Research Topics

Surveillance of language development in pre-school children

Edited by
Johannes Fellinger, Daniel Holzinger and David Saldaña

Published in
Frontiers in Pediatrics
Frontiers in Psychology
Frontiers in Public Health

ISSN 1664-8714
ISBN 978-2-83250-914-2
DOI 10.3389/978-2-83250-914-2

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

Surveillance of language development in pre-school children

Topic editors

Johannes Fellinger — Hospitaller Brothers of Saint John of God Linz, Austria
Daniel Holzinger — Hospitaller Brothers of Saint John of God Linz, Austria
David Saldaña — Sevilla University, Spain

Citation

Fellinger, J., Holzinger, D., Saldaña, D., eds. (2022). *Surveillance of language development in pre-school children*. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-83250-914-2

Table of contents

frontiers | Frontiers in Pediatrics

TYPE Editorial
PUBLISHED 17 November 2022
DOI 10.3389/fped.2022.1042094

OPEN ACCESS

EDITED AND REVIEWED BY
Tim S Nawrot,
University of Hasselt, Belgium

*CORRESPONDENCE
Daniel Holzinger
daniel.holzinger@jku.at

SPECIALTY SECTION
This article was submitted to Children and Health, a section of the journal Frontiers in Pediatrics

RECEIVED 12 September 2022
ACCEPTED 03 November 2022
PUBLISHED 17 November 2022

CITATION
Holzinger D, Saldaña D and Fellinger J (2022) Editorial: Surveillance of language development in pre-school children.
Front. Pediatr. 10:1042094.
doi: 10.3389/fped.2022.1042094

Editorial: Surveillance of language development in pre-school children

Daniel Holzinger[1,2,3]*, David Saldaña[4] and Johannes Fellinger[1,2,5]

[1]Research Institute for Developmental Medicine, Johannes Kepler University of Linz, Linz, Austria, [2]Institute of Neurology of Senses and Language, Hospital of St. John of God, Linz, Austria, [3]Institute of Linguistics, Faculty of Humanities, University of Graz, Graz, Austria, [4]Department of Developmental and Educational Psychology, Universidad de Sevilla, Seville, Spain, [5]Division of Social Psychiatry, University Clinic for Psychiatry and Psychotherapy, Medical University of Vienna, Vienna, Austria

KEYWORDS

language surveillance, language screening, preschool, public health, primary pediatric care

Editorial on the Research Topic
Surveillance of language development in Pre-school children

Language disorders are among the most frequent developmental disorders and can have a profound impact on academic and vocational development, mental health, and quality of life from childhood into adolescence and adulthood. Since there is increasing evidence for the effectiveness of intervention, particularly if offered timely and in a family-centered way, early identification of significant language delays is crucial in preventive health care. Due to the high variance of language trajectories in the early years, continuous monitoring of language development rather than single-point screening is indicated.

The collection of articles on this research topic that includes mainly empirical research, as well as a systematic review and meta-analysis by So & To, is an innovative contribution to the field of high relevance for clinical practice focusing on the type of administration (proxy vs. direct), language skills or clinical markers, the consideration of environmental factors (including special populations), the concurrent or predictive character of the screenings, and the feasibility of systematic language surveillance in total populations.

Administration of proxy or direct screening

The findings of studies including parent report screenings support the conclusion of the systematic review by So & To, who show a comparable level of accuracy of parent reports as compared to screenings administered by trained examiners. For children at the age of two years (Holzinger et al.) and three to four years (Doove et al., Holzinger et al., Dockrell et al.) parental screenings achieved a high accuracy. Holzinger et al. demonstrated, for two-year screening, that the parent report as stage 1 (followed by a stage 2 direct evaluation by the pediatrician limited to those failing at stage 1) resulted in good predictive validity of language delay, even one year after screening administration. This finding points to the effectiveness of screening instruments based on a combination of direct child assessment and proxy reports, in line with the recently reported results for a Dutch well child language screening protocol (1).

Screenings based on language ability

Various studies suggest greater precision of instruments based on the child's language ability, compared to those based on clinical markers such as non-word repetition or sentence repetition. This could be due to the high variability found in non-word and sentence repetition in children with language disorder (2). The studies included in the current research topic are mainly language-based and result in good or even excellent accuracy (Holzinger et al., Holzinger et al., Holzinger et al., Holzinger et al., Dockrell et al., Doove et al.). In addition to the assessment of child language skills, parental concerns about language development are found to be predictive of language development trajectories as shown by Holzinger et al. and Doove et al. for language development from age 2–3 and 3–4 years, respectively. Lüke et al's contribution shows for a sample of bilingual infants that the absence of the prelinguistic skill of index finger pointing at the age of 12 months, which shows intentional communication and the ability to initiate joint attention, seems to be an early indicator of language delay at the age of 2 years, in line with findings for monolingual populations (3). In conclusion, the evaluation of language abilities and proximal precursors of linguistic skills, direct or indirect, and parental concerns about their child's language development should be considered key components of effective language screening instruments.

Environmental factors

Language is the product of a complex interplay of biological and environmental factors over time. Factors related to a child's home environment can either buffer or increase biological risk and help to understand children's developmental pathways and the early identification of risk. Eadie et al. demonstrate the potential use of early cumulative risk factors related to the home learning environment (e.g., number of books in the home, frequency of reading, and maternal education, maternal language, and mental health), including parent-child interaction (in addition to characteristics of the child) in the prediction of low language outcomes at 7 years. Many of these factors could probably be included in developmental surveillance programs, although the feasibility of such a recommendation remains to be demonstrated.

Concurrent or predictive screenings

The highly dynamic nature of language development trajectories usually results in a non-satisfying rate of children with later language difficulties missed by a screening at an earlier point of time and/or in screening-fails who turn out to achieve an average language level later-on without having undergone any specific intervention (false positives). As expected, screening tools with longer screening diagnostic intervals demonstrate lower sensitivity than those using short intervals, as shown by the systematic review included in this research topic (So & To). Current findings demonstrate that, because a significant number of children with negative screening results at an earlier point of time develop language difficulties later continuous monitoring of language development (re-screening) and the capturing of environmental effects on language development are required. To avoid early over-identification associated with unnecessary irritation of parents, cost of follow-up investigation and/or interventions with high positive predictive values of the screenings are highly relevant for a population-based implementation of a screening tool. Holzinger et al. demonstrated good predictive validity of a two-stage screener that included a parent report of expressive vocabulary and two-word combinations, parent concerns about language development, and pediatric assessment of word comprehension. In summary, including language comprehension and parental concerns about language development increases the predictive quality of language screenings.

Feasibility

As pointed out in the literature (4), evidence of feasibility of screening measures in regular preventive medical care settings is insufficient. However, the proof of feasibility is essential for the introduction of universal language screening. The studies of our working group that resulted in accurate screening measures for use in pediatric primary care (at the age of 2 and 3 years; Holzinger et al. and Holzinger et al.) and pre-school settings (age of about 4 ½ years; Holzinger et al. and Holzinger et al.) were all implemented with large populations of non-selected children and within the regular system of preventive health care. Acceptability by screeners, parents, and children was rated as high in accordance with high rates of completed screening procedures. It should be noted that even the integration of screening within the time constraints of regular pediatric care was mainly estimated to be well possible. The combination of parent reports and—possibly as a second stage—assessments by trained screeners can contribute to an efficient administration in regular preventive health care. However, in their study on language screenings in disadvantaged populations including a significant number of non-English speaking parents, Dockrell et al. reported low completion rates of parent questionnaires.

Special populations

Dockrell et al. confirmed higher rates of language difficulties in socially disadvantaged populations and showed that a shortened version of a parent questionnaire on their child's language performance was an effective measure that captured the language learning needs of children before they enter nursery schools. The low rate of returned screening forms (38.6%) points to the remaining challenges involved in the use of parent reports with this population.

For children in their penultimate year of kindergarten who grow up multilingually with a minority language as their dominant language, Holzinger et al. demonstrated that a screening targeting expressive grammatical skills in the majority language with bilingual norms achieves high accuracy in the identification of children with language disorders.

Screening for increased risk for deficits in language-related skills

Schöfl et al. present a promising app-based screening tool for universal use at school entry to predict word-reading difficulties through a combination of phonological information processing and linguistic skills. In two small groups of Arabic speaking children and adults who stutter, significant correlations between non word repetition skills and the percentage of stuttered syllables indicate that nonword repetition tasks might be useful for the early identification of stuttering (Alsulaiman et al.).

Conclusions and future directions

This research topic demonstrates the availability of accurate and feasible tools for use in developmental surveillance programs to identify children at increased risk of language difficulties in the first years of life. As a synthesis, the findings indicate the high relevance of parent reports, staged combinations with screenings by practitioners, the validity of screenings based on language and language-related skills, the necessity of including home environment variables and factors such as language comprehension and parental concerns that increase predictive validity and—for some of the studies – acceptance by those involved and practicability in public health approaches. The current state of the development of screening procedures warrants their implementation and evaluation in surveillance programs in total population samples.

Author contributions

DH prepared a first version of the manuscript. JF and DS reviewed the manuscript. All authors contributed to the article 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.

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

1. Visser-Bochane M, Luinge M, Dieleman L, van der Schans C, Reijneveld S. The Dutch well child language screening protocol for 2-year-old children was valid for detecting current and later language problems. *Acta Paediatr*. (2021) 110(2):556–62. doi: 10.1111/apa.15447

2. Ahufinger N, Berglund-Barraza A, Cruz-Santos A, Ferinu L, Andreu L, Sanz-Torrent M, et al. Consistency of a nonword repetition task to discriminate children with and without developmental language disorder in Catalan–Spanish and European Portuguese speaking children. *Children*. (2021) 8(2):85. doi: 10.3390/children8020085

3. Colonnesi C, Stams GJ, Koster I, Noom MJ. The relation between pointing and language development: a meta-analysis. *Dev Rev*. (2010) 30:352–66. doi: 10.1016/j.dr.2010.10.001

4. Wallace IF, Berkman ND, Watson LR, Coyne-Beasley T, Wood CT, Cullen K, et al. Screening for speech and language delay in children 5 years old and younger: a systematic review. *Pediatrics*. (2015) 136(2):e448–62. doi: 10.1542/peds.2014-3889

frontiers in Public Health

ORIGINAL RESEARCH
published: 22 January 2021
doi: 10.3389/fpubh.2020.546536

Check for updates

Preschool Communication: Early Identification of Concerns About Preschool Language Development and Social Participation

Bernice M. Doove[1,2]*, Frans J. M. Feron[2], Jim van Os[3,4,5] and Marjan Drukker[4]

[1] Youth Health Care Division, Regional Public Health Service South Limburg, Heerlen, Netherlands, [2] Department of Social Medicine, Care and Public Health Research Institute, School for Public Health and Primary Care, Maastricht University, Maastricht, Netherlands, [3] King's Health Partners, Department of Psychosis Studies, Institute of Psychiatry, King's College London, London, United Kingdom, [4] Department of Psychiatry and Psychology, MHeNS School for Mental Health and NeuroScience, Maastricht University Medical Centre, Maastricht, Netherlands, [5] Department Psychiatry, Brain Center Rudolf Magnus, Utrecht University Medical Centre, Utrecht, Netherlands

OPEN ACCESS

Edited by:
Daniel Holzinger,
Hospitaller Brothers of Saint John of God Linz, Austria

Reviewed by:
Laura Nabors,
University of Cincinnati, United States
David Saldaña,
Sevilla University, Spain

***Correspondence:**
Bernice M. Doove
bernice.doove@ggdzl.nl;
bernice.doove@maastrichtuniversity.nl
orcid.org/0000-0002-3189-4276

Specialty section:
This article was submitted to Children and Health, a section of the journal Frontiers in Public Health

Received: 28 March 2020
Accepted: 31 December 2020
Published: 22 January 2021

Citation:
Doove BM, Feron FJM, van Os J and Drukker M (2021) Preschool Communication: Early Identification of Concerns About Preschool Language Development and Social Participation. Front. Public Health 8:546536. doi: 10.3389/fpubh.2020.546536

Background: Adverse communication development in preschool children is a risk factor influencing child health and well-being with a negative impact on social participation. Language and social skills develop and maintain human adaptability over the life course. However, the accuracy of detecting language problems in asymptomatic children in primary care needs to be improved. Therefore, it is important to identify concerns about language development as a risk factor for child health. The association between parental and professional caregivers' concerns about language development and the level of preschool social participation was assessed, as well as the possible mediating/moderating effect of the perception of social competence. In addition, validity and predictive value of parental and professional caregivers' concerns about language development were tested.

Methods: To identify emerging concerns about development and social participation, a community sample of 341 preschool children was systematically assessed with a comprehensive preventive child health care "toolkit" of instruments, including parent-completed tools like the Parents' Evaluation of Developmental Status (PEDS) and child competence Visual Analog Scales (VAS). At baseline, children were aged 3 years and at follow-up ∼4 years.

Results: There was a statistically significant association between parental and professional caregivers' concerns about language development and the level of preschool social participation, with a mediating effect of child social competence at the age of 3 years as well as 4 years. Negative predictive value of parental and professional caregiver language concerns at the age of 3 and 4 years were 99 and 97%, respectively. Furthermore, this article showed that while some preschool children grow out of language problems, others may develop them.

Conclusion: Short but valid pediatric primary care tools like the PEDS and child competence VAS can support monitoring and early identification of concerns about language development and social competence as a risk factor for preschool social participation. Personalized health care requires continued communication between parents, professional caregivers and preventive child health care about parental and professional caregiver perceptions concerning preschool language development as well as the perception of a child's social competence.

Keywords: preschool social participation, social competence, communication, early identification, language concerns, PEDS, personalized health care, monitoring

INTRODUCTION

Poor communication is a risk factor influencing child health and well-being with adverse consequences for behavior, literacy, learning, mental health, future employment, parenting, the next generation, and social inequalities (1, 2).

Effective communication is fundamental to the initiation and maintenance of successful peer relations (3, 4). The ability to interact with others and to establish relationships is of great influence on learning and development, and successful social adaptation and participation. From a dynamic perspective, health can be seen as the ability to adapt and self-manage in the face of social, physical, and emotional challenges (5). For this, language and social skills are needed; they develop and maintain human adaptability over the life course (6, 7).

From a public health perspective, preschool children represent an important group (8, 9). The preschool period is a sensitive period in language development (7, 10). Developmental growth in language skills is an important parameter of overall communication development (11). Language problems are often the first presenting symptoms of delay in the development of multiple basic functions including socialization and communication (3, 12). Early expressive and receptive language problems and behavioral problems may have long-term consequences (13). In particular, early receptive language problems are a significant risk factor for adult mental health (1).

However, the accuracy of detecting language problems in asymptomatic children in primary care is inadequate (14). Early recognition of adverse language development is challenging, given that normal development in young children is highly variable and all growth and development takes place in interaction with the environment (15–17). Differentiating between speech language delays and disorders is complicated, children with concerns about language development are a heterogeneous group with different individual and environmental characteristics. On the other hand, many children whose language development is delayed may catch up over the next few years and do not require interventions (18). Prevalence of language problems varies widely (2–25%) due to a lack of consistent definitions, the nature of the population, the diagnostic method that is utilized, and whether data were collected in a clinical sample or in the general population (19–21).

From a personalized health care "growing into deficit" model (**Figure 1**), prevention of language developmental problems requires a focus on concerns, emerging problems and symptoms at an early stage where signs and symptoms do not yet meet diagnostic criteria for a disorder (13, 22, 23). For early identification of language needs, it is important to understand the pervasive nature of language development (8, 13, 19). It is assumed that differences in young children's language development reflect differences in experience and in creating interactive routines, next to their biologically mediated genetic potential (24). If needed, early intervention has to be personalized; standard intervention programs have limited added value (25).

Previous research has shown that there is an association between language difficulties, behavioral difficulties, and social participation (26, 27). Language impairment in childhood may be related to problems with activities and social participation as defined by the International Classification of Functioning Disability and Health—Children and Youth (ICF-CY) (19, 28). Social participation is a broad concept including the objective state and the subjective experience of involvement in society. This concept has to be understood in the light of social roles (6). For young children, play is an important social activity.

Social competence development is linked with both language development and social participation. Social competence is affected when abilities or skills that are required to engage in socio-cognitive processes and to display social behaviors are limited (7). For example, not only expressive and receptive language, but also the ability to grasp non-linguistic signals is important for optimal social interaction and participation. Toddler's play has been associated with their language proficiency (4). However, this same study also showed that the child's functioning in play was better explained by their social competence than by their language skills. From a dynamic transactional developmental perspective (29), it is hypothesized that language development is mediated by social competence and social participation, and vice versa. Language development, social competence, and social participation are seen as dynamic skills simultaneously developing during the preschool period, suggesting a reciprocal model (30).

A community-based approach with a focus on personalized health care requires cooperation and communication within a public health framework (8, 22). According to a bio ecological model of development-in-context, it is important

Abbreviations: PCHC, Preventive Child Health Care; PEDS, Parents' Evaluation of Developmental Status; SDQ, Strengths and Difficulties Questionnaire; Stata, Statistical Software Package; VAS, visual analog scale.

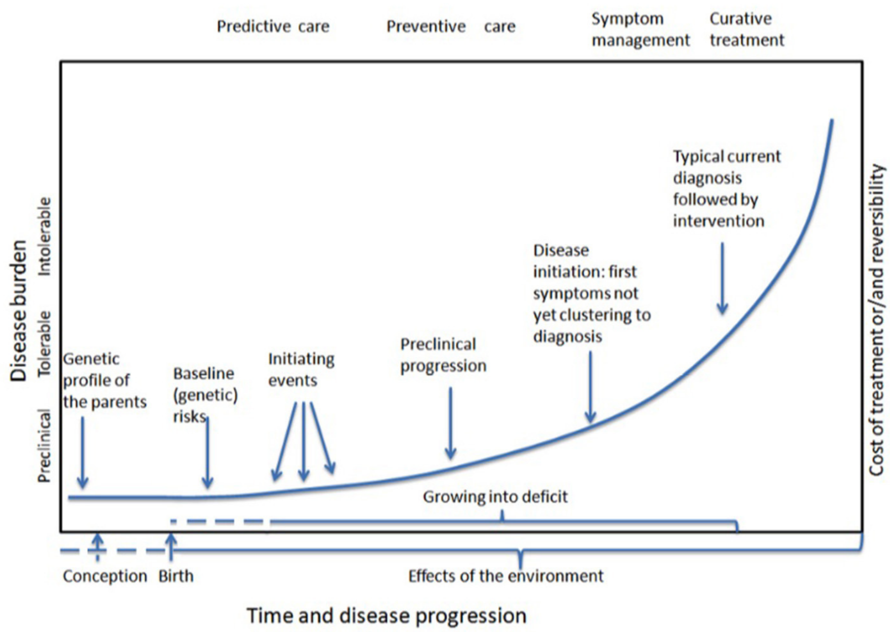


FIGURE 1 | Modification of Syurina's adaptation of Snyderman's curve representing the timelines of "growing into deficit" and developing common complex diseases.

to obtain child context-specific information (20). Teachers, employees from childcare, kindergarten, preschool or primary school (hereafter: professional caregivers) as well as parents and Preventive Child Health Care (PCHC) professionals are important perceivers with expert knowledge on child development from different perspectives.

PCHC is synonymous with Pediatric Preventive Primary Care. All children in the area are regularly invited to visit the PCHC. The Dutch PCHC system includes preventive health care doctors and has a high level of population compliance. It is a public health endeavor to provide ongoing monitoring up to the age of 18 years (31, 32). This way, the early conditions that place children at risk for less than optimal development and successful social participation can be improved (33–36). To deal with emerging problems and symptoms at an early stage where signs and symptoms do not yet meet diagnostic criteria for a disorder, systematically exploring parental as well as other caregivers' concerns is a main component in PCHC for family-centered practice and personalized health care. Knowledge and understanding of the true epidemiology of genetic and environmental risk and protective factors and their early phenotypes can help in prevention of "growing into deficit" (23, 37).

In order to document children's development over time, monitoring development at multiple time points, across informants, instruments and contexts, is more valid and accurate than a single assessment (16, 38–41). For early identification of developmental problems, special attention should be given to the validity of instruments about the perceived impact of concerns as concurrent and long-term predictors, and outcome domains such as health, well-being and social participation (42).

In a PCHC setting, monitoring instruments should: (1) easily obtain information in every day PCHC setting; (2) carry out dimensional assessment of symptoms and behavior; (3) measure the progress of development of young children and their possible determinants of influence; (4) identify general signals and symptoms indicating a possible disruption or imbalance of the educational/parent-child system, not yet related to a specific diagnosis; (5) support communication between PCHC, parents and professional caregivers about their perceptions on health and development; (6) connect to needs and demands of the child and the social system around the child; and (7) promote shared decision making (43, 44). Short instruments with a high negative predictive value are preferred; it ensures that most children who pass the developmental assessment are truly healthy. Follow-up consultations are no problem, these children can benefit from additional preventive monitoring (45).

Research has shown that parent-completed tools are highly accurate in detecting true problems, are relatively inexpensive, and promote a dialogue about concerns, needs and demands between parents and other caregivers (46–48). Therefore, incorporating tools utilizing a parent—and professional caregivers—report assessment like the Parents' Evaluation of Developmental Status (PEDS), child competence Visual Analog Scales (VAS) and the Strengths and Difficulties Questionnaire (SDQ) into a routine child monitoring toolkit could improve the rate of early identification of concerns about language development, social competence and social participation (43, 49). In this article, the concept social participation was operationalized using instruments to assess early emerging concerns about factors underlying preschool competence and social participation: a child's general competence at day care, kindergarten and preschool, the impact of distress and the total amount of concerns about child development and behavior.

This article investigates (1) the validity of the Parents' Evaluation of Developmental Status (PEDS) to assess language development concerns; (2) the cross-sectional association of language development concerns with social participation; (3) the longitudinal association of language development concerns with social participation, and (4) the possible mediating effect of social competence on the association between language development and social participation at the ages of 3 and 4 years.

METHODS

The present study was performed as part of the Monitoring Outcome Measurements of child development (MOM) study, a prospective observational study within PCHC practice. A community-based sample of 346 children was systematically assessed with a comprehensive PCHC "toolkit" of instruments using a multisource and cross-informant repeated measures design to identify developmental pathways impacting school readiness as an outcome of social participation. Children were aged 3 years at baseline and 4 years at follow up.

The Maastricht University Medical Center Medical Ethics Committee approved the MOM-study protocol under registration number MEC 09-04-018/P. Therefore, this study has been performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki and its later amendments. All participating parents gave their informed consent prior to their inclusion in the study.

Data Collection and Instruments

For this article, data of parents, professional caregivers and PCHC professionals of 341 children were analyzed. At baseline, children were aged 3 years and at follow-up ∼4 years. To assess emerging problems, signs and symptoms, perceptions, demands and concerns about development and social participation, various short instruments like the Parents' Evaluation of Developmental Status (PEDS), the Strengths and Difficulties Questionnaire (SDQ) and child competence Visual Analog Scales (VAS) were included in this study. PCHC professionals provided information about, for example, background factors, family history, child health, and development and interventions.

