003X

Gollan, T. H., Montoya, R. I., Cera, C., and Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: aging, bilingualism, and the weaker links hypothesis. *J. Mem. Lang.* 58, 787–814. doi: 10.1016/j.jml.2007.07.001

Heilmann, J., Ellis Weismer, S., Evans, J., and Hollar, C. (2005). Utility of the MacArthur-Bates communicative development inventory in identifying language abilities of late-talking and typically developing toddlers. *Am. J. Speech Lang. Pathol.* 14, 40–51. doi: 10.1044/1058-0360(2005/006)

Kašćelan, D., Prévost, P., Serratrice, L., Tuller, L., Unsworth, S., and De Cat, C. (2022). A review of questionnaires quantifying bilingual experience in children: do they document the same constructs? *Biling. Lang. Congn.* 25, 29–41. doi: 10.1017/S1366728921000390

Keller, K., Troesch, L. M., and Grob, A. (2015). A large receptive-expressive gap in bilingual children. *Front. Psychol.* 6:1284. doi: 10.3389/fpsyg.2015.01284

Larrañaga, P. (2000). Ergative Sprachen, akkusative Sprachen: Der Erwerb des Kasus bei bilingualen Kindern. *Dans Ergative Sprachen Akkusative Sprachen*. doi: 10.31819/9783964561848

Office Public de la Langue Basque/Euskararen Erakunde Publikoa. (2021). Enseignement de et en langue basque: Évolution 2004–2021 et bilan de la rentrée 2021. Available at: http://www.ikasbi.com/wp-content/uploads/2022/02/OPLB-RENTRE%CC%81E2021-BILAN-E%CC%81volutions.pdf

Paradis, J. (2010). Bilingual children's acquisition of English verb morphology: effects of language exposure, structure complexity, and task type. *Lang. Learn.* 60, 651–680. doi: 10.1111/j.1467-9922.2010.00567.x

Paradis, J. (2011). Individual differences in child English second language acquisition: comparing child-internal and child-external factors. *Linguistic Approaches Bilingualism* 1, 213–237. doi: 10.1075/lab.1.3.01par

Paradis, J. (2023). Sources of individual differences in the dual language development of heritage bilinguals. *J. Child Lang.* 50, 793–817. doi: 10.1017/S0305000922000708

Paradis, J., Emmerzael, K., and Sorenson Duncan, T. (2010). Assessment of English language learners: using parent report on first language development. *J. Commun. Disord.* 43, 474–497. doi: 10.1016/j.jcomdis.2010.01.002

Paradis, J., and Jia, R. (2017). Bilingual children's long-term outcomes in English as a second language: language environment factors predict individual differences in catching up to monolinguals. *Dev. Sci.* 20, 1–15. doi: 10.1111/desc.12433

Pourquié, M. (2017). "Erreurs produites par des enfants bilingues à développement typique et atypique du langage: qu'est-ce qui les distingue?" in *Atypies langagières de l'enfance à l'âge adulte. Apport de la psycholinguistique et des neurosciences cognitives (pp. 51–75)*. eds. I. Bogliotti and L. Dujour (Belgium: De Boeck Superieur publishers)

Restrepo, A. (1998). Identifiers of predominately Spanish-speaking children with language impairment journal of speech. *Lang. Hear. Res.* 41, 1398–1411. doi: 10.1044/jslhr.4106.1398

Rice, M. L., Taylor, C. L., and Zubrick, S. R. (2008). Language outcomes of 7-year-old children with or without a history of late language emergence at 24 months. *J. Speech Lang. Hear. Res.* 51, 394–407. doi: 10.1044/1092-4388(2008/029)

Scheele, A. F., Leseman, P. P. M., and Mayo, A. Y. (2010). The home language environment of monolingual and bilingual children and their language proficiency. *Appl. Psycholinguist.* 31, 117–140. doi: 10.1017/S0142716409990191

Schulz, P., and Tracy, R. (2011). *Linguistische Sprachstandserhebung–Deutsch als Zweitsprache (LiSE-DaZ)*. Göttingen: Hogrefe.

Surakka, S., Vehkavuori, S.-M., Saaristo-Helin, K., and Stolt, S. (2023). What kind of information do early parental report instruments provide on language ability at 3; 6 when used at 2;0? A longitudinal comparison study. *Front. Psychol.* 14:1206949. doi: 10.3389/fpsyg.2023.1206949

Thomas, E. M., Williams, N., Jones, L. A., Davies, S., and Binks, H. (2014). Acquiring complex struc-tures under minority language conditions: bilingual acquisition of plural morphology in welsh. *Biling. Lang. Congn.* 17, 478–494. doi: 10.1017/S1366728913000497

Thordardottir, E. (2014). "The typical development of simultaneous bilinguals: vocabulary, mor-phosyntax and language processing in two age groups of Montreal preschoolers" in *Input and experience in bilingual development*. eds. T. Grüter and J. Paradis (Amsterdam: John Benjamins), 141–160.

Thordardottir, E. (2015a), Proposed diagnostic procedures for use in bilingual and cross-linguistic contexts. In S. Armon-Lotem, JongJ. de and N. Meir (Eds.), *Assessing multilingual children: Disentangling bilingualism from language impairment* (pp. 331–358). Bristol: Multilingual Matters.

Thordardottir, E. (2015b). The relationship between bilingual exposure and morphosyntactic development. *Int. J. Speech Lang. Pathol.* 17, 97–114. doi: 10.3109/17549507.2014.923509

Tomblin, J. B., Records, N. L., and Zhang, X. (1996). A system for the diagnosis of specific language impairment in kindergarten children. *J. Speech Hear. Res.* 39, 1284–1294. doi: 10.1044/jshr.3906.1284

Tuller, L. (2015). Clinical use of parental questionnaires in multilingual contexts. In S. Armon-Lotem, JongJ. de and N. Meir (eds), *Assessing multilingual children: Disentangling bilingualism from language impairment* (pp. 299–328). Bristol: Multilingual Matters.

Tuller, L., Abboud, L., Ferré, S., Fleckstein, A., Prévost, P., dos Santos, C., et al. (2015). "Specific language impairment and bilingualism: assembling the pieces" in *Language acquisition and development: Proceedings of GALA 2013*. eds. C. Hamann and E. Ruigendijk (Newcastle: Cambridge Scholars Publishing), 533–567.

Tuller, L., Hamann, C., Chilla, S., Ferré, S., Morin, E., Prevost, P., et al. (2018). Identifying language impairment in bilingual children in France and in Germany. *Int. J. Lang. Commun. Disord.* 53, 888–904. doi: 10.1111/1460-6984.12397

Unsworth, S. (2016). "Quantity and quality of language input in bilingual language development" in *Lifespan perspectives on bilingualism*. eds. E. Nicoladis and S. Montanari (Berlin: De Gruyter), 136–196.

Unsworth, S., Brouwer, S., de Bree, E., and Verhagen, J. (2019). Predicting bilingual preschoolers' patterns of language development: degree or non-native input matters. *Appl. Psycholinguist.* 40, 1189–1219. doi: 10.1017/S0142716419000225

Volpin, L., Weck, G. D., and Rezzonico, S. (2020). Enfants bilingues et prise en charge logopédique: Panorama des pratiques déclarées. *Glossa-Les cahiers de L'unadreo* 129, 16–31.

World Health Organization. (2018). International classification of Deseases-11. Available at: https://icd.who.int

Zebib, R., Henry, G., Khomsi, A., Messarra, C., and Hreich, E. (2017). *Evaluation du langage oral chez l'enfant libanais: ELO-L*. Antelias: Liban Tests Editions.