Language Concerns

Parents as well as professional caregivers completed the PEDS, a 10-item standardized semi-structured questionnaire to elicit concerns regarding child development for children aged <8 years in the general population and clinical samples (40). Ten questions explore concerns in various domains: expressive and receptive language, fine motor, gross motor, behavior, socialization, self-care and learning. The PEDS-question could be answered on a trichotomous scale: "no," "a little," "yes." Subsequently, an open-ended field provides more information. The PEDS is validated for clinical samples and general population samples aged between 0 and 8 years, and is available in multiple languages. In recent validation studies from the USA for the accuracy of parental concerns in detecting children at high and/or moderate developmental risk, the PEDS has a sensitivity of 91–97% and specificity of 73–86% (50). The PEDS is less time-consuming than other instruments, emphasis is on parental and other professional caregivers' opinions, and has reasonable test characteristics for developmental screening in primary care settings (51). Furthermore, the PEDS has shown to be reliable, valid and useful as brief monitoring tools in daily Dutch PCHC practice (43, 51). This suggests that the PEDS is an accurate tool for use as an initial screening and monitoring tool in Dutch PCHC, where professionals have to deal with the time constraints of daily practice.

For the current paper, dichotomous "parental concerns" and "professional caregiver concerns" variables about expressive and/or receptive language (any concern yes/no) were constructed for use in the analyses, if any of the parents or professional caregivers scored "yes" or "a little," the answer was recoded as "yes."

Child Competence

To address the issue of the child's functional adaptation, professional caregivers were asked to indicate on 2 VAS, the degree of the child's general competence and the child's social competence (0 = not competent, 100 = very competent).

Participation

The child's general competence as described above is one of the instruments to assess the broad construct of participation. Other instruments are SDQ total score and SDQ impact.

The Dutch version of the SDQ was completed by parents as well as by professional caregivers to assess the child's behavior (46, 52–54). The SDQ is a brief behavioral screening questionnaire for children aged 3–16 years. It also includes items that identify the impact of the behavioral problems of the child. The SDQ is considered valid and reliable as a research instrument in community samples (49). For this article, the "SDQ total sum score" and the "SDQ impact of distress score" of both parents and professional caregivers were used. If any of the parents or professional caregivers scored "yes" on the impact probe question, the dichotomous overall distress variable was set at "yes."

Van Wiechen Developmental Test

In addition to the validation of the overall PEDS, validity of the PEDS language items was assessed, using the Van Wiechen developmental test as reference standard (43, 55–57). This Dutch instrument is a modification of the Gesell test and is routinely used by all PCHC Centers in the Netherlands and Belgium to monitor the development of all children from birth to the age of 4 years. It consists of a set of 57 developmental indicators to assess motor behavior, speech, communication, and social skills based on physicians' observations and interviewing the parents. A total of 23 indicators cover language development and communication and are called language milestones. All PCHC professionals are trained to asses and register mile stones in the PCHC system according to a uniform protocol. For this paper, the Van Wiechen communication and language items were used. In a large community-based sample of Dutch children, test characteristics of the Van Wiechen language items for the age group 36–48 months showed an Area Under the Curve (AUC) of 0.83%, with an average sensitivity of 66.1%, specificity of 87.5%, positive predictive value (PPV) of 29.2%, and a negative predictive value (NPV) of 98.8% (58).

In a study in Australia, agreement between ratings of parental PEDS language concerns and clinical assessment was high (86–90%); agreement between teacher PEDS language concerns and clinical assessment was lower and more varied (63–80%) (59). In this study, parental and professional caregiver PEDS language concerns were combined to provide complementary information and capture all possible language concerns of a specific child. Subsequently this combined concerns variable was validated; reference standard was the Van Wiechen developmental test, communication and language items (see above) (60, 61). For this study, the PCHC professionals were asked to judge the Van Wiechen language and communication items as "sufficient" or "not sufficient," at the age of 3 years and a year later.

Other Variables

As an indicator of socioeconomic status, the level of maternal and paternal education was assessed across three categories: low (primary education, junior vocational education), middle (general secondary education, senior vocational education) and high (preparatory university education and university education). The parent with the highest level of education determined parental educational level.

Statistical Analyses

All analyses were performed using Stata Statistical Software, version 15 (62). First, to assess the validity of the PEDS language items, positive predictive value (PPV) and negative predictive value (NPV) were obtained at the age of 3 years and at the age of 4 years. PPV and NPV were then assessed as measures of predictive validity of language concerns at the age of 3 years, using the Van Wiechen developmental test at age 4 years as reference standard. Second, logistic and linear regression analyses were performed. In the cross-sectional analyses at both ages 3 and 4 years and in the longitudinal analyses, the independent variable (X) was concerns about language development. The dichotomous dependent variable (Y) to index social participation, used in the logistic regression analyses, was SDQ impact of distress score. Continuous dependent variables (Y) to index social participation, used in the linear regression analyses, were: SDQ total score, the child's general competence VAS and the child's social competence VAS. Analyses were adjusted a priori for age, sex, and parental educational status.

Finally, to analyze the fourth research question, the child's social competence was included as a mediator (M) in the association between independent (X) and dependent (Y) variable. Mediation was assessed by analyzing a regression model with and without the mediator. The question was whether the association between X and Y after including the mediator is zero or substantially smaller than the direct association between X and Y. This is visualized in **Figure 2**. The arrows a, b, c and c' present regression coefficients or odds ratios: a represents the association between independent variable (X) and mediator (M); b represents the association between the mediator (M) and the dependent variable (Y); c represents the crude association between independent variable (X) and dependent variable (Y); and c' represents the association between X and Y after including the mediator (M) in the regression model. When the hypothesis that there can be mediation is plausible and c shows an association while c' is smaller or close to zero, there is evidence for partial or full mediation, respectively.

RESULTS

Parents of 346 children agreed to participate in the MOM study. At baseline, parents of 341 children and professional caregivers of 301 children completed the questionnaires. The mean age of the children was 3.0 years (SD 0.2, **Table 1**). For 296 of these children (86%), information from both informants was available. In the follow up, at the age of ∼4 years (mean age 3.8; SD 0.2, **Table 1**), information of both informants was available for 236 children (68%). For 32 children (9%) there was no information available from parents or professional caregivers because they did not return the questionnaire. At baseline, the total sample of children consisted of 166 boys (48%) and 180 girls (52%). Of the participating children, 60% ($n = 207$) were resident in the municipality of Maastricht, while 40% ($n = 139$) lived in the surrounding areas. At baseline, parents and/or professional caregivers of 108 (32%) of 334 children had concerns about expressive and/or receptive language development (12 missings on the PEDS). In the follow up, at the age of 4 years, the total number of children with concerns about language development was 81 (26%) of 313 children (**Table 1**).

In order to test representativeness, 40% of non-responders were randomly sampled to manually collect data on parental education from the medical files. The distribution in non-responders was minimally different from distribution in responders (responders 63, 27, and 10% and non-responders 55, 33, and 12% having high, intermediate and low parental education, respectively).

Validity of the PEDS Language Items

The prevalence of PCHC language concerns was 8% at the age of 3 and 7% one year later (**Table 2**). At the age of 3 years, PEDS language concerns had a PPV of 23% and NPV of 99%. At the age of 4, the PPV and NPV of PEDS language concerns were 19 and 97% respectively (**Table 2**). **Table 3** shows the stability of language developmental concerns at the age of 3 and 4 years. The predictive validity of the PEDS at the age of 3 years was: PPV of 14% and NPV of 97% (**Table 4**).

Association Between Language Development Concerns and Preschool Social Participation

When assessing parental SDQ impact at the age of 3 years, children with receptive language concerns had an OR of 7.3 and children with expressive language concerns had an OR of 2.4. However, confidence intervals were overlapping (**Table 5**).

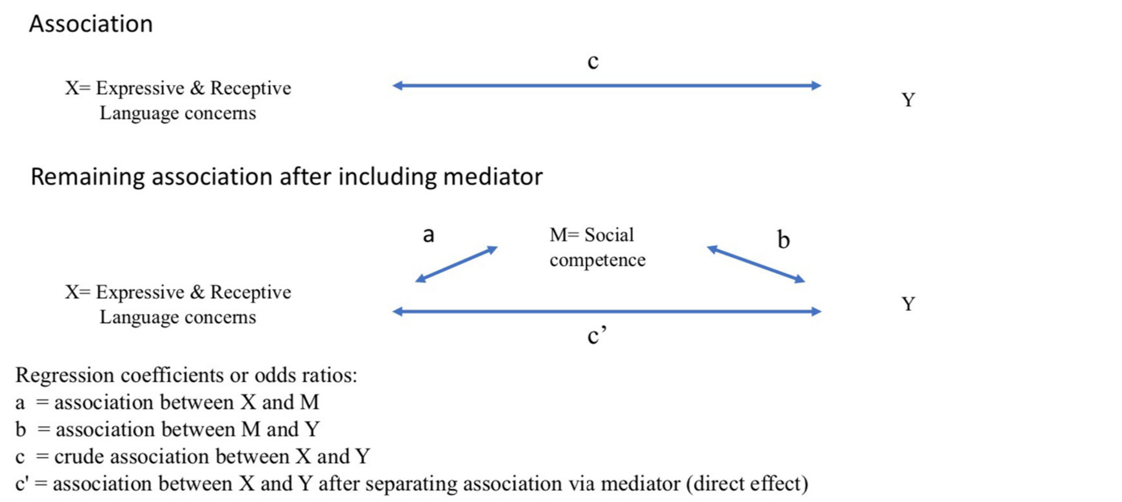


FIGURE 2 | Theoretical figure to explain analysis of mediation.

TABLE 1 | Descriptive statistics at baseline (T1) and follow-up (T2).

Variable	*N*		Mean (S.D.)		Range	
	T1	T2	T1	T2	T1	T2
Age in years	346	293	3.0 (0.2)	3.8 (0.2)	1.8–3.5	3.5–4.8
Child general competence VAS[a]	290	251	63.7 (19.7)	69.7 (16.3)	4–100	10–99
Child social competence VAS[a]	297	254	62.9 (23.4)	68.9 (20.1)	3–100	9–99
SDQ (parents)	338	293	6.8 (4.9)	6.1 (4.2)	0–28	0–27
SDQ (prof.[b])	294	256	6.1 (5.0)	5.0 (5.0)	0–27	0–29
			Normal		Atypical	
PEDS concerns about language[c]	334	313	226 (68%)	232 (74%)	108 (32%)	81 (26%)
Van Wiechen developmental test [d]	331	319	304 (92%)	298 (93%)	27 (8%)	21 (7%)
SDQ impact (parents)	340	292	307 (90%)	271 (93%)	33 (10%)	21 (7%)
SDQ impact (prof.[b])	292	254	248 (85%)	223 (88%)	44 (15%)	31 (12%)

[a]*A higher Visual Analog Scale (VAS) score means professional caregiver judges child competence more positive.*
[b]*Professional caregivers.*
[c]*Parental and/or professional caregiver's concerns about expressive and/or receptive language development.*
[d]*Speech, language and communication items.*

According to professional caregivers, the association between receptive language concerns and outcomes was stronger than the association between expressive language and outcomes (e.g., general competence VAS: B $= -21.3$, $p < 0.001$; **Table 6**). In addition, both professional caregivers and parents reported more behavioral problems when there were receptive language concerns (SDQ total score B $= 4.5$, $p < 0.001$; B $= 4.5$, $p < 0.001$, respectively). A year later, the association between language concerns, competence and behavior was less strong but still significant, except for PEDS expressive language concerns and the parental perception of child behavior (**Table 6**).

Mediating Effect of Social Competence

At baseline and at follow up a year later, there was a significant association between social competence and social participation, but also a direct association between concerns about expressive and/or receptive language development and social competence. For example, according to the professional caregiver, at age 3 years there was a significant association between language concerns and social participation (SDQ impact B $= 4.3$, $p < 0.001$; SDQ total score B $= 3.0$, $p < 0.001$) and a significant association between social competence and social participation (SDQ impact B $= 0.9$, $p < 0.001$; SDQ total score B $= -0.1$, $p < 0.001$) (**Tables 7**, **8**). There was a mediating effect of social competence: after inclusion of social competence in the regression model, the remaining association between language concerns and social participation was less strong (SDQ impact B $= 2.8$, $p = 0.014$; SDQ total score B $= 1.2$, $p = 0.016$). A year later, at the age of 4 years, the mediating effect of social competence was even stronger with non-significant regression coefficients (B =

TABLE 2 | Prevalence, Positive, and Negative Predictive Value of PEDS concerns about language development at the age of 3 and 4 years.

Expressive language and/or receptive language	T1 reference standard[b]		
T1PEDS concerns[a]	**Yes**	**No**	**Total**
Yes	24	82	106
No	3	222	225
Total	27	304	331
Expressive language and/or receptive language	**T2 reference standard[b]**		
T2 PEDS concerns[a]	**Yes**	**No**	**Total**
Yes	15	65	80
No	6	225	231
Total	21	290	311

T1 Prevalence 8.2%; Specificity 73%; Sensitivity 88.9%; PPV 22.6%; NPV 98.7%.
T2 Prevalence 6.8%; Specificity 77.6%; Sensitivity 71.4%; PPV 18.8%; NPV 97.4%.
[a]*Parental and/or professional caregiver's concerns about language development.*
[b]*Van Wiechen developmental test, speech, language and communication items.*

TABLE 3 | Stability of PEDS concerns about language development at the age of 3 and 4 years.

Expressive and/or receptive language	T2 PEDS concerns[a]		
T1 PEDS concerns[a]	**Yes**	**No**	**Total**
Yes	55 (18%)	45 (14%)	100 (32%)
No	26 (8%)	186 (60%)	212 (68%)
Total	81 (26%)	231 (74%)	312(100%)

Prevalence 26.0%; Specificity 80.5%; Sensitivity 68.0%; PPV 55.0%; NPV 87.7%.
[a]*Parental and/or professional caregiver's concerns about language development.*

TABLE 4 | The predictive validity of PEDS concerns about language development at the age of 3 years.

Expressive and/or receptive language	T2 reference standard[b]		
T1 PEDS concerns[a]	**Yes**	**No**	**Total**
Yes	14	89	103
No	7	209	216
Total	21	298	319

Prevalence 6.6%; Specificity 70.1%; Sensitivity 66.7%; PPV 13.6%; NPV 96.8%.
[a]*Parental and/or professional caregiver's concerns about language development.*
[b]*Van Wiechen developmental test, speech, language, and communication items.*

1.1, $p = 0.884$ and B = 0.8, $p = 0.198$, respectively). When tested separately, the mediating effect was found both for expressive and receptive language concerns (data not shown). Mediating effects of social competence were also found between language concerns and general competence (data not shown).

TABLE 5 | Logistic regression analysis (significant interaction with one or both risk factors): association between PEDS concerns about expressive and receptive language development and SDQ impact according to parents and professional caregivers at the age of 3 and 4 years; odds ratios (OR) and 95% confidence intervals (CI).

	SDQ impact (Parents)		SDQ impact (prof.[b])	
	T1	**T2**	**T1**	**T2**
	OR (95% CI)	**OR (95% CI)**	**OR (95% CI)**	**OR (95% CI)**
PEDS concerns[a] Expressive language	2.4 (1.2; 5.1)*	2.5 (1.0; 6.3)*	3.0 (1.5; 5.8)**	2.3 (1.1; 5.0)**
PEDS concerns[a] Receptive language	7.3 (3.1; 17.3)†	5.3(1.7; 16.3)**	10.5 (4.6; 24.1)†	2.6 (0.8; 8.6)

**p <0.05; **p <0.01; †p <0.001; 95% CI, 95% confidence interval; OR, odds ratio (obtained from logistic regression).*
[a]*Parental and/or professional caregiver's concerns about language development.*
[b]*Professional caregivers.*

DISCUSSION

The results in this paper suggest concurrent and predictive validity of the PEDS to assess parental and professional caregivers' language development concerns, as well as the mediating effect of professional caregivers' perception of the child's social competence in the association between these concerns and social participation at the age of 3 years as well as the age of 4 years.

Validity of Preschool Language Development Concerns

Prevalence of language delay (7–8%) in the present study is within the international reported range of prevalence of atypical language delay (7–15%) (20, 21, 38). In addition, the association between parental and professional caregiver concerns on the one hand, and not meeting the expected milestones for language on the other, was statistically significant. These results are in line with other studies where parental concerns were consistently associated with preschool language development (40, 58, 63). Moreover, this confirms the value of including parents' and professional caregivers' expert knowledge in the assessment and clinical decision-making process for personalized support (28, 59). The high NPV of parental and professional caregiver language concerns validate a strategy of exclusion of children without concerns from extra monitoring. The PEDS language screening items appear to be very good in identifying children who do not have any language delay. Current assessment tools are still not sufficiently specific to discriminate between delayed language that will resolve naturally and delayed speech or language that will develop into persistent problems. The relatively low PPV in the present study implies a high percentage of false positives. Earlier research has shown that children with false positive screening results differ from children with true negative scores. These children had more risk factors and their performance on diagnostic measures was less (45). As

TABLE 6 | Linear regression analysis: association between PEDS concerns about expressive and receptive language development and child general competence, child social competence, and total score SDQ according to parents and professional caregivers at the age of 3 and 4 years; b- coefficient (B) and 95% confidence intervals (CI).

	Child general competence VAS[b] (prof.[c])		Child social competence VAS[b] (prof.[c])		Total score SDQ (Parents)		Total score SDQ (prof.[c])	
	T1	T2	T1	T2	T1	T2	T1	T2
	B (95% CI)	B (95% CI)	B (95% CI)	B (95% CI)	B (95% CI)	B (95% CI)	B (95% CI)	B (95% CI)
PEDS concerns[a] Expressive language	−11.4 (−16.4 −6.8)†	−10.7 (−15.1; −6.2)†	−12.1(−17.9; −6.4)†	−10.3 (−15.9; −4.7)†	2.6 (1.5; 3.7)†	1.1 (−0.0; 2.3)	2.8 (1.5; 4.0)†	2.2 (0.8; 3.6)**
PEDS concerns[a] Receptive language	−21.3 (−28.7; −13.9)†	−13.3 (−21.7; −4.9)**	−19.1 (−27.8; −10.4)†	−16.1 (−26.5; −5.7)**	4.5 (2.7; 6.3)†	4.2 (2.3; 6.0)†	4.5 (2.7; 6.3)†	3.2 (0.8; 5.7)**

**p* < 0.05; ***p* < 0.01; †*p* < 0.001; *95% CI, 95% confidence interval; b–coefficient (B) (obtained from linear regression).*
[a]*Parental and/or professional caregiver's concerns about language development.*
[b]*A higher VAS score means professional caregiver judges child competence more positive.*
[c]*Professional caregivers.*

TABLE 7 | Mediating effect of social competence on the association of PEDS concerns about language development with SDQ impact score at T1 and T2.

		Path	Odd ratios			
			T1		T2	
			Parent	Prof.[b]	Parent	Prof.[b]
PEDS concerns[a] language (X)	Social competence (M)	A	−13,5†[c]	−13,5†[c]	−10.8†[c]	−10.8†[c]
Social competence (M)	SDQ Impact score (Y)	B	1.0**	0.9†	1.0†	0.9†
PEDS concerns[a] language (X)	SDQ Impact score (Y)	C	4.2†	4.3†	2.8*	2.4*
PEDS concerns[a] language (X)	SDQ Impact score (Y)	C'	3.0**	2.8*	0.9	1.1

**p* < 0.05; ***p* < 0.01; †*p* < 0.001.
[a]*Parental and/or professional caregiver's concerns about expressive and/or receptive language development.*
[b]*Professional caregivers.*
[c]*B regression coefficients.*
Odd ratios unless otherwise indicated.
A = Association between X and M.
B = Association between M and Y.
C = Crude association between X and Y.
C' = Association between X and Y after separating association via mediator (direct effect).

TABLE 8 | Mediating effect of social competence on the association of PEDS concerns about language development with SDQ total score at T1 and T2.

		Path	B regression coefficient			
			T1		T2	
			Parent	Prof.[b]	Parent	Prof.[b]
PEDS concerns[a] language (X)	Social competence (M)	A	−13,5†	−13,5†	−10.8†	−10.8†
Social competence (M)	SDQ total score (Y)	B	−0.0†	−0.1†	−0.1†	−0.1†
PEDS concerns[a] language (X)	SDQ total score (Y)	C	3.2†	3.0†	1.5**	2.4 **
PEDS concerns[a] language (X)	SDQ total score (Y)	C'	2.1†	1.2*	0.1	0.8

**p* < 0.05; ***p* < 0.01; †*p* < 0.001.
[a]*Parental and/or professional caregiver's concerns about expressive and/or receptive language development.*
[b]*Professional caregivers.*
B regression coefficients.
A = Association between X and M.
B = Association between M and Y.
C = Crude association between X and Y.
C' = Association between X and Y after separating association via mediator (direct effect).

confirmed by other population studies (17, 18, 26, 64), this article showed that while some preschool children grow out of language problems, others may develop them (**Table 3**). From a classical screening point of view, children crossing back and forth over the threshold would impact sensitivity and specificity. However, PCHC repeated monitoring concerns of language development and if necessary, extra follow up can make a distinction between children "growing into or out of deficit."

So, PCHC monitors the true positives as well as the false positives and refers when needed, even if the child did not score on the reference standard. Language tests may not capture important aspects of everyday communication. In addition, a language problem may not always look like a language problem: underlying comprehension impairment can present as poor academic attainment, impaired social interaction, or behavioral difficulties (12, 65). Furthermore, due to the variation in the cut-off points of different "reference standard" measures in research, interpretation of parents' and professional caregivers' information is complicated. In addition, there is no agreement on different definitions of language disorders and what proportion of the population should be considered cases that need intervention (65).

Prediction of Preschool Social Participation: Mind the Communication

Parental and professional caregiver concerns were associated with altered social participation at home as well as in preschool. This association was seen both in cross-sectional and in longitudinal analyses.

Language concerns seem to be predictive for altered social participation as early as in preschool. Earlier research showed that especially children who experienced language impairment that persisted into the school years are at risk for adult mental health problems and substandard social participation (66). The strongest association was seen between receptive language delay concerns and behavioral problems. Odds ratio confidence intervals of children with receptive and expressive language concerns were overlapping, thus were not statistically significant, except for the association with parental SDQ total score at the age of 4 years (**Table 5**). From PCHC practice it is recognizable that receptive and expressive language development are closely linked, with more problems in social participation because of language comprehension problems. This confirms the observation that needs of children with receptive language problems are complex and call for extra monitoring of the child's developmental pathway (1). Listening to parental and professional caregivers' concerns with avoidance of diagnostic labels is an important aspect of PCHC clinical judgement and pre-screening. It may identify other developmental problems without potential stigmatization (23, 37). Avoidance of diagnostic labels is not the same as denying any role of biological risk factors in causing health problems; children vary in their biological as well in their social backgrounds and life events (65).

Mediating Effect of Social Competence

Language delay in itself may not be a risk factor for later behavioral and emotional disturbances (67). The present results showed that concerns about language development may reflect the effect of other developmental problems (68). There was a mediating effect of child social competence on the association between receptive and expressive language concerns and social participation at the age of 3 and 4. While at age 3 years social competence was a partial mediator, at age 4 it was a full mediator. So, at age 4, social competence seems to play a more important role in the association between concerns and participation. After inclusion of social competence score, language concerns seem to lose their predictive value but these factors might be related to each other.

The expansion of this mediating effect between ages 3 and 4 years emphasizes once again that all children with language concerns can benefit from additional monitoring to prevent "growing into deficit," especially concerning interpersonal relationships. There is a role for enhanced monitoring in which the primary care professional responds to parental concerns about language development and social skills (3, 6, 23). The group of children with symptoms of mental problems may be twice as large as the group of children meeting formal diagnostic criteria for a mental disorder (69–77). Inefficiency can arise if educational and medical support is restricted to those who meet arbitrary cut-offs as a result of discrepancy in criteria used for diagnostic labels (65). Therefore, a PCHC "toolkit" with short instruments for regular short parental and professional caregivers' reports can serve as a first step in PCHC monitoring procedures to select children who require further support in the form of a "watch and wait" strategy, assessment of other developmental domains, or referral to a specialist. Professional support can then be tailored to the needs, conform the child's development.

PCHC professionals have to deal with emerging problems and symptoms at a stage where signs and symptoms do not yet meet diagnostic criteria, but already give rise to early impairment and distress for both the children and their context, at home as well as in preschool. Both parental and professional caregiver concerns are relevant for early detection of problems, because they both know the child and their perception is from a different perspective (41). The PEDS: (1) facilitates monitoring of parental and professional caregivers concerns; (2) identify general signals and symptoms not yet related to a specific diagnosis; (3) support communication between PCHC, parents and professional caregivers about their perceptions on health and development; and (4) promote shared decision making (23, 44).

METHODOLOGICAL ISSUES

Strength of the study is that a community sample of preschool children was systematically assessed using a comprehensive PCHC "toolkit" of instruments designed for the purpose of monitoring in a public health setting. The study was integrated in real life practice. No children were excluded for not meeting inclusion criteria. In addition, the child's development and participation was evaluated across different settings with cross-sectional and longitudinal information from different instruments and multiple informants. With emphasis on their perception, information was obtained through hetero-anamnesis of parents and professional caregivers (78). PCHC professionals provided data as well; with exception of the Van Wiechen developmental test, these data were not used in the present paper. Furthermore, the Van Wiechen language milestones were collected in a uniform manner by trained professionals.

The present paper has some limitations. First, response rates were difficult to establish. In the MOM region over the study period, 1,692 children were born and, therefore, were within the caseload of the PCHC professionals participating in the MOM study. However, not all PCHC professionals participated in MOM. Consequently, parents of non-participating PCHC doctors were asked to participate by another PCHC doctor (BD), who did not know these families. During the baseline inclusion, the number of participating PCHC professionals increased. Response from one PCHC doctor who participated from the beginning (BD) was 70%. Because response in participating doctors was relatively high and because distribution in socio-economic status was comparable, results presented in this article can be considered approximately representative for the general population. If the PEDS and different VAS are implemented in general PCHC practice, a response rate higher than 70% is expected, because a possible barrier for parents to participate in MOM was the number of questions added for research purposes (e.g., additional instruments for the purpose of validation of VAS). Usually, short form questionnaires collected in PCHC have response rates between 80 and 90% (79).

Second, the MOM data are limited to the city of Maastricht and surrounding areas. This part of the Netherlands is quite similar to the rest of the country. However, there are some differences. In Maastricht, the proportion of non-European inhabitants (about 10%) is less than in the larger cities in the north west of the country (about 30%) (80). In addition, the proportion of highly educated parents participating the MOM is quite large (63%). For this reason, the MOM study findings may not necessarily be valid for large cities with ethnically mixed populations and areas with a larger proportion of low educated parents.

Third, assessment tools were general in nature and did not reveal specific information to assist with remediation of deficits. The measures find general delay and it is not necessarily clear that the delays are clinically significant. PCHC professionals must remain aware they have to deal with emerging problems and symptoms at a stage where signs and symptoms do not yet meet diagnostic criteria. Although each type of parental concern can be associated with validated tests on the same developmental domain, studies about the validity of the PEDS showed that parents often have concerns in seemingly unrelated domains [i.e., parents often reflect on not just the apparent problem but also its impact on other aspects of development (40)]. Revelation of parental and professional caregivers' concerns are a first step in PCHC monitoring procedures to select children who require further assessment of other developmental domains, or referral to a specialist.

Fourth, social competence was included as a mediator and this was based on judgement of the professional caregiver. However, if a similar instrument was available from the parents, we would expect to find similar mediating effects.

Finally, because of the small sample size, analyses were not adjusted for risk factors (e.g., family history of language or literacy problems, health or developmental problems).

CONCLUSIONS

The individual's social and educational environment, including interpersonal relationships, is hypothesized to be instrumental for PCHC professionals wishing to provide personalized preventive public health care for successful participation for all children (31). In order to identify emerging problems at an early stage where signs and symptoms do not yet meet diagnostic criteria for a disorder, short but valid PCHC monitoring tools like the PEDS and different VAS are required (43). Within this PCHC "toolkit," parental and professional caregivers' perception and concerns about language development take an important position. Language development can be seen as the outcome of the mental processes set in motion when the child meets the social and linguistic world (24, 81). The analyses presented here uncovered significant associations between parental and professional caregiver concerns about language development, the child's social competence and the level of preschool social participation. Therefore, pediatric primary care professionals may productively use parental and professional caregiver perceptions concerning preschool language development in clinical practice. Equally important is the perception of a child's social competence. In children not meeting the expected milestones for language development, a comprehensive developmental evaluation and additional monitoring of child development may be required, particularly concerning interpersonal relationships. Consequently, personalized health care requires cooperation within the public health frame. Monitoring of language and social competence development in preschool children can profit from continued communication between parents, professional caregivers and preventive child health care.

DATA AVAILABILITY STATEMENT

The datasets for this article are not publicly available to conform with the General Data Protection Regulation, a European regulation regarding the processing of personal data. Requests to access the datasets should be directed to Bernice Doove, MD, bernice.doove@ggdzl.nl.

ETHICS STATEMENT

The Maastricht University Medical Center Medical Ethics Committee approved the MOM-study under registration number MEC 09-04-018/P. Therefore, this study has been performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki and its later amendments.

AUTHOR'S NOTE

Underlying research materials related to this paper are available on request by the corresponding author.

AUTHOR CONTRIBUTIONS

BD drafted the initial manuscript. FF, JO, and MD reviewed and revised the manuscript. All authors approved the final manuscript as submitted, agreed to be accountable for all aspects of the work, and conceptualized and designed the study.

FUNDING

ZonMw, the Netherlands Organization for Health Research and Development, funded this study (No. 7125.0001).

ACKNOWLEDGMENTS

Parents, professional caregivers, and colleagues are kindly thanked for their participation. Thanks to Frances Page Glascoe for permission to use the PEDS.

REFERENCES

1. Schoon I, Parsons S, Rush R, Law J. Children's language ability and psychosocial development: a 29-years follow-up study. *Pediatrics.* (2010) 126:e73–80. doi: 10.1542/peds.2009-3282
2. Schoon I, Parsons S, Rush R, Law J. Childhood language skills and adult literacy: a 29-years follow-up study. *Pediatrics.* (2010) 125:e459–66. doi: 10.1542/peds.2008-2111
3. Mok PL, Pickles A, Durkin K, Conti-Ramsden G. Longitudinal trajectories of peer relations in children with specific language impairment. *J Child Psychol Psychiatry*. (2014) 55:516–27. doi: 10.1111/jcpp.12190
4. Stangeland EB. The impact of language skills and social competence on play behaviour in toddlers. *Eur Early Childh Educ Res J.* (2017) 25:106–21. doi: 10.1080/1350293X.2016.1266224
5. Huber M, Knottnerus JA, Green L, van der Horst H, Jadad AR, Kromhout D, et al. How should we define health? *BMJ.* (2011) 343:d4163. doi: 10.1136/bmj.d4163
6. Zubrick SRT, Catherine L, Lawrence D, Mitrou F, Christensen D, Dalby R. The development of human capability across the lifecourse: perspectives from childhood. *Australasian Epidemiol.* (2009) 16:6–10.
7. Beauchamp MH, Anderson V. SOCIAL: an integrative framework for the development of social skills. *Psychol Bull.* (2010) 136:39–64. doi: 10.1037/a0017768
8. Law J, Reilly S, Snow PC. Child speech, language and communication need re-examined in a public health context: a new direction for the speech and language therapy profession. *Int J Lang Commun Disord.* (2013) 48:486–96. doi: 10.1111/1460-6984.12027
9. Jones DE, Greenberg M, Crowley M. Early social-emotional functioning and public health: the relationship between kindergarten social competence and future wellness. *Am J Public Health.* (2015) 105:2283–90. doi: 10.2105/AJPH.2015.302630
10. Cunningham BJ, Rosenbaum PL. A Bioecological framework to evaluate communicative participation outcomes for preschoolers receiving speech-language therapy interventions in Ontario, Canada. *Int J Lang Commun Disord.* (2015) 50:405–15. doi: 10.1111/1460-6984.12145
11. Goorhuis-Brouwer SM, Knijff WA. Efficacy of speech therapy in children with language disorders: specific language impairment compared with language impairment in comorbidity with cognitive delay. *Int J Pediatr Otorhinolaryngol.* (2002) 63:129–36. doi: 10.1016/S0165-5876(02)00004-6
12. O'Hare, Bremner L. Management of developmental speech and language disorders: part 1. *Arch Dis Child.* (2016) 101:272–7. doi: 10.1136/archdischild-2014-307394
13. Hawa VV, Spanoudis G. Toddlers with delayed expressive language: an overview of the characteristics, risk factors and language outcomes. *Res Dev Disabil.* (2014) 35:400–7. doi: 10.1016/j.ridd.2013.10.027
14. Siu AL, Force USPST. Screening for speech and language delay and disorders in children aged 5 years or younger: US preventive services task force recommendation statement. *Pediatrics.* (2015) 136:e474–81. doi: 10.1542/peds.2015-1711
15. Masten AS. Global perspectives on resilience in children and youth. *Child Dev.* (2014) 85:6–20. doi: 10.1111/cdev.12205
16. Rutter M. Resilience as a dynamic concept. *Dev Psychopathol.* (2012) 24:335–44. doi: 10.1017/S0954579412000028
17. Taylor CL, Zubrick SR. Predicting children's speech, language and reading impairment over time. *Int J Speech Lang Pathol.* (2009) 11:341–3. doi: 10.1080/17549500903161561
18. Ukoumunne OC, Wake M, Carlin J, Bavin EL, Lum J, Skeat J, et al. Profiles of language development in pre-school children: a longitudinal latent class analysis of data from the Early Language in Victoria Study. *Child Care Health Dev.* (2012) 38:341–9. doi: 10.1111/j.1365-2214.2011.01234.x
19. McCormack JMS, Mc Allister L, Harrison L. A systematic review of the association between childhood speech impairment and participation across the lifespan. *Int J Speech-Language Pathol.* (2009) 2:155–70. doi: 10.1080/17549500802676859
20. McLeod S, Harrison LJ. Epidemiology of speech and language impairment in a nationally representative sample of 4 to 5 years old children. *J Speech Language Hearing Res.* (2009) 52:1213–29. doi: 10.1044/1092-4388(2009/08-0085)
21. Tomblin JB, Records NL, Buckwalter P, Zhang X, Smith E, O'Brien M. Prevalence of specific language impairment in kindergarten children. *J Speech Lang Hear Res.* (1997) 40:1245–60. doi: 10.1044/jslhr.4006.1245
22. Snyderman R. Personalized health care: from theory to practice. *Biotechnol J.* (2012) 7:973–9. doi: 10.1002/biot.201100297
23. Syurina E, Hens K, Feron F. Literature review: use of family history for primary paediatric care as the next step towards use of genomics in healthcare. *Curr Pediatr Rev.* (2013) 9:353–72. doi: 10.2174/157339630904131223112559
24. Hoff E. How social contexts support and shape language development. *Dev Rev.* (2006) 26:55–88. doi: 10.1016/j.dr.2005.11.002
25. Fukkink R, Jilink L, Oostdam R. A meta-analysis of the impact of early childhood interventions on the development of children in the Netherlands: an inconvenient truth? *Eur Early Childh Educ Res J.* (2017) 25:656–66. doi: 10.1080/1350293X.2017.1356579
26. Levickis P, Sciberras E, McKean C, Conway L, Pezic A, Mensah FK, et al. Language and social-emotional and behavioural well-being from 4 to 7 years: a community-based study. *Eur Child Adolesc Psychiatry.* (2018) 27:849–59. doi: 10.1007/s00787-017-1079-7
27. Conway LJ, Levickis PA, Mensah F, McKean C, Smith K, Reilly S. Associations between expressive and receptive language and internalizing and externalizing behaviours in a community-based prospective study of slow-to-talk toddlers. *Int J Lang Commun Disord.* (2017) 52:839–53. doi: 10.1111/1460-6984.12320
28. McCormack J, McLeod S, Harrison LJ, McAllister L. The impact of speech impairment in early childhood: investigating parents' and speech-language pathologists' perspectives using the ICF-CY. *J Commun Disord.* (2010) 43:378–96. doi: 10.1016/j.jcomdis.2010.04.009
29. Sameroff A. A unified theory of development: a dialectic integration of nature and nurture. *Child Dev.* (2010) 81:6–22. doi: 10.1111/j.1467-8624.2009.01378.x
30. Pakarinen E, Salminen J, Lerkkanen M, Von Suchodoletz A. Reciprocal associations between social competence and language and pre-literacy skills in preschool. *J Early Childh Educ Res.* (2018) 7:207–34.
31. Williams PG, AAP Council on Early Childhood and AAP Council on School Health. The pediatrician's role in optimizing school readiness. *Pediatrics.* (2016) 138:2293. doi: 10.1542/peds.2016-2293

32. van Esso D, del Torso S, Hadjipanayis A, Biver A, Jaeger-Roman E, Wettergren B, et al. Primary-Secondary Working Group of European Academy of Paediatric primary care in Europe: variation between countries. *Arch Dis Child.* (2010) 95:791–5. doi: 10.1136/adc.2009.178459
33. Garner AS. Home visiting and the biology of toxic stress: opportunities to address early childhood adversity. *Pediatrics.* (2013) 132(Suppl.2):S65–73. doi: 10.1542/peds.2013-1021D
34. Kuo AA, Etzel RA, Chilton LA, Watson C, Gorski PA. Primary care pediatrics and public health: meeting the needs of today's children. *Am J Public Health.* (2012) 102:e17–23. doi: 10.2105/AJPH.2012.301013
35. Gan SM, Tung LC, Yeh CH, Chang HY, Wang CH. The ICF-CY-based structural equation model of factors associated with participation in children with autism. *Dev Neurorehabil.* (2014) 17:24–33. doi: 10.3109/17518423.2013.835357
36. Theunissen MHC. *The Early Detection of Psychosocial Problems in Children Aged 0–6 Years by Dutch Preventive Child Health Care: Professionals and Their Tools*. Groningen: University of Groningen (2013).
37. Syurina EV, Feron FJM. Advances in genetics and neuroscience: A challenge for personalizing child and youth health care? In: Hens K, Cutas D, Horstkötter D editors. *Parental Responsibility in the Context of Neuroscience and Genetics.* Maastricht: Springer International Publishing (2017). p. 55–72.
38. Reilly S, McKean C, Morgan A, Wake M. Identifying and managing common childhood language and speech impairments. *BMJ.* (2015) 350:h2318. doi: 10.1136/bmj.h2318
39. van der Vleuten CP, Schuwirth LW, Driessen EW, Govaerts MJ, Heeneman S. 12 Tips for programmatic assessment. *Med Teach.* (2015) 37:641–6. doi: 10.3109/0142159X.2014.973388
40. Glascoe FP. *Collaborating with Parents: Using Parents' Evaluation of Developmental Status (PEDS) to Detect and Address Developmental and Behavioural Problems.* Nolensville, TN: Pedstest (2013).
41. Kaurin A, Egloff B, Stringaris A, Wessa M. Only complementary voices tell the truth: a reevaluation of validity in multi-informant approaches of child and adolescent clinical assessments. *J Neural Transm.* (2016) 123:981–90. doi: 10.1007/s00702-016-1543-4
42. Thompson EJ, Beauchamp MH, Darling SJ, Hearps SJC, Brown A, Charalambous G, et al. Protocol for a prospective, school-based standardisation study of a digital social skills assessment tool for children: the Paediatric Evaluation of Emotions, Relationships, and Socialisation (PEERS) study. *BMJ Open.* (2018) 8:e016633. doi: 10.1136/bmjopen-2017-016633
43. Doove B, Feron J, Feron F, van Os J, Drukker M. Validation of short instruments assessing parental and caregivers' perceptions on child health and development for personalized prevention. *Clin Child Psychol Psychiatry.* (2019) 24:608–30. doi: 10.1177/1359104518822673
44. Marks KP, Glascoe FP, Macias MM. Enhancing the algorithm for developmental-behavioral surveillance and screening in children 0 to 5 years. *Clin Pediatr.* (2011) 50:853–68. doi: 10.1177/0009922811406263
45. Glascoe FP. Are overreferrals on developmental screening tests really a problem? *Arch Pediatr Adolesc Med.* (2001) 155:54–9. doi: 10.1001/archpedi.155.1.54
46. Theunissen MHC, Vogels AG, de Wolff MS, Crone MR, Reijneveld SA. Comparing three short questionnaires to detect psychosocial problems among 3 to 4-years olds. *BMC Pediatr.* (2015) 15:84. doi: 10.1186/s12887-015-0391-y
47. Potijk MR, de Winter AF, Bos AF, Kerstjens JM, Reijneveld SA. Co-occurrence of developmental and behavioural problems in moderate to late preterm-born children. *Arch Dis Child.* (2016) 101:217–22. doi: 10.1136/archdischild-2015-308958
48. Kerstjens JM, Bos AF, ten Vergert EM, de Meer G, Butcher PR, Reijneveld SA. Support for the global feasibility of the Ages and Stages Questionnaire as developmental screener. *Early Hum Dev.* (2009) 85:443–7. doi: 10.1016/j.earlhumdev.2009.03.001
49. Sim F, Thompson L, Marryat L, Ramparsad N, Wilson P. Predictive validity of preschool screening tools for language and behavioural difficulties: a PRISMA systematic review. *PLoS ONE.* (2019) 14:e0211409. doi: 10.1371/journal.pone.0211409
50. Woolfenden S, Eapen V, Williams K, Hayen A, Spencer N, Kemp L. A systematic review of the prevalence of parental concerns measured by the Parents' Evaluation of Developmental Status (PEDS) indicating developmental risk. *BMC Pediatr.* (2014) 14:231. doi: 10.1186/1471-2431-14-231
51. Limbos MM, Joyce DP. Comparison of the ASQ and PEDS in screening for developmental delay in children presenting for primary care. *J Dev Behav Pediatr.* (2011) 32:499–511. doi: 10.1097/DBP.0b013e31822552e9
52. Goodman R. Psychometric properties of the strengths and difficulties questionnaire. *J Am Acad Child Adolesc Psychiatry.* (2001) 40:1337–45. doi: 10.1097/00004583-200111000-00015
53. van Widenfelt BM, Goedhart AW, Treffers PD, Goodman R. Dutch version of the Strengths and Difficulties Questionnaire (SDQ). *Eur Child Adolesc Psychiatry.* (2003) 12:281–9. doi: 10.1007/s00787-003-0341-3
54. Stone LL, Otten R, Engels RC, Vermulst AA, Janssens JM. Psychometric properties of the parent and teacher versions of the strengths and difficulties questionnaire for 4- to 12-years-olds: a review. *Clin Child Fam Psychol Rev.* (2010) 13:254–74. doi: 10.1007/s10567-010-0071-2
55. Laurent di Angulo MS. *Ontwikkelingsonderzoek in de jeugdgezondheidszorg.* Van Gorcum: Assen (2008).
56. Jacobusse G, van Buuren S, Verkerk PH. An interval scale for development of children aged 0–2 years. *Stat Med.* (2006) 25:2272–83. doi: 10.1002/sim.2351
57. van Buuren S. Growth charts of human development. *Stat Methods Med Res.* (2014) 23:346–68. doi: 10.1177/0962280212473300
58. van Agt HM, van der Stege HA, de Ridder-Sluiter JG, de Koning HJ. Detecting language problems: accuracy of five language screening instruments in preschool children. *Dev Med Child Neurol.* (2007) 49:117–22. doi: 10.1111/j.1469-8749.2007.00117.x
59. Harrison LJ, McLeod S, McAllister L, McCormack J. Speech sound disorders in preschool children: correspondence between clinical diagnosis and teacher and parent report. *Austr J Learn Difficult.* (2017) 22:35–48. doi: 10.1080/19404158.2017.1289964
60. Diepeveen FB, Dusseldorp E, Bol GW, Oudesluys-Murphy AM, Verkerk PH. Failure to meet language milestones at 2 years of age is predictive of specific language impairment. *Acta Paediatr.* (2016) 105:304–10. doi: 10.1111/apa.13271
61. Diepeveen FB, van Dommelen P, Oudesluys-Murphy AM, Verkerk PH. Concise tool based on language milestones identifies children with specific language impairment at 24–45 months of age. *Acta Paediatr.* (2018) 107:2125–30. doi: 10.1111/apa.14596
62. StataCorp., *Stata Statistical Software: Release 15*. TX: StataCorp LP, College Station, Texas (2017).
63. Klee T, Pearce K, Carson DK. Improving the positive predictive value of screening for developmental language disorder. *J Speech Lang Hear Res.* (2000) 43:821–33. doi: 10.1044/jslhr.4304.821
64. Roulstone S, Peters TJ, Glogowska M, Enderby P. A 12-months follow-up of preschool children investigating the natural history of speech and language delay. *Child Care Health Dev.* (2003) 29:245–55. doi: 10.1046/j.1365-2214.2003.00339.x
65. Bishop DV. Ten questions about terminology for children with unexplained language problems. *Int J Lang Commun Disord.* (2014) 49:381–415. doi: 10.1111/1460-6984.12101
66. Snowling MJ, Bishop DV, Stothard SE, Chipchase B, Kaplan C. Psychosocial outcomes at 15 years of children with a preschool history of speech-language impairment. *J Child Psychol Psychiatry.* (2006) 47:759–65. doi: 10.1111/j.1469-7610.2006.01631.x
67. Whitehouse AJ, Robinson M, Zubrick SR. Late talking and the risk for psychosocial problems during childhood and adolescence. *Pediatrics.* (2011) 128:e324–32. doi: 10.1542/peds.2010-2782
68. McLaughlin MR. Speech and language delay in children. *Am Fam Physician.* (2011) 83:1183–8.
69. Carter AS, Briggs-Gowan MJ, Davis NO. Assessment of young children's social-emotional development and psychopathology: recent advances and recommendations for practice. *J Child Psychol Psychiatry.* (2004) 45:109–34. doi: 10.1046/j.0021-9630.2003.00316.x
70. Hielkema M, de Winter AF, de Meer G, Reijneveld SA. Effectiveness of a family-centered method for the early identification of social-emotional and behavioral problems in children: a quasi experimental study. *BMC Public Health.* (2011) 11:636. doi: 10.1186/1471-2458-11-636
71. Spijkers W, Jansen DE, de Meer G, Reijneveld SA. Effectiveness of a parenting programme in a public health setting: a randomised controlled trial of the

positive parenting programme (Triple P) level 3 versus care as usual provided by the preventive child healthcare (PCH). *BMC Public Health.* (2010) 10:131. doi: 10.1186/1471-2458-10-131
72. Lavigne JV, Binns HJ, Christoffel KK, Rosenbaum D, Arend R, Smith K, et al. Behavioral and emotional problems among preschool children in pediatric primary care: prevalence and pediatricians' recognition. Pediatric Practice Research Group. *Pediatrics.* (1993) 91:649–55.
73. Lavigne JV, Lebailly SA, Hopkins J, Gouze KR, Binns HJ. The prevalence of ADHD, ODD, depression, and anxiety in a community sample of 4-years-olds. *J Clin Child Adolesc Psychol.* (2009) 38:315–28. doi: 10.1080/15374410902851382
74. American Academy of Pediatrics. Identifying infants and young children with developmental disorders in the medical home: an algorithm for developmental surveillance and screening. *Pediatrics*. (2006) 118:405–20. doi: 10.1542/peds.2006-1231
75. Luby JL, Gaffrey MS, Tillman R, April LM, Belden AC. Trajectories of preschool disorders to full DSM depression at school age and early adolescence: continuity of preschool depression. *Am J Psychiatry*. (2014) 171:768–76. doi: 10.1176/appi.ajp.2014.13091198
76. Gill AM, Hyde LW, Shaw DS, Dishion TJ, Wilson MN. The Family Check-Up in early childhood: a case study of intervention process and change. *J Clin Child Adolesc Psychol.* (2008) 37:893–904. doi: 10.1080/15374410802359858
77. Bayer JK, Rapee RM, Hiscock H, Ukoumunne OC, Mihalopoulos C, Wake M. Translational research to prevent internalizing problems early in childhood. *Depress Anxiety*. (2011) 28:50–7. doi: 10.1002/da.20743
78. Wolraich ML, Felice ME, Drotar D. *The Classification of Child and Adolescent Mental Diagnosis in Primary Care: Diagnostic and Statistical Manual for Primary Care (DSM-PC) Child and Adolescent Version*. Elk Grove Village, IL: American Academy of Pediatrics (1996).
79. de Wolff MS, Theunissen MH, Vogels AG, Reijneveld SA. Three questionnaires to detect psychosocial problems in toddlers: a comparison of the BITSEA, ASQ:SE, and KIPPPI. *Acad Pediatr*. (2013) 13:587–92. doi: 10.1016/j.acap.2013.07.007
80. Tjin-A-Tsoi TBPM. *Jaarrapport integratie 2016.* Den Haag: Centraal Bureau voor de Statistiek (2016).
81. Bishop DV, Price TS, Dale PS, Plomin R. Outcomes of early language delay: II. Etiology of transient and persistent language difficulties. *J Speech Lang Hear Res.* (2003) 46:561–75. doi: 10.1044/1092-4388(2003/045)

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.

frontiers in Psychology

BOOK REVIEW
published: 25 June 2021
doi: 10.3389/fpsyg.2021.710903

Book Review: Current Perspectives on Child Language Acquisition: How Children Use Their Environment to Learn

Xiaoling Zhang[1,2], *Xiaoxiang Chen*[1]* *and Fei Chen*[1]*

[1] *School of Foreign Languages, Hunan University, Changsha, China,* [2] *Department of Foreign Languages, College of Arts and Sciences, National University of Defense Technology, Changsha, China*

Keywords: child language acquisition, usage-based, grammatical constructions, levels of variation, language disorders

A Book Review on

Current Perspectives on Child Language Acquisition: How Children Use Their Environment to Learn

Caroline F. Rowland, Anna L. Theakston, Ben Ambridge, and Katherine E. Twomey (Amsterdam; Philadelphia, PA: John Benjamins Publishing Company), 2020, 330 pages, ISBN: 9789027207074 (hardback), 9789027261007 (e-book)

OPEN ACCESS

Edited and reviewed by:
David Saldaña,
Sevilla University, Spain

Correspondence:
Xiaoxiang Chen
xiaoxiangchensophy@hotmail.com
Fei Chen
chenfeianthony@gmail.com

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

Received: *17 May 2021*
Accepted: *31 May 2021*
Published: *25 June 2021*

Citation:
Zhang X, Chen X and Chen F (2021) Book Review: Current Perspectives on Child Language Acquisition: How Children Use Their Environment to Learn. Front. Psychol. 12:710903. doi: 10.3389/fpsyg.2021.710903

For child language development, the nature vs. nurture debate has been ongoing for a long time, but neither side is presumably persuaded by the other side. Nativists assume children are equipped with learning mechanisms for morpho-syntax innately and the lexicon with access to innate linguistic representations (Chomsky, 2011). In contrast, the usage-based theorists claim that language development is the product of language input as well as children's internal learning mechanism which is domain-general (Lieven, 2016). However, as Karmiloff-Smith (1998) mentioned that "all scientists from the staunchest Chomskian nativist to the most domain-general empiricist agree that development involves contributions from both genes and environment." The main disputes lie in the relative contribution of genes and environment and their interactions. How children interact with the environment in their language development has long fascinated researchers from different language and cultural backgrounds. The book *Current Perspectives on Child Language Acquisition: how children use their environment to learn* is a collection of essays of some researchers who have focused on child-environment interaction, and been influenced by the studies of Elena Lieven (Rowland et al., 2020), to whom this book is dedicated.

The book consists of two parts with a total of 13 essays, 7 in the first part "Levels of Acquisition," which centers on child-environment interactions across different levels of development. The starting chapter provides readers a review of the early communicative development of infants and some of the theoretical perspectives. Chapter 2 discusses several developmental robotics models of the language acquisition, showing the significance for cross-disciplinary collaboration in future theory development and a rising new outlook of language development in which both non-linguistic and linguistic input triggers language development. Then, the following five chapters (Ch3 to Ch7) elaborate the child-environment interaction in morphosyntactic and semantic acquisition from diverse facets, such as grammatical categorization, transitive-causative overgeneralization errors, construction of form-meaning mappings and complex syntactic structures, and Theory of Mind development, mainly based on the previous research of

German and English. Generally speaking, the findings of these research all indicate the significance of language input context.

The second part of the book, "Levels of Variation," includes 6 essays, which discuss variations across individuals, languages and cultures in language development. Among them, two (Ch8&9) focus on language acquisition of typical developing children. Chapter 8 elaborates a dynamic, interactive explanation of gesture development at prelinguistic stage, in which infants' gestures become social over time via interaction with other more experienced speakers. Chapter 9 illustrates how individual differences provide an essential perspective in the language acquisition from gestures to morphosyntax and proposes three casual factors to explain individual differences. Chapters 10 and 13 introduce a few research into Developmental Language Disorders (DLD) and autism and the implications from usage-based theories of language development. Chapter 11 proposes a maximal diversity method, which samples from structurally diversified languages to avoid sampling bias. Chapter 12 discusses language development in bilingualism by detecting factors that account for children's bilingual experience and its position in the input, the interaction between the processing skills and pragmatic skills, and representations across two languages.

As mentioned in the book's introduction, it's the first attempt to bring some of the new perspectives together in one place. It comprises multiple dimensions of child language development: theories vs. empirical evidence, lexical vs. morphosyntactic development, monolingualism vs. bilingualism, and healthy children vs. children with language disorders and etc. It explores how children make use of collective sources of information from environment, construct linguistic representations at several diverse aspects, and learn how to incorporate these representations to conduct effective communication. These findings have enlightened fresh theoretical perspectives which concentrate more on interpreting learning as a complicated dynamic child-environment interaction.

It would be more comprehensive if the book could include more studies of tone languages. Languages of the world exhibit a natural diversity. It is important to note that the existing theories explaining the child language acquisition are mostly based on findings in children from Romance and Germanic language backgrounds such as English, Spanish, German, French, etc. (Singh and Fu, 2016). Actually, not all theories operate universally regardless of the language background. A natural consequence is that the existing theories explaining child language development may not be generalizable to the tone-language-speaking children.

AUTHOR CONTRIBUTIONS

XZ and FC selected the book. XZ drafted the book review. XC and FC provided valuable suggestions and revised it. All authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by Hunan Provincial Education Reform Program (2019).

REFERENCES

Chomsky, N. (2011). Language and other cognitive systems. What is special about language? *Lang. Learn. Dev.* 7, 263–278. doi: 10.1080/15475441.2011.584041

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

Lieven, E. (2016). Usage-based approaches to language development: where do we go from here? *Lang. Cogn.* 8, 346–368. doi: 10.1017/langcog.2016.16

Rowland, C. F., Theakston, A. L., Ambridge, B., and Twomey, K. E. (2020). *Current Perspectives on Child Language Acquisition: How Children Use Their Environment to Learn*. Amsterdam; Philadelphia, PA: John Benjamins Publishing Company. doi: 10.1075/tilar.27

Singh, L., and Fu, C. S. (2016). A new view of language development: the acquisition of lexical tone. *Child Dev.* 87, 834–854. doi: 10.1111/cdev.12512

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.

frontiers in Pediatrics

ORIGINAL RESEARCH
published: 23 November 2021
doi: 10.3389/fped.2021.752141

Language Screening in 3-Year-Olds: Development and Validation of a Feasible and Effective Instrument for Pediatric Primary Care

Daniel Holzinger[1,2,3]*, *Christoph Weber*[2,4], *William Barbaresi*[2,5], *Christoph Beitel*[1] and *Johannes Fellinger*[1,2,6]

[1] Institute of Neurology of Senses and Language, Hospital of St. John of God, Linz, Austria, [2] Research Institute for Developmental Medicine, Johannes Kepler University Linz, Linz, Austria, [3] Institute of Linguistics, University of Graz, Graz, Austria, [4] Department for Inclusive Education, University of Education Upper Austria, Linz, Austria, [5] Division of Developmental Medicine, Department of Pediatrics, Boston Children's Hospital and Harvard Medical School, Boston, MA, United States, [6] Division of Social Psychiatry, University Clinic for Psychiatry and Psychotherapy, Medical University of Vienna, Vienna, Austria

OPEN ACCESS

Edited by:
Satinder Aneja,
Sharda University, India

Reviewed by:
Ramesh Konanki,
Rainbow Children's Hospital, India
Ali Idrissi,
Qatar University, Qatar

***Correspondence:**
Daniel Holzinger
daniel.holzinger@bblinz.at

Specialty section:
This article was submitted to Children and Health, a section of the journal Frontiers in Pediatrics

Received: 02 August 2021
Accepted: 01 November 2021
Published: 23 November 2021

Citation:
Holzinger D, Weber C, Barbaresi W, Beitel C and Fellinger J (2021) Language Screening in 3-Year-Olds: Development and Validation of a Feasible and Effective Instrument for Pediatric Primary Care. Front. Pediatr. 9:752141. doi: 10.3389/fped.2021.752141

Objective: The study was aimed at evaluating the validity and feasibility of SPES-3 (Sprachentwicklungsscreening), a language screening in 3-year-old children within the constraints of regular preventive medical check-ups.

Methods: A four-component screening measure including parental reports on the child's expressive vocabulary and grammar based on the MacArthur Communicative Development Inventory and pediatrician-administered standardized assessments of noun plurals and sentence comprehension was used in a sample of 2,044 consecutively seen children in 30 pediatric offices. One-hundred forty-four children (70 who failed and 74 who passed the screener) comprised the validation sample and also underwent follow-up gold standard assessment. To avoid verification and spectrum bias multiple imputation of missing diagnosis for children who did not undergo gold standard assessment was used. Independent diagnoses by two experts blinded to the screening results were considered gold standard for diagnosing language disorder. Screening accuracy of each of the four subscales was analyzed using receiver operator characteristic (ROC) curves. Feasibility was assessed by use of a questionnaire completed by the pediatricians.

Results: The two parental screening subscales demonstrated excellent accuracy with area under the curve (AUC) scores of 0.910 and 0.908 whereas AUC scores were significantly lower for the subscales directly administered by the pediatricians (0.816 and 0.705). A composite score based on both parental screening scales (AUC = 0.946) outperformed single subscales. A cut off of 41.69 on a *T*-scale resulted in about 20% positive screens and showed good sensitivity (0.878) and specificity (0.876). Practicability, acceptability and sustainability of the screening measure were mostly rated as high.

Conclusion: The parent-reported subscales of the SPES-3 language screener are a promising screening tool for use in primary pediatric care settings.

Keywords: pediatric, feasibility, validity, pre-school age, language disorder, language-delayed children

INTRODUCTION

Depending on the definition used, 2–10% of pre-school-age children experience delayed language acquisition, which makes language disorder (LD) one of the most prevalent developmental disorders (1, 2). However, there is no generally accepted definition of what constitutes a LD. A recent consensus statement on terminology and criteria for language problems in children (3) has resulted in the endorsement of the term Developmental Language Disorder (DLD) for language difficulties that are associated with functional impairment and poor prognosis but have no known biomedical etiology. DLD continues to be a clinical diagnosis since functional impairment and prognosis need to be assessed by appropriately trained clinicians. In addition, the degree of language delay and the linguistic dimensions (phonology, vocabulary, morphology, syntax, pragmatics) and modalities (expressive and receptive) encompassed have not been specified in the consensus document.

In an English population study, Norbury et al. (2) found a prevalence of DLD (of unknown origin) of 7.58%, while 2.34% of LDs were associated with intellectual disability and/or a medical diagnosis (total approximately 10% of LDs from all causes). They defined DLD as scores of −1.5 standard deviations (SD) and below on at least two of five language domains. Similarly, other researchers (4–6) classified a child as "specifically language impaired" whenever language performance was below −1.25 SD in at least two language domains measured by norm-referenced tests. In the absence of a generally accepted measurable gold standard for the definition of LDs, we based our definition on the previously mentioned classifications, which are commonly used in research, with an expected prevalence rate of about 10%.

LDs can affect multiple domains of development through adolescence and adulthood. Children with delayed language development are at increased risk for poor socio-emotional, health (7), behavior and academic outcomes (8, 9) and later unemployment (10) with corresponding costs and loss of human potential.

There is growing evidence that intervention for children with LD may be effective. Direct treatments by a specialist and indirect treatments mediated by caregivers have been shown to have positive effects (11–14). Hence, it is essential to identify children who require educational or therapeutic support in their language learning in order to offer timely and effective intervention.

Based on a number of methodological problems identified in their systematic review of language screening (e.g., lack of information on the effects of age, setting and administrator on screening accuracy) and insufficient evidence for long-term outcomes of language interventions, Nelson et al. (15) did not recommend universal language screening. In a more recent systematic review, Wallace et al. (16) reported on the accuracy of some screening measures for identification of children with language impairment although evidence of feasibility in primary-care settings remained inadequate. Consequently, the U.S. Preventive Taskforce did not recommend universal screening for language delay (17). Given the international systematic reviews (1, 15) and following the National Health and Medical Research Council (18), the German Institute for Quality and Efficiency in Health Care (19) also considered the available evidence insufficient to recommend the implementation of language screening in Germany. In the same vein, a systematic review evaluating the effectiveness of systematic population-based screening for specific language impairment in pre-school children in Germany (20) concluded that the accuracy of German screening measures had not yet been sufficiently examined.

Systems for general health check-ups have been established in many countries in the Western world and could be suitable opportunities for identification of language delays. However, due to a lack of accurate and feasible instruments, routine well-baby check-ups are generally not used for systematic language screening. One exception is in the Netherlands, where a five-minute interview/test procedure (VTO language screening) administered by a youth health care physician to 24 month-old children led to more cases with language impairment being identified than by the regular procedure (0.4–2.4%), and at a significantly younger age than in regions in which the regular detection procedures were used (21). However, as demonstrated by the low number of cases, many 2-year-old children with language impairment were not identified (low sensitivity of 24–52%). Given the high instability of language development trajectories in young children, Law et al. (1) recommended the age span of 3–5 years as the optimal period for language screening.

A combination of (i) observations by parents with extensive long-time knowledge of their children's behaviors in everyday life and (ii) standardized assessments by pediatricians is essential to avoid assessment bias (22). A comparative study of direct assessments and parent reports of language and pragmatics revealed patterns of difference that indicated a need to collect assessment data from multiple informants (23). Nevertheless, convergent validity of parent-reported tools in comparison to direct assessments of language is usually high (24, 25). Mere elicitation of parental concerns has therefore also been described as a valid method for identifying an increased risk for developmental disorders (26–28).

Particularly at the age of 3 years, grammatical knowledge is a good marker of language development (29–31). Furthermore, grammatical skills can usually be reliably assessed within a shorter time than expressive or receptive vocabulary knowledge. In addition, a comprehensive measure for identifying language disorders must refer to both production and comprehension

(4). Another reason for including language reception in a screening measure is its ability to predict the persistence of language difficulties (32). As parental reports of language comprehension have often been found to overestimate children's skills (33), a direct assessment of language comprehension should be considered.

Upper Austria is a federal state with a population of 1.45 million inhabitants (with 13.297 births in 2007). Well-child visits are provided free of charge by community pediatricians and general practitioners. In 2010, when the validation study was conducted, 9,125 health check-ups (68.6% of the children born in 2007) were carried out with 3-year-old children, 66% of which by pediatricians and the remainder by general practitioners.

The aim of this study was the validation of a screening tool for LDs for 3-year-old German-speaking children in a pediatric primary-care setting in Upper Austria. The new instrument, the SPES-3 (Sprachentwicklungsscreening) that has been developed in a pilot study includes both parent observations and direct assessments of the child by the primary-care pediatrician. As required for comprehensive language assessments, the screening measure takes various linguistic domains (grammar and vocabulary) and modes (receptive and expressive) into account. For use in primary-care settings, the screening tool must have high acceptability and require little time to administer while maximizing accuracy.

METHODS AND PROCEDURES

Construction of the Screening Measure and Pilot Testing

Both parent-administered screening scales are based on the MacArthur-Bates Communicative Development Inventories (MCDI), Level III (34), and require parents to systematically report their observations of their child's development of expressive grammar and expressive vocabulary. Inspired by the MCDI concept for grammar assessment a number of morphosyntactic structures of German that typically emerge around age 3 were selected to be presented to the child in the form of 27 pairs of correct and incorrect options. Parents are asked to select the option they are more likely to observe in their child's spontaneous use of language. For expressive vocabulary, 100 words from the MCDI-III English word list (34) translated into Austrian German (35) were chosen. Parents are asked to indicate whether their child uses the words in their expressive communication or not.

The screening subscales administered by the pediatricians were compiled from pre-existing subtests of a German standardized language test (Sprachentwicklungstest SETK 3-5; (36), The first subscale includes 20 items that assess the production of noun plurals: The pediatrician presents and names a pictured item in the singular and asks the child to produce the respective plural form supported by a picture that shows several identical items. The second subscale assesses sentence comprehension: Single sentences are read aloud (9 items), and the child is asked to point to the corresponding picture from a selection of four. For an overview of all screening scales (see **Table 1**).

TABLE 1 | Screening subscales.

	Parent report	Pediatric assessment
Vocabulary expressive	Expressive vocabulary (MCDI-3 word list, Austrian version)	
Grammar expressive	Expressive grammar (inspired by MCDI-3)	Noun plurals (SETK 3-5)
Grammar receptive		Sentence comprehension (SETK 3-5)

MCDI-3, MacArthur-Bates Communicative Development Inventories Level III; SETK 3-5, Sprachentwicklungstest (language test for age 3–5 years).

The preliminary version of the screening procedure was first used in 2006 in a pilot study in close cooperation with the Pediatric Association and the Department of Health and Social Affairs in Upper Austria, with the aim of developing language screening instruments to be used in pediatric primary care within regular health check-ups at ages 2 and 3. The pilot study included 1,730 non-preselected 3-year-old children with German as their first language, who were consecutively assessed by a group of 24 primary care pediatricians recruited from across the state of Upper Austria (whose participation was voluntary).

Based on the pilot study, the *parental report of expressive grammar* scale, which initially contained 27 items, was shortened by excluding 14 items with low response rates, low difficulty and low item-scale correlation. Thus, the final screening measure used in the validation study covered 13 items for expressive grammar, 100 items for expressive vocabulary, 20 items for noun plurals and 9 items for sentence comprehension. All items were scored either as 0 (does not apply/false) or 1 (applies/correct), and for each subscale a sum over the item scores was computed. Internal consistency (Kuder and Richardson Formula 20) was excellent for the parental screening scales ($\rho_{\text{KR20}} = 0.98$ for expressive vocabulary and 0.90 for expressive grammar) and adequate for the screening scales directly administered with the children [$\rho_{\text{KR20}} = 0.78$ for noun plurals and sentence comprehension; (36)] For the questionnaires to be completed by the parents (expressive grammar and expressive vocabulary) cut-off scores for the validation study were derived from the 10th percentiles (SD 1.25) based on the final instruments, as suggested by the literature (4–6). Normative data to determine cut-off scores were available for both of the standardized assessments to be completed by pediatricians (noun plurals and sentence comprehension).

Gold Standard Assessment of Language Disorder

Given the lack of well-defined standards for the diagnosis of LDs, independent expert diagnoses by two experienced clinical linguists blinded to the language-screening results were considered the gold standard. Their diagnostic decisions on LD were based on the results of two standardized tests assessing the linguistic domains of expressive sentence grammar, noun plural production, sentence comprehension [SETK 3-5; (36)] and expressive vocabulary [AWST-R; (37)] that had been administered by other linguists and a short video sample (5–10 min) of spontaneous language of each child in a play and/or dialogic picture-book situation. For their decision on LD both clinical linguists followed international research (4–6) and classified a child as language delayed when language performance was at about the 10th percentile or lower in at least two of the four measured language domains and observations of spontaneous language production confirmed the significant language difficulties. Inter-rater reliability (kappa) between the linguists' diagnoses (+/− LD) was 0.95. Discrepancies were resolved by consensus decisions between the two raters.

For the assessment of non-verbal intelligence, the Snijders Oomen Non-verbal Intelligence Test [SON-R 2 ½-7; (38)] was administered by clinical psychologists. Pure tone audiometry was used to assess hearing.

Feasibility Measures

Feasibility was measured primarily by use of a questionnaire completed by the pediatricians who participated in the study and by the completeness of screening tests administered. Following the guidelines suggested by Bowen et al. (39), four dimensions of feasibility were investigated. All questions of the pediatric questionnaire were rated on a 4-point Likert scale (very good-good-difficult-very difficult).

Practicality

Practicality was operationalized by the extent to which administration of the screening was considered possible within the time constraints of pediatric primary care. In addition, the pediatricians were asked about the ease of administration of each of the two screening measures (noun plurals and sentence comprehension) and to evaluate parental difficulties in completing the two parental subscales of the screening. In addition, the pediatricians ranked five pre-specified factors that might challenge the completion of the screening measure within their respective settings.

Acceptability

Acceptability refers to the children's, parents' and pediatricians' reactions to the screening measures. Child acceptance was measured by the percentage of screening subscales administered by the pediatricians that were fully completed, as this reflects child compliance with the test. In addition, the pediatricians assessed parental acceptance of the inclusion of language screening in the 3-year medical check-up. Finally, the pediatricians were asked to rate the meaningfulness of including a language screening within the regular well-baby check-ups.

Sustainability

Sustainability refers to the likelihood of language screening in this form being continued within the present system of preventive medical care in Austria. Pediatricians were asked in the questionnaire whether they intended to continue the language screening after the study ended (yes, to a limited extent, no).

Study Procedures and Recruitment

In 2009, all primary care pediatricians of the province of Upper Austria were invited to participate in the validation study of the language screening. Thirty-six out of 60 pediatricians participated in a half-day training and in the subsequent implementation of the screening procedures. The participating pediatricians served all major geographical areas of Upper Austria. Over a 1-year period (2010) 2,635 3-year-old children (19.8% of the entire 2007 Upper Austrian birth cohort) were screened by the 30 pediatricians who ultimately participated in the study.

Overall, 591 children were excluded from the study: 31 children with missing data on their date of birth, one child with missing data on the screening date, 95 children who were outside the age range (34–38 months), 349 children from families whose primary family language was other than German, 95 children with missing data on both parental screening subscales and 22 with missing data on both pediatric screening subscales. The remaining sample ($n = 2,044$) equaled 15.4% of the entire 2007 birth cohort and about 22% of the children born in 2007 by native-born mothers. The mean age was 36.03 months (SD = 0.994). 50.9% were boys, 4.9% were multiple births, 50.7% of the children had older siblings, and 8.6% of children were born prematurely. Compared to the 2007 birth cohort for Upper Austria, the sample was representative in terms of sex ratio and prematurity rate. Multiple births were slightly overrepresented (4.9% vs. 3.4% in the birth cohort; $\chi^2(1) = 14.003, p < 0.001$). To assess the representativeness of the sample in terms of maternal education (proportion of mothers with university entrance qualification or tertiary degree), we calculated an age-adjusted comparison value based on the educational level of all women (i.e., not only mothers) from Upper Austria. Among the participating families the percentage of mothers with either a university entrance qualification or a tertiary degree was comparable to the age-adjusted population (39.4% vs. 41.6%; $\chi^2(1) = 4.072, p < 0.05$).

The full screening was administered in the course of well-child visits in the pediatric medical practices. Parents completed a screening package that included demographic information in addition to their language observations (expressive grammar and expressive vocabulary). All parents gave written permission for scientific use of their children's anonymized data for scientific purposes. The study was approved by the ethics committee of the Hospital of St. John of God, Linz. Pediatricians administered two screening subscales (noun plurals and sentence comprehension). All screening data were sent back to the clinic conducting the research. In accordance with Tomblin et al. (40), a screening test was considered a fail if the results of at least two of the four subtests were 1.25 standard deviations below the age norm. Using this ex-ante definition for positive screening results, 21.7% of the sample was considered a screening fail.

To validate the screening measures, a sample for full gold standard assessment was recruited in a two-step procedure. All pediatricians were instructed to refer children with positive screening results to a single specialized program for gold standard examination, which was performed in 70 children. Second, to evaluate specificity of the screening tool, four pediatricians from different regions were asked to invite a random subsample of children from the whole cohort, excluding those ($n = 70$) who had already undergone the gold standard assessment. The sample was stratified by gender, maternal education, position among siblings and single parenthood. The recruitment procedures (see **Figure 1**) led in total to a validation sample of $n = 144$ (i.e., 7% of the total sample), that did not deviate from the remaining sample in terms of demographic variables (see sample description).

Analytic Strategy

First, we report descriptive statistics for the total sample and the validation sample. Second, we present results on the screening accuracy of each subscale based on receiver operator characteristic (ROC) analyses. Areas under the curve (AUCs) $\geq$0.9 are regarded as excellent, AUCs $\geq$0.8 and <0.9 as good, AUCs $\geq$0.7 and <0.8 as fair, and tests with AUCs <0.7 as poor (41). DeLong's method for comparing AUCs of different tests (42) was applied to analyze differences in diagnostic accuracy between the subtests. Further, logistic regression was applied to identify subscales with significant independent contributions to the prediction of the gold standard diagnosis. The aim of this analytic step was to reduce the number of screening subtests while optimizing diagnostic accuracy. Lastly, we evaluated possible cut-off scores for the final screening composite by estimating sensitivity, specificity, positive predictive values (PPV), negative predictive values (NPV), and diagnostic likelihood ratios for positive and negative screening results (DLR+ and DLR–, respectively). DLR+ and DLR– are alternative measures of the accuracy of a diagnostic test and have the advantage, unlike predictive values, not to depend on the prevalence of the disorder under investigation [(43); for an explanation see **Supplementary Material**]. DLR+ is the multiplicative change in the pre-screening odds of having a LD given a positive screening result (i.e., post-screening odds = DLR+ × pre-screening odds) and DLR– is the change in the pre-screening odds of having a LD given a negative screening result (post-screening odds = DLR– × pre-screening odds). Following Jaeschke et al. (44), DLR+ values $\geq$ 10 and DLR– $\leq$ 0.1 indicate large changes in pre-screening odds, DLR+ $\leq$ 10 and >5, and DLR– > 0.1 and $\leq$0.2 indicate moderate changes, DLR+ $\leq$ 5 and >2, and DLR– > 0.2 and $\leq$0.5 indicate small changes. DLR+ < 2 and DLR– > 0.5 are rarely important. Logistic regression models were conducted using M*plus* 8 (45), for the ROC analysis the pROC package (46) in R was employed, and cut-off values were determined with the R-OptimalCutpoints package (47).

Notably, the recruitment procedure led to oversampling of children with positive screening results, which is not uncommon for screening validation studies (48). Ideally, the validation sample should reflect the patient population, the given overrepresentation of positive screening results induces bias in measures of screening accuracy [i.e., verification bias; (49)]. In order to deal with this bias, we used multiple imputation (MI) of missing diagnosis status for children who did not undergo gold standard assessment (50). Missing values on screening subscales (ranging between 0.4 and 12.8% in the full sample) were also imputed. We used the Blimp imputation software (51, 52) to generate 50 imputed data sets using a chained equation imputation procedure that takes the clustering of children within pediatricians into account. Beside the study variables (i.e., sociodemographic variables and screening results) we used various auxiliar variables (e.g., parental concern about language development) that were predictive of diagnosis status. Estimates and their standard errors were computed according to Rubin's combining rules (53). Even though recent simulation studies do not indicate that high proportions (up to 90%) of missing data might bias estimates based on MI data sets (54), we also report results for the original validation sample ($n = 144$) as **Supplementary Material** for completeness.

RESULTS

Figure 2 shows the distributions of the screening subscales. As expected, the scores on all scales were skewed to the left. Means and standard deviations were $M = 10.22$ (SD = 3.67) for expressive grammar and $M = 73.00$ (SD = 21.76) for expressive vocabulary, $M = 12.05$ (SD = 5.19) for noun plurals, $M = 7.48$ (SD = 1.48) for sentence comprehension. The correlations between screening subscales were moderate to high. The parental scales correlated with $r = 0.56$ ($p < 0.001$). The correlation of the pediatricians' scales was $r = 0.40$ ($p < 0.001$). Further correlations were: $r_{\text{expressive vocabulary, sentence comprehension}} = 0.30$, $p < 0.001$; $r_{\text{expressive vocabulary, noun plurals}} = 0.40$, $p < 0.001$; $r_{\text{expressive grammar, sentence comprehension}} = 0.36$; $p < 0.001$; $r_{\text{expressive grammar, noun plurals}} = 0.43$; $p < 0.001$.

Moreover, analyses showed that children with positive screening results (ex-ante definition) who attended the gold standard assessment showed lower scores in three subscales than children with positive screening results who did not attend the assessment (expressive grammar: $d = 0.60$, $p < 0.001$; sentence comprehension: $d = 0.29$, $p < 0.05$; noun plurals: $d = 0.51$; $p < 0.001$). Thus, the validation sample may also be subject to spectrum bias (i.e., screening positives include primarily the "sickest of the sick" and not the full spectrum of positive screens), which is associated with overestimation of sensitivity, specificity, and AUC (49). However, the used multiple imputation procedure is also suitable to counteract spectrum bias. Based on the gold standard, 27.8% of the validation sample ($n = 144$) had a LD. Notably, all children with LD had positive screening results (i.e., sensitivity = 1.00). After imputation 11.7% of the children are classified as having a LD.

As children are clustered within pediatricians, we estimated the intraclass correlation (ICC) to evaluate whether there are differences between pediatricians in the subscales. We found that pediatricians accounted for ≈14% of the differences in the pediatrician-reported subscales ($ICC_{\text{noun plurals}} = 0.144$, $p < 0.001$; $ICC_{\text{sentence comprehension}} = 0.139$, $p < 0.001$). If differences between pediatricians were due to population differences in the catchment areas, we would also expect comparable ICCs for the parent reports. However, ICCs for parent reports were smaller ($ICC_{\text{expressive vocabulary}} = 0.049$, $p = 0.005$, $ICC_{\text{expressive grammar}} =$

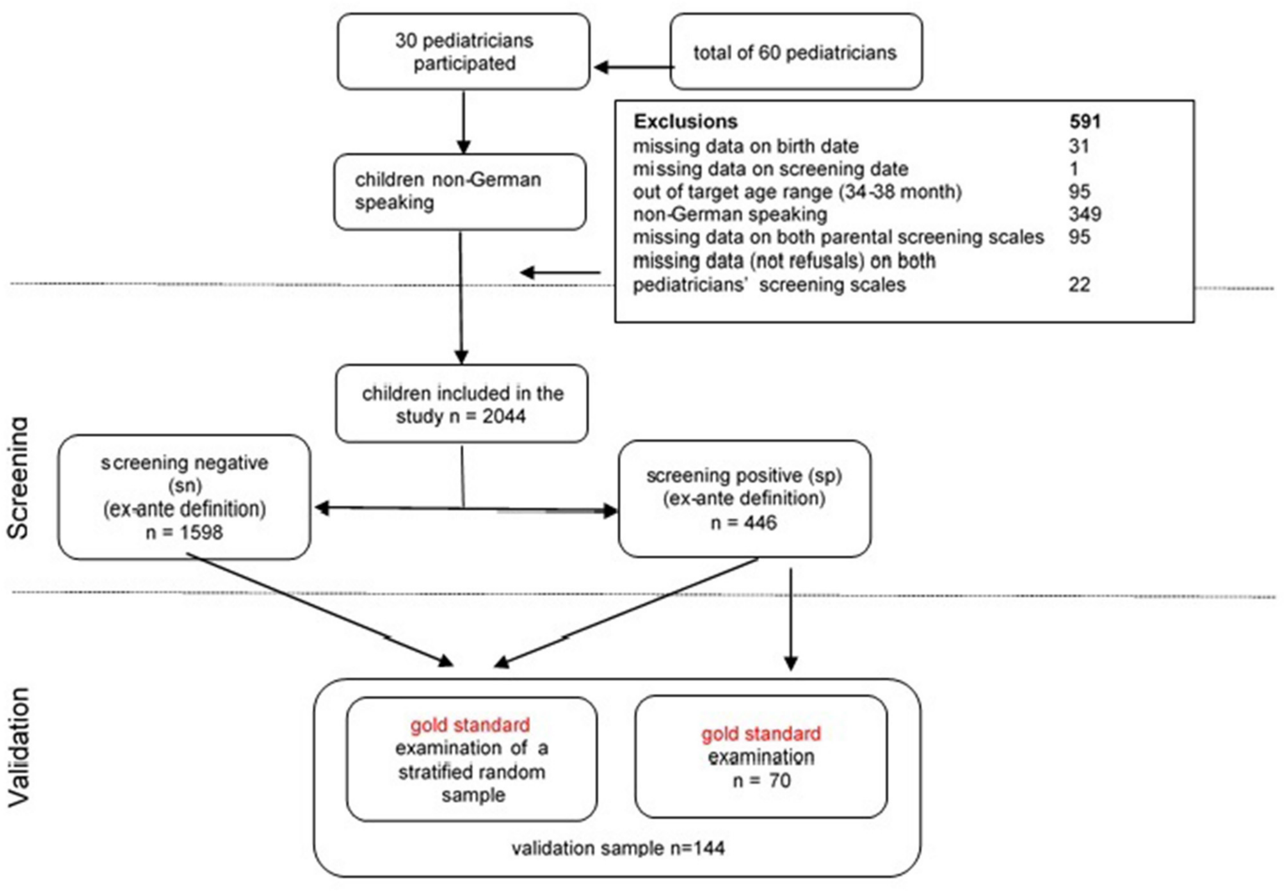


FIGURE 1 | Recruitment and participation.

0.021, $p = 0.05$). Thus, these findings indicate that pediatricians differed significantly in their application of the screening tools, which calls into question the objectivity of implementation.

Diagnostic Accuracy of Subscales

Results of the ROC analysis for the screening subscales are shown in **Table 2**. The AUCs ranged from fair ($AUC_{sentence\ comprehension} = 0.705$, DeLong 95% CI = [0.623, 0.786]) to excellent ($AUC_{expressive\ grammar} = 0.910$, DeLong 95% CI = [0.859, 0.960]). The second parent-reported subscale showed an almost identical excellent AUC value ($AUC_{expressive\ vocabulary} = 0.908$, DeLong 95% CI = [0.864, 0.952]). The AUC for noun plurals was good ($AUC_{noun\ plurals} = 0.816$, DeLong 95% CI = [0.745, 0.887]). As indicated by DeLong tests for paired ROC curves, parent-reported scales outperformed the screening subscales administered by pediatricians. All AUC differences between parent-reported and pediatrician-administered scales were significant, and noun plurals outperformed sentence comprehension.

To examine independent contributions of subscales to predicting the gold-standard diagnosis logistic regression analyses were performed (**Table 3**). The two parent-reported subscales independently predicted LD. After controlling for parent-reported screening tests, none of the pediatrician-administered subscales significantly predicted LD. Standardized coefficient (b) for the expressive vocabulary subscale was −0.408 ($p < 0.001$) and −0.388 ($p < 0.001$) for the expressive grammar subscale. Notably, as indicated by the overlapping confidence intervals of the standardized logistic regression coefficients, both parent reported scales had roughly the same weight in predicting LD.

Diagnostic Accuracy of the Composite Screening Score

Given the results of the logistic regression models, a composite screening score based on both significant predictors (expressive vocabulary and expressive grammar) was computed. As both parent reported scales contributed almost equally to the prediction of LD, a composite score was computed as the mean of the z-scores of expressive vocabulary and expressive grammar. The AUC for the composite score was excellent at 0.946 (DeLong 95% CI = [0.883, 1.000]). DeLong tests for paired ROC curves indicate that the composite outperformed the single parent reported scales (composite score vs. expressive vocabulary: $\Delta AUC = 0.038$, t-value = 2.380, $p = 0.019$; composite score vs. expressive grammar: $\Delta AUC = 0.036$, t-value = 2.102, $p = 0.037$).

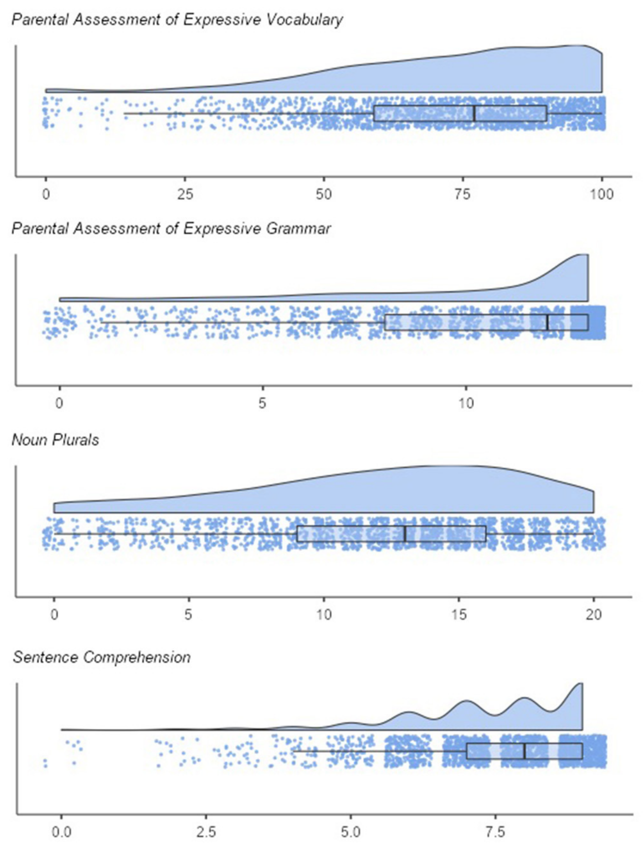


FIGURE 2 | Distribution of the screening subscales (raw scores).

Cut-Off Estimation

In the next step, cut-off values for the screening composite were estimated (**Table 4**). For ease of interpretation, the composite score was transformed into a T metric (i.e., $M = 50$, SD = 10). First, we estimated the cut-off by setting sensitivity equal to specificity using the "SpEqualSe" criterion in the Optimal Cutoff Package (47). A cut-off at 41.69 was most efficient. **Table 5** reports the classification results for this cut-off that resulted in satisfactory accuracy statistics: sensitivity = 0.878 (95%-CI = [0.770, 0.985]), specificity = 0.876 (95%-CI = [0.856, 0.895]), PPV = 0.438 (95%-CI = [0.333, 0.544]), NPV = 0.984 (95%-CI = [0.967, 1.000]), DLR+ = 7.078 (95%-CI = [5.779, 8.378]), DLR– = 0.140 (95%-CI = [0.018, 0.261]). The cut-off would have resulted in 20.0% screening fails and consequently in a relatively high number of clinical evaluations required. Lower cut-offs resulted in fewer screening fails and higher PPV, DLR+, DLR– and specificity, but also in lower sensitivity (see **Table 4**). Thus, lower cut-offs would yield more false-negative results.

Feasibility

Completed questionnaires that addressed feasibility were returned by 23 (77%) out of the 30 pediatricians participating in the study.

TABLE 2 | Diagnostic accuracy of the screening subscales.

	Language disorder		r_{pb}	AUC	95%-CI	AUC-differences—*t*-values (DeLong)		
	Yes	No				Expressive Vocabulary	Expressive grammar	Sentence comprehension
	M (SD)	*M* (SD)						
N = 2,044								
Expressive vocabulary	33.656 (17.563)	70.810 (19.864)	−0.515***	0.908	(0.864; 0.952)			
Expressive grammar	1.851 (2.906)	9.165 (4.103)	−0.556***	0.910	(0.859; 0.960)	−0.039		
Sentence comprehension	5.462 (2.388)	7.026 (1.734)	−0.251***	0.705	(0.623; 0.786)	4.711***	5.804***	
Noun plurals	3.426 (4.492)	10.056 (5.903)	−0.363***	0.816	(0.745; 0.887)	2.440*	2.629*	−2.848**

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$. r_{pb}, point biserial correlation.

TABLE 3 | Logistic regression predicting LD on the basis of screening subtests.

	b (SE)	Stand. *b* [95%-CI]	OR
Expressive vocabulary	−0.057*** (0.015)	−0.408 [−0.579, −0.237]	0.945
Expressive grammar	−0.315*** (0.066)	−0.388 [−0.536, −0.241]	0.730
Sentence comprehension	0.052 (0.153)	0.028 [−0.132, 0.188]	1.053
Noun plurals	−0.092 (0.058)	−0.160 [−0.351, 0.031]	0.912
Threshold	−3.917		
R^2		0.634***	

*OR, odds ratio; ***p <0.001.*

Practicality

According to the pediatricians, 13.6% of the parents considered completion of the questionnaire to be "difficult". However, the remainder rated it most often as "very easy" (47.8%) or "easy" (34.8%). Practicality of the screening measures within the time limits of preventive medical care was rated mostly as good (69.6%) or very good (8.7%). In 21.7% of the cases, it was considered "difficult", but never "very difficult". Ease of administration of the sentence comprehension screening scale was assessed as significantly higher than that for noun plurals (expressive grammar). Ease of administration of the receptive measure was rated in most cases as good (60.9%) and very good (34.8%), and as difficult by only one pediatrician (4.3%). In contrast, 47.8% rated the screening of noun plurals as difficult, 39.1% as good and only three pediatricians (13.0%) as very good. Among the factors that might complicate or even prevent administration of language screening, lack of or insufficient follow-up was ranked highest (40%), followed equally by insufficient training of pediatricians (20%) and insufficient funding (20%), and then by limited meaningfulness of language screening (13.3%). Only one respondent cited time constraints (4.3%).

Acceptability

General parental acceptance of language screening included in preventive health care was rated as "very good" by 43.5% and as "good" by all remaining pediatricians (52.2%). The percentage of children refusing their cooperation in the pediatrician's administration of the assessment of noun plurals was 12.8%, and for the assessment of sentence comprehension 5.5%, indicating satisfying acceptance of the assessment of language reception by the children, although child participation is no longer required for the final screening package that is exclusively based on parent report. A great majority of pediatricians rated the meaningfulness of language screening within the regular medical checkups as very good, another three (13.9%) as good, and only 2 expressed concerns (9.1% "difficult").

Sustainability

Most pediatricians (82.6%) reported that they would continue the SPES-3 language screening beyond the study, while the

TABLE 4 | Diagnostic accuracy statistics for various cut-offs.

Cutoff	%-Screening positives	Sensitivity	Specificity	PPV	NPV	DLR+	DLR-
35	0.092	0.609 (0.475, 0.742)	0.965 (0.954, 0.977)	0.658 (0.538, 0.779)	0.956 (0.929, 0.983)	17.813 (10.418, 25.187)	0.406 (0.268, 0.543)
36	0.111	0.661 (0.532, 0.790)	0.950 (0.936, 0.963)	0.591 (0.470, 0.713)	0.961 (0.936, 0.986)	13.294 (8.735, 17.843)	0.357 (0.211, 0.493)
37	0.124	0.697 (0.568, 0.828)	0.939 (0.924, 0.954)	0.557 (0.438, 0.677)	0.965 (0.941, 0.988)	11.518 (7.936, 15.100)	0.323 (0.187, 0.460)
38	0.138	0.742 (0.615, 0.868)	0.929 (0.913, 0.945)	0.536 (0.0.418, 0.654)	0.969 (0.947, 0.992(	10.547 (7.545, 0.13.548)	0.278 (0.143, 0.413)
39	0.150	0.779 (0.655, 0.903)	0.919 (0.902, 0.936)	0.516 (0.399, 0.632)	0.973 (0.952, 0.995)	9.704 (7.156, 12.252)	0.240 (0.106, 0.375)
40	0.169	0.820 (0.701, 0.939)	0.904 (0.886, 0.922)	0.485 (0.374, 0.596)	0.978 (0.958, 0.997)	8.553 (6.620, 0.10.480)	0.199 (0.068, 0.330)
41	0.189	0.865 (0.755, 0.976)	0.886 (0.867, 0.905)	0.456 (0.348, 0.564)	0.983 (0.965, 1.000)	7.612 (6.127, 9.907)	0.152 (0.028, 0.276)
42	0.202	0.879 (0.773, 0.985)	0.873 (0.853, 0.892)	0.433 (0.328, 0.538)	0.984 (0.967, 1.000)	6.918 (5.678, 8.159)	0.139 (0.018, 0.259)

95%-CI in brackets.

TABLE 5 | Classification table.

	Language disorder No	Language disorder Yes	Total
Screening pass	1,609.90	26.18	1,636.08
Screening fail	229.25	178.67	407.92
Total	1,839.15	204.85	2,044

Decimal absolute values are presented, because results are based on the average across 50 imputed data sets.

remainder would stop. It is noteworthy that no specific additional funding for language screening could be provided.

DISCUSSION

This study evaluated the diagnostic accuracy and feasibility of a newly developed language screening measure (SPES-3) for children aged 3 years in pediatric primary care within regular well-baby check-ups. In addition to the implementation by community pediatricians in consecutively evaluated children, the extensive sample size and the use of multiple imputation to avoid verification and spectrum bias are strengths of this study. The two parent-reported subscales (expressive grammar and vocabulary development) showed excellent accuracy (AUC.910, 0.908). Although AUC scores for pediatric subscales were significantly lower, they were good for the noun-plural subscale (0.816) and fair for the sentence comprehension subscale (0.705). Nevertheless, logistic regression analysis showed that none of the pediatrician-administered subscales significantly increased the diagnostic accuracy achieved by the parent scales. A composite based on both parent reports showed excellent accuracy (AUC = 0.946) and outperformed the single parental subscales. Our findings are consistent with those of a systematic review on the predictive validity of preschool screening tools for language by Sim et al. (55). Their results exhibited higher sensitivity, specificity and negative predictive value of parent-report screening tools as compared to direct child assessment. This finding supports the richness of parental information that is based on long-time observation of their child's language use in a variety of everyday situations. In contrast, the quality of screening tools based on direct assessments is influenced by the brevity required for use in the regular pediatric office and might be influenced by the relationship between child and examiner. The higher intraclass correlations of the screening tests administered by the pediatricians compared to the parent-reported screening tests may, at least partly, reflect differences in building rapport with the child.

Since screening duration in community settings is critical for their large-scale implementation, stronger validity of parent-reported screening in place of direct assessment by a pediatrician can be viewed as a positive finding in terms of feasibility of the screening tool. Practicability of the new screening measure was assessed to be high, even when direct assessments were included in the screening measure. Pediatricians also rated the acceptability by parents and children as high, and the majority of them regarded the inclusion of a language screening measure within their regular well baby check-ups was meaningful.

By use of ROC analysis a cut off of 41.69 is optimal when equal values for sensitivity and specificity are desired. Predictive values, which depend on the prevalence of the disorder, were NPV = 0.984 and PPV = 0.438. Moreover, we evaluated a broader range of cut-offs. As illustrated by **Table 4**, there is a trade-off between sensitivity and specificity. The more sensitive is a cut off value, the less specific it is. Due to increasing risk of persistence of language difficulties by increasing age and for reasons of practicability we suggest the cut-off of 41 for this new language screening measure for 3-year-olds to reduce the number of missed children with delayed language development. The relatively low PPV would lead to overreferrals but it must be interpreted with regard to the relatively low rate of diagnoses of LD (11.7%). As language difficulties have a dimensional rather than categorical character, children with false-positive scores in language skills have been shown to perform significantly lower on diagnostic measures than children with true-negative scores (56). Since children "overreferred" for diagnostic testing perform lower on language, they are very likely to also carry more psycho-social and cognitive risk factors associated with language delay. Therefore, diagnostic testing should not be regarded as an unnecessary inconvenience to the family and expense to society, but as an opportunity to identify children with unmet needs and the interventions required to improve their language and social and academic learning even though they are probably less severely affected than children with LD.

Even though direct assessments by pediatricians did not contribute to the predictive quality of the screening measure its administration within existing systems for general health check-ups might still be considered. Medical reports and recommendations are often highly valued by families. In case of screening fails (parent reported scales) a medical professional may still administer a brief assessment of language comprehension (sentence comprehension subscale), that was very well accepted by the children. A delay in language comprehension can be indicative of a more severe and persisting language problem (3) and may be associated with other more general developmental problems (e.g., general developmental delay, autism spectrum disorder, or hearing loss) and thus require follow-up at a multi-professional diagnostic center. Notably, among the possible barriers to universal language screening, lack of insufficient follow-up was ranked highest by the pediatricians who participated in the study, followed by insufficient training of pediatricians and lack of funding. Screening guided tiered referral pathways to local speech-language therapists and/or specialized multi-professional diagnostic centers (for the smaller number of children with complex needs) might be a cost-efficient approach.

The limited size of the original validation sample, the high proportion of screening fails in the validation sample and a more severe character of the language delay as compared to those without clinical follow-up are limitations of this study. However, multiple imputation of missing diagnosis for children without gold standard assessment was used to deal with these biases. Another limitation is the lack of baseline data on other developmental disorders. Whereas, we have provided data on the representativity of the study sample, we do not have information on differences between participating and non-participating pediatricians.

CONCLUSION

Parent-reported screening measures for expressive vocabulary and grammar (SPES-3) administered within regular well-baby check-ups in pediatric primary care have been found to be accurate in identifying LDs in 3-year-olds. Administration of screening subscales for noun-plurals and sentence comprehension by pediatricians showed lower specificity and sensitivity and, when added to the parental assessments, did not improve overall accuracy of the screening package. Feasibility within regular preventive check-ups was rated as mostly good or very good by the pediatricians. Ease of administration and acceptance by parents and pediatricians demonstrated in implementation with a large cohort of non-preselected children that the SPES-3 screening measure is valuable and can be recommended for universal language screening at age three in pediatric primary care.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because parents of the children participating in the study have not given their consent to make the data publicly available. Requests to access the datasets should be directed to daniel.holzinger@bblinz.at.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethikkommission Konventhospital Barmherzige Brüder Linz. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

DH and JF: conceptualization. DH, JF, CW, WB, and CB: methodology. CW and CB: formal analysis. DH, JF, and CB: investigation. DH and CW: data curation. DH, JF, WB, and CW: writing—original draft preparation and writing—review and editing. DH: project administration. All authors have read and agreed to the published version of the manuscript.

FUNDING

The development of the SPES-3 screener was financially supported by the Department of Health and Social Affairs Upper Austria. Article Processing Charge is funded by the Johannes Kepler University Open Access Publishing Fund.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fped.2021.752141/full#supplementary-material

REFERENCES

1. Law J, Boyle J, Harris F, Harkness A. Prevalence and natural history of primary speech and language delay: findings from a systematic review of the literature. *Int J Lang Commun Disord.* (2000) 35:165–88. doi: 10.1080/136828200247133
2. Norbury CF, Vamvakas G, Gooch D, Baird G, Charman T, Simonoff E. Language growth in children with heterogeneous language disorders: a population study. *J Child Psychol Psychiatry.* (2017) 58:1092–105. doi: 10.1111/jcpp.12793
3. Bishop DV, Snowling MJ, Thompson PA, Greenhalgh T, and CATALISE consortium (2016). CATALISE: a multinational and multidisciplinary Delphi Consensus Study. Identifying language impairments in children. *PloS One* 11:e0158753. doi: 10.1371/journal.pone.0158753
4. Tomblin JB, Records NL, Buckwalter P, Zhang X, Smith E. Prevalence of specific language impairment in kindergarten children. *J Speech Lang Hear Res.* (1997) 40:1245–60. doi: 10.1044/jslhr.4006.1245
5. Leonard LB. Children with specific language impairment and their contribution to the study of language development. *J Child Lang.* (2014) 41:38–47. doi: 10.1017./S0305000914000130
6. Thordardottir E. 12. Proposed diagnostic procedures for use in bilingual and cross-linguistic contexts. In: Armon-Lotem S, de Jong J, Meir N, editors. *Assessing Multilingual Children*. (2015). pp. 331–58. Bristol, Blue Ridge Summit: Multilingual Matters. doi: 10.21832/9781783093137-014
7. Clegg J, Hollis C, Mawhood L. Developmental language disorders—a follow-up in later adult life. Cognitive, language and psychosocial outcomes. *J Child Psychol Psychiatry Allied Discip.* (2005) 46:128–49. doi: 10.1111/j.1469-7610.2004.00342.x
8. Lindsay G, Dockrell JE, Strand S. Longitudinal patterns of behaviour problems in children with specific speech and language difficulties: child and contextual factors. *Brit J Educ Psychol.* (2007) 77:811–28. doi: 10.1348./000709906X171127
9. Catts HW, Fey ME, Tomblin JB. A longitudinal investigation of reading outcomes in children with language impairments. *J Speech Lang Hear Res.* (2002) 45:1142–57. doi: 10.1044/1092-4388(2002/093)
10. Law J, Rush R, Schoon I. Modeling developmental language difficulties from school entry into adulthood: literacy, mental health, and employment outcomes. *J Speech Lang Hear Res.* (2009) 52:1401–16. doi: 10.1044/1092-4388(2009/08-0142)
11. Buschmann A, Jooss B, Rupp A, Feldhusen F, Pietz J. Parent based language intervention for 2-year-old children with specific expressive language delay: a randomised controlled trial. *Arch Dis Child.* (2009) 94:110–6. doi: 10.1136/adc.2008.141572
12. Law J, Garrett Z, Nye C. The efficacy of treatment for children with developmental speech and language delay/disorder: a meta-analysis. *J Speech Lang Hear Res.* (2004). 47:924–43. doi: 10.1044/1092-4388(2004/069)
13. Roberts MY, Kaiser AP. The effectiveness of parent-implemented language interventions: a meta-analysis. *Am J Speech Lang Pathol.* (2011) 20:180–99. doi: 10.1044/1058-0360(2011/10-0055)
14. Roberts MY, Kaiser AP. Early intervention for toddlers with language delays: a randomized controlled trial. *Pediatrics.* (2015) 135:686–93. doi: 10.1542/peds.2014-2134
15. Nelson HD, Nygren P, Walker M. Screening for speech and language delay in preschool children: systematic evidence review for the US Preventive Services Task Force. *Pediatrics.* (2006) 117:e298–319. doi: 10.1542/peds.2005-1467

16. Wallace IF, Berkman ND, Watson LR, Coyne-Beasley T, Wood CT, Cullen K. Screening for speech and language delay in children 5 years old and younger: a systematic review. *Pediatrics.* (2015) 136:e448–62. doi: 10.1542/peds.2014-3889
17. Siu AL, US Preventive Services Task Force. Screening for speech and language delay and disorders in children aged 5 years or younger: US Preventive Services Task Force Recommendation Statement. *Pediatrics* 136:e474–81. doi: 10.1542/peds.2015-1711
18. National Health and Medical Research Council (Australia). *Child Health Screening and Surveillance: A Critical Review of the Evidence*. (2002). Canberra: National Health and Medical Research Council.
19. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. *Früherkennungsuntersuchung auf umschriebene Entwicklungsstörungen des Sprechens und der Sprache: Abschlussbericht*. (2009); Version 1, 0. Nr. 57. Retrieved from: https://www.iqwig.de/download/s06-01_abschlussbericht_frueherkennung_umschriebener_stoerungen_des_sprechens_und_der_sprache.pdf (accessed September 15, 2021).
20. Kasper J, Kreis J, Scheibler F, Möller D, Skipka G, Lange S, et al. Population-based screening of children for specific speech and language impairment in Germany: a systematic review. *Folia Phoniatr Logop*. (2011) 63:247–63. doi: 10.1159/000321000
21. Uilenburg N, Wiefferink K, Verkerk P, van Denderen M, van Schie C. Accuracy of a screening tool for early identification of language impairment. *J Speech Lang Hear Res.* (2018) 61:104–13. doi: 10.1044/2017_JSLHR-L-16-0173
22. Voigt RG, Llorente AM, Jensen CL, Fraley JK, Barbaresi WJ. Comparison of the validity of direct pediatric developmental evaluation versus developmental screening by parent report. *Clin Pediatr.* (2007) 46:523–9. doi: 10.1177/0009922806299100
23. Sturrock A, Marsden A, Adams C. Observational and reported measures of language and pragmatics in young people with autism: a comparison of respondent data and gender profiles. *J Autism Dev Disord.* (2020) 50:812–30. doi: 10.1007/s10803-019-04288-3
24. Ebert KD. Convergence between parent report and direct assessment of language and attention in culturally and linguistically diverse children. *PLoS ONE.* (2017) 12:e0180598. doi: 10.1371/journal.pone.0180598
25. Miller LE, Perkins KA, Dai YG. (2017). Comparison of parent report and direct assessment of child skills in toddlers. *Res Autism Spectr Disord.* (2017) 41–42:57–65. doi: 10.1016/j.rasd.08, 002.
26. Glascoe FP. Evidence-based approach to developmental and behavioural surveillance using parents' concerns. *Child Care Health Dev.* (2000) 26:137–49. doi: 10.1046/j.1365-2000.00173.x
27. Glascoe FP. Parents' evaluation of developmental status: how well do parents' concerns identify children with behavioral and emotional problems? *Clin Pediatr*. (2003) 42:133–8. doi: 10.1177/000992280304200206
28. Glascoe FP, Dworkin PH. The role of parents in the detection of developmental and behavioral problems. *Pediatrics.* (1995) 95:829–36.
29. Hamann C. Specific Language Impairment in German speaking children. In: Editor Stavrakaki S, editors. *Specific Language Impairment: Current trends in research*. (2015). pp. 215–252. Available from: https://benjamins.com/catalog/lald.58, 10ha.m. doi: 10.1075/lald.58.10ham
30. Clahsen H, Richman K, Richman K. *Child Language and Developmental Dysphasia: Linguistic Studies of the Acquisition of German*. Amsterdam: John Benjamins Publishing Company. (1991)
31. Hamann C, Penner Z, Lindner K. German impaired grammar: the clause structure revisited. *Language Acquisition*. (1998) 7:193–245. Retrieved May 27, 2021. Available from: http://www.jstor.org/stable/20000285. doi: 10.1207/s15327817la0702-4_5
32. Rescorla L. Late talkers: do good predictors of outcome exist? *Develop Disabil Res Rev*. (2011) 17:141–50. doi: 10.1002/ddrr.1108
33. Möller D, Furche G, Slabon-Lieberz S, Gaumert G, Breitfuss A. Blickdiagnose Sprachverständnisstörungen–Die diagnostische Güte von Experten-und Elternurteilen. *Sprache· Stimme· Gehör*. (2008) 32:129–35. https://10.1055/s-0028-1085435. doi: 10.1055/s-0028-1085435
34. Fenson L, Marchman VA, Thal DJ, Dale PS, Reznick JS. MacArthur-Bates communicative development inventories. 2nd edition. (2007). Baltimore: Paul H. Brookes Publishing Co. doi: 10.1037./t11538-000
35. Marschik PB, Einspieler C, Garzarolli B. Events at early development: are they associated with early word production and neurodevelopmental abilities at the preschool age? *Early Hum Develop*. (2007) 83, 107–14. doi: 10.1016/j.earlhumdev.2006.05.009
36. Grimm H, Aktas M, Frevert S. *SETK 3-5. Sprachentwicklungstest für drei-bis fünfjährige Kinder. Diagnose von Sprachverarbeitungsfähigkeiten und auditiven Gedächtnisleistungen*. (2001). Göttingen: Hogrefe.
37. Kiese-Himmel C. *AWST-R-Aktiver Wortschatztest für 3-bis 5-jährige Kinder*. Göttingen: Beltz Test. (2005).
38. Tellegen PJ, Laros JA. *SON-R 2 ½ - 7. Nonverbaler Intelligenztest*. (2007). Göttingen: Hogrefe.
39. Bowen DJ, Kreuter M, Spring B, Cofta-Woerpel L, Linnan L, Weiner D, et al. How we design feasibility studies. *Am J Prev Med.* (2009) 36:452–7. doi: 10.1016/j.amepre.2009.02.002
40. Tomblin JB, Records NL, Zhang X. A system for the diagnosis of specific language impairment in kindergarten children. *J Speech Hear Res*. (1996) 39:1284–94. doi: 10.1044/jshr.3906.1284
41. Swets JA. Measuring the accuracy of diagnostic systems. *Science.* (1988) 240:1285–93. doi: 10.1126/science.3287615
42. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics.* (1988) 44:837–45. doi: 10.2307/2531595
43. Pepe MS. *The Statistical Evaluation of Medical Tests for Classification and prediction*. Oxford University Press, New York. (2003).
44. Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. *JAMA.* (1994) 271:703–7. doi: 10.1001/jama.271.9.703
45. Muthén LK, Muthén B. *Mplus User's Guide*. (1998–2017). Los Angeles, CA: Muthén and Muthén.
46. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics.* (2011) 12:77. doi: 10.1186/1471-2105-12-77
47. López-Ratón M, Rodríguez-Álvarez M, Cadarso-Suárez C. Optimal cutpoints: an R package for selecting optimal cutpoints in diagnostic tests. *J Stat Softw*. (2014) 61:1–36. doi: 10.18637/jss.v061.i08
48. Alonzo T, Pepe MS. Assessing accuracy of a continuous screening test in the presence of verification bias. *J Royal Stat Soc: Series C (Appl Stat)*. (2005) 54:173–90. doi: 10.1111/1460-6984.12242
49. Hall MK, Kea B, Wang R. Recognising bias in studies of diagnostic tests part 1: patient selection. *Emerg Med J: EMJ.* (2019) 36:431–4. doi: 10.1136/emermed-2019-208446
50. Cho H, Matthews GJ, Harel O. Confidence intervals for the area under the receiver operating characteristic curve in the presence of ignorable missing data. *Int Stat Rev*. (2019) 87:152–77. doi: 10.1111/insr.12277
51. Enders CK, Keller BT, Levy R. A chained equations imputation approach for multilevel data with categorical and continuous variables. Psychol Methods. (2018) 23:298–317. doi: 10.1037/met0000148
52. Keller BT, Enders CK. *Blimp User's Manual (Version 1, 0.)*. (2017). Los Angeles, CA.
53. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. New York, NY: Wiley. (2004).
54. Madley-Dowd P, Hughes R, Tilling K. (2019). The proportion of missing data should not be used to guide decisions on multiple imputation. *J Clin Epidemiol.* (2019) 110:63–73. doi: 10.1016/j.jclinepi.02,016.
55. Sim F, Thompson L, Marryat L, Ramparsad N. Predictive validity of preschool screening tools for language and behavioural

difficulties: A PRISMA systematic review. *PLoS One.* (2019) 14:e0211409. doi: 10.1371/journal.pone.0211409

56. Glascoe FP. Are overreferrals on developmental screening tests really a problem?. *Arch Pediatr Adolesc Med.* (2001). 155, 54–9. doi: 10.1001/archpedi.155.1.54

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.

frontiers in Pediatrics

ORIGINAL RESEARCH
published: 04 February 2022
doi: 10.3389/fped.2022.826817

Developing Preschool Language Surveillance Models - Cumulative and Clustering Patterns of Early Life Factors in the Early Language in Victoria Study Cohort

Patricia Eadie[1*], *Penny Levickis*[1,2], *Cristina McKean*[3], *Elizabeth Westrupp*[4,5], *Edith L. Bavin*[2,6], *Robert S. Ware*[7], *Bibi Gerner*[1,2] and *Sheena Reilly*[7]

[1] Melbourne Graduate School of Education, University of Melbourne, Melbourne, VIC, Australia, [2] Genetics, Murdoch Children's Research Institute, Melbourne, VIC, Australia, [3] School of Education, Communication and Language Sciences, Newcastle University, Newcastle upon Tyne, United Kingdom, [4] Deakin University, Centre for Social and Early Emotional Development, School of Psychology, Geelong, VIC, Australia, [5] Judith Lumley Centre, La Trobe University, Melbourne, VIC, Australia, [6] School of Psychology and Public Health La Trobe University, Melbourne, VIC, Australia, [7] Menzies Health Institute Queensland, Griffith University, Gold Coast, QLD, Australia

OPEN ACCESS

Edited by:
*Daniel Holzinger,
Hospitaller Brothers of Saint John of God Linz, Austria*

Reviewed by:
*Kate Crowe,
University of Iceland, Iceland
Gisela Szagun,
University of Oldenburg, Germany*

***Correspondence:**
*Patricia Eadie
peadie@unimelb.edu.au*

Specialty section:
This article was submitted to Children and Health, a section of the journal Frontiers in Pediatrics

Received: *01 December 2021*
Accepted: *10 January 2022*
Published: *04 February 2022*

Citation:
Eadie P, Levickis P, McKean C, Westrupp E, Bavin EL, Ware RS, Gerner B and Reilly S (2022) Developing Preschool Language Surveillance Models - Cumulative and Clustering Patterns of Early Life Factors in the Early Language in Victoria Study Cohort. Front. Pediatr. 10:826817. doi: 10.3389/fped.2022.826817

Background: Screening and surveillance of development are integral to ensuring effective early identification and intervention strategies for children with vulnerabilities. However, not all developmental skills have reliable screening processes, such as early language ability.

Method: We describe how a set of early life factors used in a large, prospective community cohort from Australia are associated with language abilities across the preschool years, and determine if either an accumulation of risk factors or a clustering of risk factors provide a feasible approach to surveillance of language development in preschool children.

Results: There were 1,208 children with a 7-year language outcome. The accumulation of early life factors increased the likelihood of children having low language skills at 7-years. Over a third of children with typical language skills (36.6%) had ≤ two risks and half of the children with low language (50%) had six or more risks. As the number of factors increases the risk of having low language at 7-years increases, for example, children with six or more risks had 17 times greater risk, compared to those with ≤ two risks. Data collected from 1,910 children at 8- to 12-months were used in the latent class modeling. Four profile classes (or groups) were identified. The largest group was developmentally enabled with a supportive home learning environment (56.2%, $n = 1,073$). The second group was vulnerable, both developmentally and in their home learning environment (31.2%, $n = 596$); the third group was socially disadvantaged with a vulnerable home learning environment (7.4%, $n = 142$); the final group featured maternal mental health problems and vulnerable child socio-emotional adjustment (5.2%, $n = 99$). Compared to developmentally enabled children, the risk of low language at 7-years was greater for children in the three other groups.

Conclusion: The cumulative and cluster risk analyses demonstrate the potential to use developmental surveillance to identify children within the first years of life who are at risk of language difficulties. Importantly, parent-child interaction and the home learning environment emerged as a consistent cluster. We recommend they be adopted as the common focus for early intervention and universal language promotion programs.

Keywords: language, development, preschool, surveillance, risk factors

INTRODUCTION

Language skills emerge during the first year of life, characterized by periods of swift growth and a relatively consistent sequence of development. Despite these commonalities, a hallmark feature of children's early language development is noticeable individual variability (1). What drives individual child differences in language ability has been a core question of language development research over many decades. Like other developmental domains, language skills are shaped by biological and environmental factors, and while some of the variability in children's language skills can be attributed to both from early in life (2), the ongoing challenge has been to better understand and predict different developmental pathways.

While the earliest years of life provide a crucial window to impact children's developmental pathways (3), building policy for service provision which maximizes learning opportunities in the preschool years requires a comprehensive understanding of a child's development and likely trajectories. Mapping individual developmental pathways involves identification of biological and environmental factors which either serve to buffer a child's development or puts them at risk and in need of preventative or early interventions. Yet ensuring effective early identification processes for some developmental skills has proven elusive. Early language ability is one such example (4) where the current evidence is insufficient to recommend screening in children from birth to 5-years.

Screening and surveillance of development are an integral part of public health approaches which focus on the social and environmental determinants of population health. Public health is defined as "the science of protecting and improving the health of people and their communities" through "detecting, preventing, and responding to disease" (5). The challenge is to ensure the accuracy and reliability of processes for detecting or screening of disease. Understanding the developmental pathways of language ability well enough to inform public health approaches for the management of early language vulnerability and thereby minimize the life-long consequences associated with language disorder, has to date proved impracticable.

This paper draws on almost two decades of data from the Early Language in Victoria Study (ELVS), an Australian longitudinal cohort study, in discussing what we have learnt about predicting language outcomes from early life factors and to question whether novel approaches examining accumulating risks or clustering of risks offer alternative possibilities to the surveillance of preschool language development.

Developmental Surveillance and Screening

Effective early prevention and intervention efforts have potential for significant positive impacts on children's health and development (6). However, detection processes rely on accurate screening tools to identify those vulnerable children who most stand to benefit from the health and education services available. Currently, not all developmental difficulties have adequately sensitive indicators of vulnerability or measurement tools, and consequently universal screening is not recommended.

Although screening, monitoring and surveillance are terms often used interchangeably, the distinction is critical for complex developmental skills, such as language. Developmental screening involves the use of brief questionnaires and/or standardized tools, used at a designated point in time, to assist in the identification of children who are vulnerable to developmental difficulty. The purpose of developmental screening is to identify children at increased risk of a disorder and who need referral for further in-depth diagnostic assessment (7, 8). Universal screening is recommended when there is a tool sufficiently accurate to ensure confidence in identifying children truly at risk from all children in a certain age group of the population regardless of symptomatology (8), for example, with infant hearing screening. Neither universal nor targeted speech and language screening are currently recommended; two systematic reviews conducted for the US Preventative Services Task Force (4, 9), concluded there was inadequate evidence on the accuracy of screening instruments for speech and language delay. Consequently, research efforts need to consider alternatives to screening, examples of which may include identifying early life factors, and the potential for developmental monitoring (or surveillance) to better detect vulnerable children.

Developmental surveillance refers to a broader, flexible and ongoing process of observing children regularly over time, as well as eliciting and attending to parents' concerns (7, 10). Developmental surveillance, therefore, provides a continuous, collaborative and cumulative process to document developmental history and identify children who may be at risk for developmental problems due to highly individualized and contextual factors. It is the ongoing monitoring of a child's developmental progress which sets surveillance apart from screening (8). While developmental surveillance holds the best opportunity to identify children with vulnerable language skills, there remain significant challenges in what we know, and gaps in how to translate knowledge to primary health and clinical settings.

Emergence of Language Skills and Language Screening

Language learning occurs within the context of the social interactions and relationships children have with adults and peers in the environment/s in which they live and learn (11–13). Responsive and reciprocal adult-child interactions are critical to the language learning process (14) which best occurs in the context of nurturing, predictable and contingent early experiences with adults (11).

Large scale, longitudinal studies have examined children's developmental pathways and determined a range of factors associated with good vs. poor language outcomes. Over the last 30 years there has been an accumulation of rich language data from international cohort studies, which track the speech and/or language, academic and social-emotional outcomes of participants from early childhood into adolescence and adulthood (15, 16). All of these cohort studies demonstrate that weaknesses in language learning in the preschool and early school years substantially increase the risks of significant later difficulties in education attainment, employment, mental health and wellbeing (17). They also all point to substantial challenges in the precision of early language skills (usually measured by vocabulary at 2-years), to accurately identify those children who will have persistent language difficulties at school entry. Most cohort studies started to document language skills when children were 4-years or older, collecting earlier communication and language milestones retrospectively (18). Many focused on prevalence within clinical samples (19), with few reporting on the association of early life factors with language difficulties (20, 21). These studies consistently found that being male, having a family history of language difficulties, and early neurobiological risks (e.g., low weight for gestational age) were predictors of language status at 2- and 5-years (20, 22).

It was clear that despite the significant work undertaken by the beginning of the twenty-first century, there were still many unanswered questions regarding early life factors, early communication milestones, and emerging language skills at a population level. Equally, early detection was still being considered from a developmental screening perspective, despite emerging concerns regarding the specificity and sensitivity of screening tools (23, 24). While there were a few brief structured screening tools that showed promise (25), many were less accurate at a 2-year follow up or longer (26, 27). There was also the vexed question of whether early detection of language delay actually resulted in short-term health benefits. de Koning et al.'s (28) cluster-randomized trial conducted in the Netherlands concluded that the large-scale introduction of screening for language disorders in toddlers could not be recommended based on both screening and intervention outcomes.

With an understanding of both the findings and gaps from these studies, in 2002 the Early Language in Victoria Study (ELVS) was established to examine the natural history of language development and language difficulties from infancy. Our purpose was to inform policy and practice regarding the promotion of language and communication development, prevention strategies and intervention for young children at risk for language difficulties. In beginning to track children's development in the first year of life in a community representative sample, our aim was to address some of the limitations in the literature at the time.

The Early Language in Victoria Study

The Early Language in Victoria Study (ELVS), commenced in 2002 with a focus on epidemiology and language skills. ELVS has a prospective, longitudinal cohort design. Our approach was to collect comprehensive information *via* survey from multiple informants that was inclusive of many development domains, as well as through direct assessment of children repeated at several salient ages using "gold standard" measures (29).

ELVS adopted Bronfenbrenner's ecological model to frame and describe the dynamic interactions between child, family, community and broad social-economic and cultural contexts that influence children's learning and development (30). Factors within each of these areas can be either protective, buffering development, or exposing children to risks that can leave them developmentally vulnerable. Consequently, a set of early life risk and protective factors, with proven associations with language outcomes, were derived from the literature and systematic reviews (4, 9) that focused on the child, the family environment, and the primary caregiver (mother).

Background

The overall aims of ELVS were to: (i) describe the natural history and clinical course of childhood language disorders; (ii) determine the extent to which language trajectories are fluid, and identify developmental pathways to good vs. low language; (iii) identify which environmental, social and family factors predict variation in these language pathways; and (iv) examine how language pathways are associated with children's social, behavioral and educational outcomes (29). Our objective was to build clinically applicable evidence for the best age at which to accurately identify children who are likely to experience persistent language difficulties. ELVS has followed participants from infancy through to adolescence (13-years). Its current phase is collecting data as participants exit formal schooling (18–19-years). The first phase of ELVS focused on the emergence of language skills up to 4-years, with a second phase extending the research to 7-years.

ELVS analyses draw on data in which language was repeatedly assessed in the preschool and early school years, *via* parent report using the Communication and Symbolic Behavior Scales: Infant-toddler Checklist (CSBS:ITC) (31) and the Macarthur Bates Communicative Development Inventory (CDI) (32) and direct assessment using the Clinical Evaluation of Language Fundamentals (CELF) Australian Adaptations, the Preschool 2nd Edition (33) and Fourth Edition (34). Here, we provide a summary of major themes from the findings when the children were aged between 8-months and 11-years.

Most Variation in Language Outcomes Is Unexplained Using Early Life Factors

At 12-months, 2-, 4-, 7, and 11-years (2, 35–38) language outcomes were predicted based on the 12 child, family and maternal factors, and earlier communication and language skills. Our aim was to identify factors that would contribute at key ages to the identification of children with vulnerable language skills. While many of the early life factors remained associated with language outcomes across time, at no point did they provide enough accuracy to identify children with either vulnerable language skills or language difficulties.

A major finding from the 12-month and 2-year analyses was that only a small amount of the total variation in communication and expressive vocabulary scores at each age was explained by the set of 12 risk variables: at 12-months <6.0%, and at 2-years 4.3% (37, 38). Similarly, for vocabulary (number of words produced), the amount of variation contributed by the risk factors was small (7.0% at 2-years). Earlier communication scores made the major contribution to the variance in 2-year vocabulary (38). Notwithstanding this, some of the factors had a significant association with children's communication and vocabulary scores at one or more of the time-points. **Table 1** summarizes the significant early life factors at the different ages.

More variation in language outcomes was explained by the 12 factors at 4-years than at the earlier ages. For receptive language, the 12 factors together explained 18.9% of the variation, and for expressive language, 20.9%. When we included a measure of Late Talking status at 2-years (based on the 10th percentile cut point for vocabulary), the variance explained increased to 23.6% for receptive language and 30.4% for expressive. Nine of the 12 risk factors were significantly associated with the language scores (see **Table 1**). Of interest in these 4-year predictive models was the shift from predominantly child factors predicting early communication and vocabulary at 12-months and 2-years, to mainly family and maternal factors significantly predicting language at 4-years. We concluded that the biological (child) factors drive the earliest development (e.g., male sex and birth order), similar to studies that were interested in predicting late talking at 2-years (22). Our 4-year outcomes suggested that the impact of social and environmental factors may take longer to accumulate but are detectable by 4-years. These factors included socio-economic status, family history of language difficulties, and non-English speaking background.

As children progressed through school the same set of potential predictors explained less of the overall variance in language, 9–13% for the receptive and expressive language scores at 7-years and 11–12% at 11-years. Not surprisingly, children's earlier language skills made a greater contribution than the early life factors, low language at 7-years was more accurately predicted by the 4-year language scores, and 11-year language outcomes were predicted more reliably when 7-year language scores were added to regression models, with the variance explained up to 47 and 64%, respectively.

Early Variability in Children's Language Profiles

In reporting regularly and systematically on the emergence of children's language skills, findings from ELVS have demonstrated, that despite common assumptions, developmental trajectories fluctuate considerably in the preschool years. This is the case for children with early typical development, as well as for those with early vulnerabilities. In the ELVS cohort, less than half of the children identified at 4-years with language difficulties were identified at 2-years as late talkers (39). These figures are remarkably similar to those from other international cohorts (40, 41). Furthermore, 6% of children who had typical skills at 2-years had language difficulties by 4-years. ELVS analyses also demonstrated that the stability in language classification was low between 4- and 5-years, with 36% of 5-year-olds with low language scores classified as typical at 4-years (42). The variability observed consistently across studies in developmental language pathways to 5-years is the result of a combination of fluctuations in children's abilities, the changing nature of the language skills measured at different ages (e.g., gestures and vocabulary, semantics and grammar), limitations in measurement instruments, and the arbitrary nature of the boundaries defining language difficulties.

Developmental Profiles or Sub-groups of Language Trajectories

Latent class modeling has been used to identify sub-groups of developmental trajectories for children within the ELVS cohort using data from 8-months to 4-years and from 4- to 11-years. Developmental profiles were derived from early communication and language measures at 8-months through to direct assessment at 4-years. Five developmental profiles were identified (43): (i) typical group had age-expected language scores at each age (68.5%); (ii) precocious (late) group showed typical development initially but precocity in development from 24-months on (15.0%); (iii) impaired (early) group had high probability of impairment up to 12-months and then typical language development (6.1%); (iv) impaired (late) group, showed early typical development but delay from 2-years (6.1%); and (v) precocious (early) group showed early precocity and typical language by 4-years (4.3%). From these five profiles it was evident that there was considerable variability in the early developmental trajectories in the ELVS cohort. In addition, those profiles in which improvement was shown were more likely to be associated with higher maternal education and vocabulary and less disadvantage, supporting the view that environmental factors have continued impact through the preschool years as language continues to develop. Importantly, the predictors of the developmental profiles pointed to the importance of language enrichment initiatives for more vulnerable children.

Using data from 4- to 11-years, three language trajectory groups were identified (44): (i) a stable group that comprised 94.0% of participants; (ii) a low-decreasing group which included 4.0% of the cohort; and (iii) a low-improving group which included 2.0% of the cohort. The stability in these language trajectories provided confidence for services identifying children with low language at 4-years that they were likely to remain low to 11 years. Further analysis revealed that the low-decreasing group was associated with mainly biological risks, while the low-improving group was associated with mostly environmental risks. In summary, we had demonstrated that by 4-years trajectories stabilize but that prior to that ongoing monitoring rather than discrete screening points was likely to be the best approach to early identification.

TABLE 1 | Summary of significant early risk factors included in regression analyses of language outcomes at ages 2-, 4-, 7-, and 11-years.

Predictor	2-years (CDI)	4-years (CELF-P2)		7-years (CELF-4)		11-years (CELF-4)	
	Expressive (N = 1,570)[a]	Receptive (N = 1,473)[a]	Expressive (N = 1,442)[a]	Receptive (N = 1,132)[a]	Expressive (N = 1,132)[a]	Receptive (N = 839)[a]	Expressive (N = 839)[a]
	P	***P***	***P***	***P***	***P***	***P***	***P***
Child							
Male sex	✓	✓	✓	✓	✓		
Birth weight (per kg)		✓	✓				
Twin birth					✓		✓
Preterm birth							
Birth order	✓	✓	✓	✓	✓	✓	✓
Family							
Non-English- speaking background	✓	✓	✓				
Socioeconomic disadvantage (SEIFA score)		✓	✓	✓	✓		
Family history of speech-language difficulties	✓	✓	✓	✓	✓	✓	✓
Mother							
Maternal education		✓	✓	✓	✓	✓	✓
Maternal mental health							
Maternal vocabulary		✓	✓	✓	✓	✓	✓
Maternal age		✓	✓				✓

SEIFA, Socio-Economic Indexes for Areas.
[a]Children with complete predictor and outcome data.

The Potential of Family and Parent Factors Measured in the First 12-Months

Findings from ELVS across the first 4-years demonstrated that all three sets of early life factors had an impact on language outcomes, but none were as explanatory as vocabulary measured at 2-years, on 4-year language ability. While this finding was not entirely surprising, it did raise the question of whether there was any information collected at 12-months that could have identified children as accurately and earlier than vocabulary at 2-years. A set of risk factors measured at 12-months, derived from the literature and broadly representing child, family and parenting characteristics (45) were used to predict language difficulties at 4-years. The set comprised three child factors: whether the child had started showing objects to adults; the number of words/phrases understood; and the number of words used meaningfully; three family factors: whether there was a family history of speech, language or communication difficulties; maternal education; socio-economic status (SES) quintile; and one factor related to parental communicative behavior. Using these items, measured at 12-months, the model distinguished children with and without language difficulties at 4-years with acceptable discrimination (AUC of 0.73). Whilst by no means diagnostic, this model was substantially better than late talker status at 2-years and had the added advantage of potentially providing an additional 12-month window in which to provide preventative interventions. Given this evidence, we continued to explore these additional family and parental behavior factors.

Predictors of Language Growth Over Time

In an effort to better understand the focus and timing of prevention and intervention strategies, ELVS data was used to examine the individual differences in children's language growth over time and identify the factors that best predicted this growth. Twenty-two variables included in the analyses were related to early life factors: child, family and environmental variables; as well as parent reported items from the 4-year parent questionnaire and child assessment, and information collected at other waves (46). The variables were classified into three groups based on whether they were (a) least mutable (not changeable through intervention—for a number of reasons); (b) mutable-distal (could be changed but at a population level through social policy); and (c) mutable-proximal (potential to be modified by direct family or child interventions and with strong evidence that by modifying them a positive impact on children's language can be made). The 22 predictors explained 67.0% of the variability in rate of language growth between 4- and 7-years, with 23.0% contributed by mutable proximal factors, including number of books in the home at 2-years, frequency of shared book reading from 8-months to 4-years, TV viewing at 4-years, and pro-social behavior scores at 4-years. Importantly, not only did the trajectories to 7-years indicate a continued influence of the home environment on children's language development but these mutable proximal factors could be modified through intervention.

Summary

After two decades of work with the ELVS cohort we have investigated early life factors, family and environmental characteristics, and expanded factors of interest to parenting communication and behaviors in our efforts to identify those factors that significantly explain later language growth and outcomes at 4- and 7-years. In addition, developmental language profiles to 4-years and language growth trajectories between 4- and 7-years have provided some insights as to who and which factors might be the targets of intervention. It is clear that if developmental surveillance is to successfully identify young children with vulnerable language skills, then we need to identify alternative approaches that consider both cumulative risk and clusters of risk factors. It is our hypothesis based on ELVS data and recent research on developmental vulnerability (47) and language and reading difficulties (48) that developmental surveillance holds the most promise for detecting risk of language difficulties using this cumulative and cluster risk approach.

Cumulative and Cluster Risk Frameworks Applied to Language Development and Disorder

Any framework to document the most meaningful factors related to language skills needs to be broad and capture those genetic and environmental aspects of the child, their family, and community that impact development. In considering cumulative risk, it is possible to extend our understanding of the complex interactions of biological and environmental factors through the lens of Pennington's multiple deficit model (49). This model has been applied to a range of developmental disorders, including dyslexia (50) and language and reading disorders (49), providing a framework to describe the multifactorial etiologies, while accounting for the accumulation of risk and protective factors determined by multiple and concurrent influences on development. The model assumes that cumulative risk increases the likelihood of emerging developmental difficulties, in a probabilistic rather than deterministic approach. This is compatible with current knowledge on the drivers of language vulnerability.

Furthermore, Shonkoff's bio-developmental framework is complementary and valuable in understanding clustering of risks. This framework is structured into three domains which, across the lifespan, capture the: (a) interactions among foundations of healthy development and sources of early adversity, (b) measures of physiological adaptation and disruption, and (c) both positive and negative outcomes in learning, behavior, and health. Focusing on the "interaction" domain enables us to capture the gene-environment interactions that shape early brain architecture and subsequent outcomes in cognitive, language and social-emotional skills. The interaction domain provides a way of conceptualizing the likely factors that will cluster together to increase language vulnerability and potentially be the levers in successful intervention.

The application of cumulative and cluster risk frameworks has only rarely been used to explain early language vulnerability (48, 51, 52) and few if any studies have had prospective data from the first years of life to determine what factors may be meaningful for developmental surveillance, build a cumulative risk index and determine threshold/s or trigger points at which early intervention is recommended. Moreover, identifying the dynamic and complex interactions early enough that impact language development (i.e., the early life factors which cluster together), can inform the investments and likely targets in interventions for children who are at greatest risk.

Here we want to determine the utility of developmental surveillance, as a way of identifying children at-risk for language difficulties. Using a cumulative risk approach, we wanted to inform the processes of monitoring language development over the first years of life. Moreover, identifying a cluster/s of influential early life factors, we aimed to provide guidance for more customized early interventions.

The Current Study

The analyses reported here aim to describe how a set of early life factors defined initially in ELVS (9) and added to using a bio-developmental framework (53) are associated with language abilities across the preschool years, and to determine if either an accumulation of risk factors or a clustering of risk factors provide a feasible approach to surveillance of language development in preschool children.

This paper draws on data from ELVS to:

i) Describe the impact of accumulation of a broad set of early life factors up to 7-years of life on language outcomes; that is, does language ability vary based on the number of risks children are exposed to, and does cumulative risk (i.e., an increasing number of factors) improve the accuracy of predicting outcomes?
ii) Investigate clusters of a broad set of early life factors, through latent class analysis, and determine if the latent classes contribute to the accuracy of prediction of language outcomes at 7-years.

MATERIALS AND METHODS

The Early Language in Victoria Study (ELVS) commenced in 2002. A community sample of 1,910 infants aged 7.5- to 10-months was recruited between September 2003 and April 2004 from 6 of 31 local government areas (LGAs) in metropolitan Melbourne, Victoria, Australia. The LGAs were selected to represent high, medium, and low SES according to the Australian census-based Socio-Economic Indexes for Areas (SEIFA) Index for Relative Socio-Economic Disadvantage (54).

Infants were recruited through the Maternal and Child Health Service, with supplementary recruitment *via* universally available hearing screening sessions and local newspaper advertising. Infants with developmental delay (e.g., Down syndrome), cerebral palsy, or other serious intellectual or physical disability were excluded, as were parents unable to speak and/or understand English sufficiently to respond to the questionnaires. Further sampling methods and study protocols are reported elsewhere (29). **Figure 1** shows participant retention and attrition across the first 8 waves of the study to 7-years. The in-scope sample for this study comprised the 1,208 children from the ELVS cohort who completed direct language assessment at 7-years. This study was approved by the ethics committees of the Royal Children's Hospital (Melbourne) (#23018 and #27078) and La Trobe University (#03-32), and all parents provided written, informed consent.

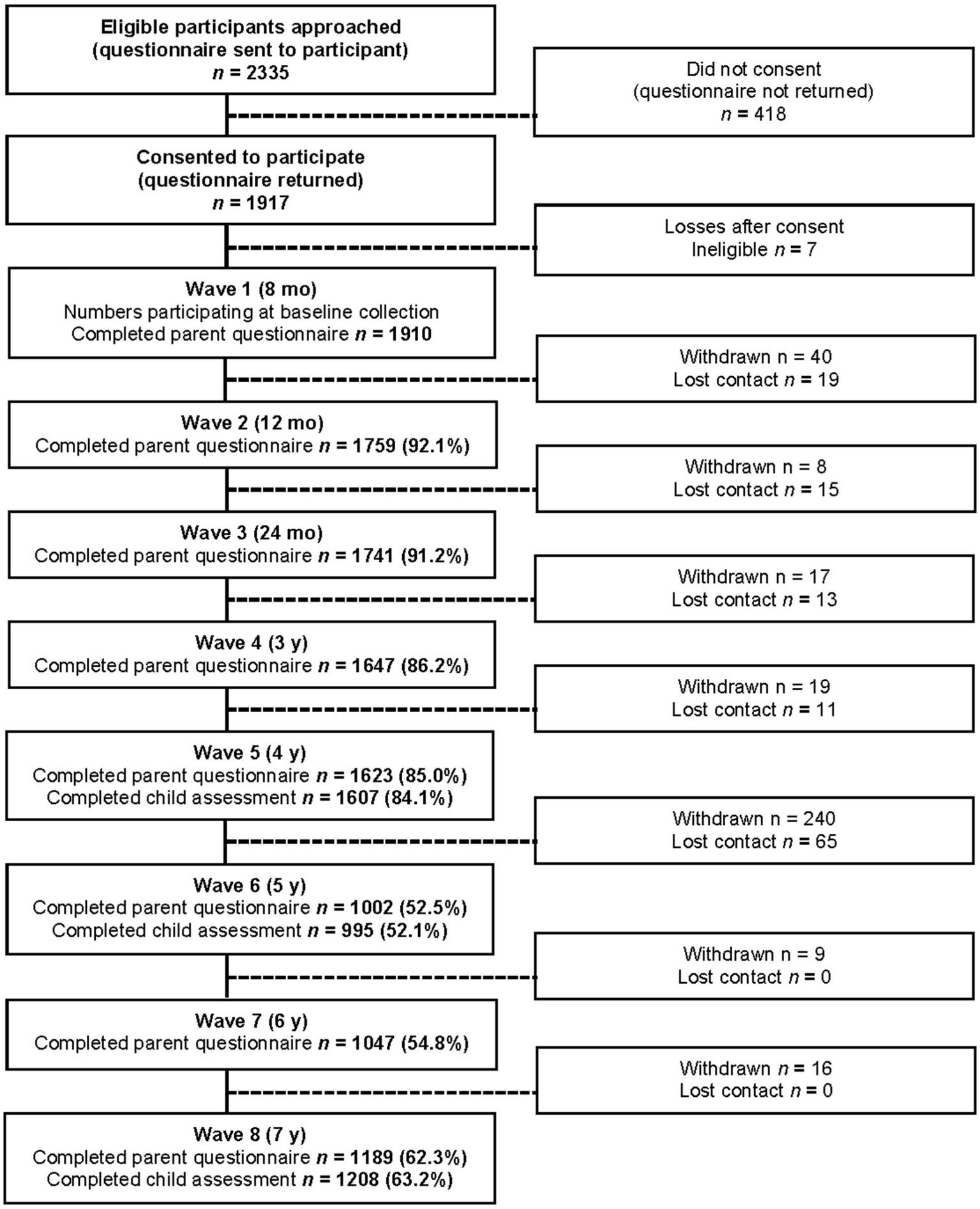


FIGURE 1 | Participant flowchart from baseline (8-months) to 7-years (denominator for percentages is number participating at baseline [$N = 1{,}910$]). Numbers vary at each wave because of participants' withdrawing, losing contact, or not participating in a particular wave but returning at a later stage.

Parent questionnaire data was collected annually from 1 to 7-years, with parents sent questionnaires within a month of their child's birthday. Face-to-face assessments occurred when children were 4, 5 and 7-years. The face-to-face assessments were administered individually to each child by an experienced trained researcher, usually in a single sitting at the child's local health center, school or home. For these analyses data were drawn from the first eight waves of ELVS questionnaire data and the 4- and 7-year assessments.

Measures

Early Life Factors

A comprehensive set of early life factors included a combination of child (birth weight, non-verbal cognition), family (history of speech/language difficulties, socio-economic disadvantage), maternal (mental health, responsivity) and environmental (home learning environment) characteristics. **Table 3** provides a description of the 16 early life factors and when they were collected. Further details are provided in the following text.

Parents completed study generated questions measuring their child's general health and development, birthweight, family history of speech, language and literacy problems, highest level of parental education, and the main language spoken in the home to the child. Families who reported a main language other than English spoken to the child at home were classified as Non-English-Speaking Background (NESB). Family history of speech and language difficulties was reported at 12-months and coded as positive if the child's father, mother or siblings was reported to have either "been late to talk," "had ongoing problems with speech or language during childhood," "had problems with stuttering," or "had problems learning to read."

SES was measured using the Australian census-based SEIFA Index of Relative Socio-Economic Disadvantage at the local government area level (mean $\pm$ SD of 1,000 $\pm$ 100) (54), with lower scores representing greater disadvantage compared with other geographic areas. Maternal mental health was determined by using the Kessler Non-specific Psychological Distress Scale (K-6) (55). Scores were defined as below 4 ("no mental health problem") and 4 to 24 ("likely mental health problem"). Maternal vocabulary was measured with the written multiple-choice modified version of the Mill Hill Vocabulary Scale (56). Each correct answer is tallied to provide a raw score with a possible maximum of 44, with high scores indicating better vocabulary.

The home learning environment was captured at 2-years by asking parents to report the number of books in the home. Having more than 30 books at home has been found to be an important indicator of child literacy practices at home (57). A study using data from the Longitudinal Study of Australian Children (LSAC) has shown a significant association between the number of children's books available at home and children's reading and numeracy performance (dichotomized variable as 0–30 books and more than 30 books in the home) (58).

Ten items from the Brigance Parent-Child Interactions Scale (BPCIS) (59), shown to predict language outcomes in infants and toddlers, were included in the ELVS parent questionnaire at 12-months, 2- and 3-years. The BPCIS is an 18-item parent-reported measure of parenting behaviors and parents' perceptions about their child, drawn from relevant literature (60). The BPCIS items capture parental responsiveness and responding contingently to a child's needs and interests. The BPCIS total raw scores at each time point were dichotomized to create a high vs. low BPCIS variable. A final categorial variable using the dichotomized BPCIS at 12-months, 2- and 3-years was generated. These groups were categorized as: high responsive parental behaviors consistency score (at or above the median BPCIS total score at three time points); inconsistent responsive parental behaviors score (at or above the median BPCIS total score at one or two time points); and low responsive parental behaviors score (below the median BPCIS total score at three times points).

Child temperament at 12-months was measured using parental ratings on the Approach/Withdrawal scale of the Australian normed Short Infant and Toddler Temperament Questionnaires (61). The Approach/Withdrawal scale produces a total score, with high scores indicating high levels of shyness or low sociability. Social, emotional and behavioral difficulties and prosocial behavior were measured at 4-years using the parent-reported Strengths and Difficulties Questionnaire (SDQ) (62). The SDQ produces a Total Difficulties score (possible range 0–40) and Prosocial Behavior score (possible range 0–10). Prosocial behavior is a protective factor for children with DLD (63), so both scores were included as variables of interest. At 4-years non-verbal IQ was measured by the matrices subtest of the Kaufman Brief Intelligence Test, Second Edition (K-BIT2) (64). The average range for K-BIT2 scores was defined as values not more than 1.25 SDs below the ELVS cohort mean; the internal cutoff point was used because the US normative sample included only 100 children at 4-years. Trained research assistants assessed children in their local child health center or in children's homes.

Early Communication and Language Measures

Language abilities were measured using a combination of parent report instruments from 8-months and standardized assessment at 4-and 7-years. At 12-months and 2-years, parents completed the CSBS I-TC (31). This provided a standardized total score (normative mean = 100, SD = 15) and three composite scores for the domains of social, speech, and symbolic skills. The composite domains broadly relate to infants' prelinguistic, linguistic, and cognitive abilities, respectively, each of which has been demonstrated to relate to later expressive language development (65).

Children's gestures were measured at 12-months using the parent-reported CDI Words and Gestures, and included three gesture components: First Communicative Gestures, Games and Routines and Actions with Objects (32). The first two components make up "early gestures," while the third component is considered "later gestures." At 2-years vocabulary was measured by the Words and Sentences version of the CDI for infants. Only the expressive vocabulary production percentile was used in this study. Permission was obtained from the authors to substitute 24 vocabulary items to accommodate Australian usage (e.g., "footpath" instead of "sidewalk").

At 4- and 7-years language was assessed individually by trained research assistants. At 4-years the Australian adaptation of the Clinical Evaluation of Language Fundamentals-Preschool, Second Edition (CELF-P2) (33) was administered and at 7-years children completed the Clinical Evaluation of Language Fundamentals Fourth Edition (CELF-4) Australian Standardization (34). All subtests of the CELF-P2 and 4 were completed. Both the CELF-P2 and CELF-4 composite scores are standardized with a mean of 100, and a standard deviation of 15. Low language outcome was based on a cut point of >1.25 SD below the mean on the CELF-4 Core Language standard score. This cut point has been used in previous ELVS analyses and is in line with other population-based studies in the literature.

Analysis Plan

Cumulative Risk Analysis

For this analysis, cut-off criteria had to be determined for all continuous risk factor variables. As the purpose of dichotomizing the risk factor variables was to describe the impact of cumulative risk factors across the first 7-years of life on language outcomes, we used a more generous cut point of the 20th percentile for continuous variables, rather than clinical cut points. As we wanted to determine whether clusters of risk increase the accuracy of predicting later language ability, we used this cut point to ensure that those children at risk of later language problems would be included. Our approach for managing missing data was a complete case analysis of the data set.

As 16 risk factors were included in our analyses and the proportions of children with some were quite small, we used a grouping strategy reported in Hayiou-Thomas et al. (48) to determine risk categories. This required the smallest risk category was at least as large as the percentage of participants observed with a low language outcome at 7-years (i.e., 10.5%). To examine the association between the cumulative risk categories and low language at 7-years, binomial regression was completed to produce risk ratios (i.e., the risk of low language for children with 3 or more risk factors compared with children with <2 risk factors as the reference group).

Latent Class Analysis

We used latent class analysis in Mplus version 8.3 (66) to identify unique subgroups of participants based on a broad set of early life factors. Where possible, we included continuous variables to maximum variation in the model, and otherwise included as binary variables. For consistency, we recoded continuous variables so that high scores equated to higher risk. To identify the optimal number of latent profiles, we began with a two-profile model and added one profile at a time. We selected the optimal number of profiles based on three criteria; first, visual examination of elbow plots of the Akaike Information Criteria (AIC), Bayesian Information Criteria (BIC), and sample-adjusted BIC (SABIC) (67, 68); second, we considered results for the Bootstrapped likelihood ratio test and Lo-Mendell-Rubin Adjusted test (lower p-values preferred) (69); and third, we used Nylund et al.'s (70) criterion that the posterior probabilities should be >0.70 as evidence that an individual belongs to their assigned profile and no other. We then ran multiple regression analyses to examine whether the latent classes (with the most advantaged class in regard to early life factors as the reference category) predicted language outcomes at 7-years.

RESULTS

Of the original ELVS cohort of 1,910 participants, 1,208 (63.2%) completed a direct assessment at 7-years; they are considered the in-scope sample for the analyses reported here. In the cumulative risk analyses, 966 (50.6%) participants had complete data across the 16 early life factors. All 1,910 original ELVS participants were included in the latent class model, however, we predicted risk ratios for the different classes in the model based on the 1,208 who had completed a 7-year language assessment.

The characteristics of the in-scope participant ($n = 1{,}208$) and non-participant ($n = 702$) groups are presented in **Table 2** which illustrates that attrition resulted in a shift in the 7-year cohort characteristics to the original sample. At 7-years, participants were more likely to be English speaking, live in areas of comparatively less disadvantage, and have mothers who were more educated than non-participants and who had higher vocabulary scores.

TABLE 2 | Characteristics at baseline of participants and non-participants at 7-years.

Characteristics[a]	Non-participants at 7-years ($n = 702$)	Participants at 7-years ($n = 1{,}208$)	*P*
Child			
Female, %	329 (46.9)	616 (51.0)	0.08
Birth weight (kg), mean ± SD	3.40 (0.6)	3.45 (0.5)	0.04
Twin birth, %	25 (3.6)	28 (2.3)	0.11
Preterm birth, %	20 (2.9)	39 (3.2)	0.64
Birth order, %			0.08
First	340 (48.9)	613 (50.8)	
Second	265 (38.1)	407 (33.7)	
Third	69 (9.9)	157 (13.0)	
Fourth or later	21 (3.0)	30 (2.5)	
Family			
Non-English-speaking background, %	80 (11.4)	46 (3.8)	<0.001
Socioeconomic disadvantage (SEIFA score), mean ± SD	1028.02 (67.3)	1040.71 (56.1)	<0.001
Family history of speech-language difficulties, %	181 (25.8)	294 (24.3)	0.48
Mother			
Maternal education, % years of completed schooling			
≤12 y reference	196 (28.3)	247 (20.7)	<0.001
13 y	285 (41.2)	472 (39.5)	
Degree/postgraduate	211 (30.5)	475 (39.8)	
Maternal mental health, %	178 (30.2)	377 (32.4)	0.35
Maternal vocabulary mean ± SD	26.17 (5.6)	28.19 (4.7)	<0.001
Maternal age mean ± SD	30.94 (4.7)	31.85 (4.4)	<0.001

P-values were derived through comparisons between those completing 7-year-old assessments and those lost to follow-up by using either χ^2-tests for categorical variables or t-tests for continuous variables.

[a]*Baseline represents data collected at 8–10 and 12-months.*

Cumulative Early Life Factors for Language Development and Difficulties Across the First 7-Years

In the cumulative risk analysis, the proportion of participants presenting with low language abilities was 8.7% ($n = 84$). The CELF core language score of the typical group ($n = 882$) was 100.9, with a standard deviation of 10.1 and a range of 82–134. In contrast, the group with low language had a CELF core score of 72.8 (SD 9.1) and a range from 40 to 81.

Table 3 presents the proportion of the total number of children ($n = 966$) with an individual risk factor, as well as the number of children with only that particular risk factor (e.g., 3.7% of the sample are male but have no other identified risk factor). Other than male sex, the frequency of children with only one individual risk factor was small, ranging from 0 to 1.2%.

TABLE 3 | Occurrence of identified risk factors in the overall sample ($N = 966$)[a].

Risk factor	Age of measure	Dichotomized variable	Children with this risk factor, *n* (%)	Children with only this risk factor, *n* (%)
Child				
Sex	8-months	Males at increased risk compared with females	476 (49.3)	36 (3.7)
Birth weight	8-months	Coded as low birth weight if <2,500 g	39 (4.0)	5 (0.5)
Temperament	12-months	Coded as high shyness/low sociability if in the top 20th percentile	177 (18.3)	6 (0.6)
Emotional and behavioral development	4-years	Coded as emotional and behavioral difficulties if in the top 20th percentile	165 (17.1)	4 (0.4)
Prosocial behavior	4-years	Coded as poor prosocial skills if in the bottom 20th percentile	283 (29.3)	9 (0.9)
Gestures	12-months	Low gestures, bottom 20th percentile	272 (28.2)	7 (0.7)
Non-verbal cognition	4-years	Low non-verbal IQ as standard score, bottom 20th percentile	180 (18.6)	3 (0.3)
Early vocabulary skills	2-years	Coded as low expressive vocabulary if in the bottom 20th percentile	256 (26.5)	10 (1.0)
Family				
Non-English- speaking background	4-years	Coded as NESB if main language spoken to child is not English	16 (1.7)	0 (0.0)
Socioeconomic disadvantage (SEIFA score)	8-months	Coded as more disadvantaged if in the bottom 20th percentile	169 (17.5)	2 (0.2)
Family history of speech-language difficulties	8-months	Coded 'yes' if the child's father, mother or siblings was reported to have either "been late to talk", "had ongoing problems with speech or language during childhood", "had problems with stuttering", or "had problems learning to read"	236 (24.4)	12 (1.2)
Parent-child interaction	12-months, 2-years, 3-years	Coded as low parent responsivity if low on parent-child interaction measure at all three time points	227 (23.5)	4 (0.4)
Mother				
Maternal education	8-months	Coded as low maternal education if did not completed year 12	192 (19.9)	9 (0.9)
Maternal mental health	8-months	Coded as "likely to have mental health problem" for those scoring 4 to 24	310 (32.1)	8 (0.8)
Maternal vocabulary	12-months	Coded as low vocabulary if in the bottom 20th percentile	128 (13.3)	4 (0.4)
Number of books in the home (home learning environment)	2-years	Coded as negative if parent reported 0–30 children's books in the home	292 (30.2)	12 (1.2)

[a]*Sample size includes no missing data on risk factor variables and outcome data at age 7.*

Factors related to the child (e.g., early communication and vocabulary skills, children's prosocial capacity), family (e.g., number of books in the home, parent-child interaction), and mother (e.g., maternal education and vocabulary) were represented in the factors with the highest proportions. Early risk factors related to the mother (e.g., maternal mental health) and family (e.g., history of speech-language difficulties) were also represented in factors with the highest proportions.

The proportion of children with low language presenting with each risk factor and the ranking of factors is presented in **Figure 2**. A similar set of factors related to the child (e.g., early communication and vocabulary skills), family (e.g., number of books in the home parent-child interaction), and mother (e.g., maternal education and mental health) were present in more than a third of children with low language.

Data is presented for five cumulative risk categories that were created based on the strategy described earlier (48), with the smallest risk category created representing 14% of the total sample (i.e., five risks). The cumulative risk categories represent the proportion of children with less than or equal to two risks; three; four; five; and six or more risks. Over a third of children with typical language skills (36.6%) were in the $\leq$ two risks category and half of the children with low language (50%) had six or more risks. Further detail for the other cumulative risk categories is provided in **Table 4**, along with the risk ratios associated with each category. The results demonstrate that the risk of having low language at 7-years is 17 times more likely for those children with six or more risk factors compared to those with $\leq$ two risks. As the number of risk factors increased from $\leq$ two, the risk of having low language at 7-years increased, for example children with four or five risk factors were 5 and 7 times more likely to have low language compared to those with ≤ 2 risk factors.

Latent Class Modeling—4 Groups

Latent class modeling identified four profile groups. **Figure 3** displays the selection process, with information criteria visualized and test results presented for 2–6 profiles, showing

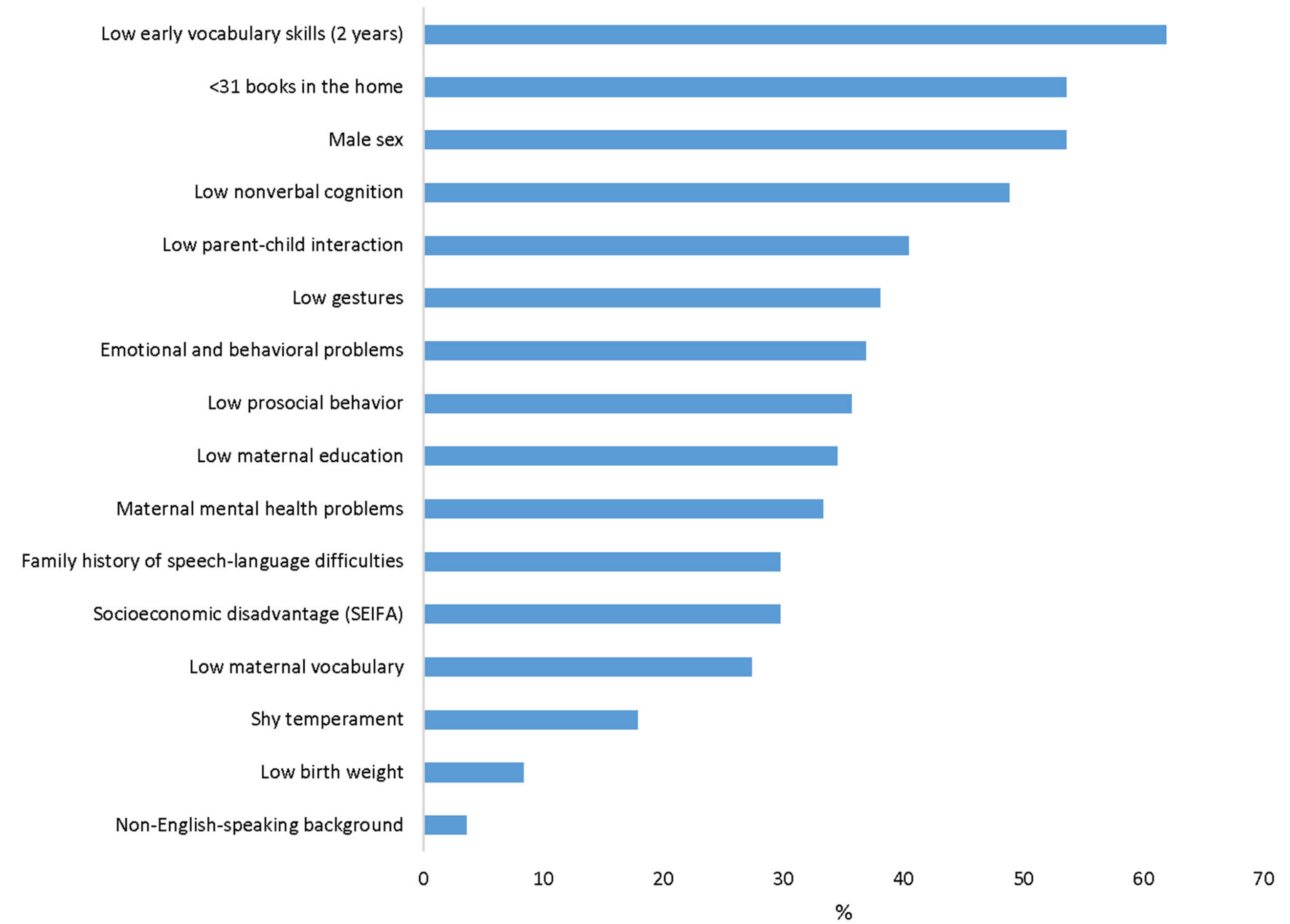


FIGURE 2 | Proportion of children with low language at age 7 by risk factor.

TABLE 4 | Association between number of early life factors and low language outcome at 7-years.

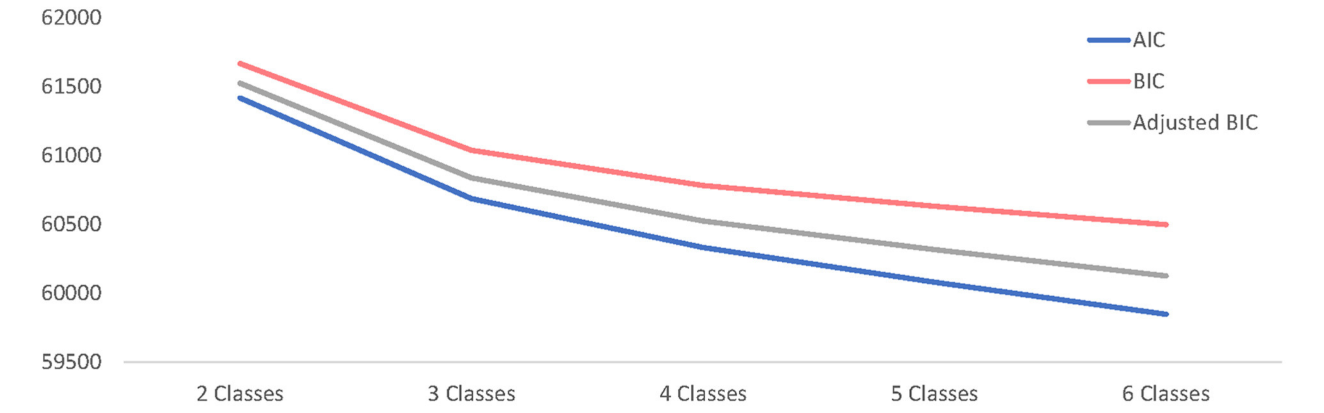
	Total *N*	Low language *N* (%)	Risk ratio (95% CI), *p*
0-2 risks	328	5 (1.5)	1.00 (reference)
3 risks	180	9 (5.0)	3.28 (1.12–9.64), 0.03
4 risks	162	13 (8.0)	5.26 (1.91–14.51), 0.001
5 risks	135	15 (12.5)	7.29 (2.70–19.66), <0.001
6 or more	161	42 (26.1)	17.11 (6.90–42.42), <0.001
Total N	966	84 (8.7)	

an "elbow" at profiles three and four. The four profile-solution was chosen for subsequent analyses because of the large entropy and results from the Bootstrapped likelihood ratio and Lo-Mendell-Rubin Adjusted tests.

The proportions and mean values of the early life factors for the four classes are displayed in **Figure 4**, and the distribution of risk factors across the four profiles is shown in **Table 5**. The first and largest class is developmentally enabled with a supportive home learning environment (56.2%, $n = 1{,}073$). In comparison to the developmentally enabled group (class 1), the three other groups consistently included more males, low scores for parent-child interactions and low maternal education. In addition, class 2 and 3 both had fewer books in the home and lower language scores at 4-years.

The second class is described as vulnerable, both developmentally and in their home learning environment (31.2%, $n = 596$); the third class is described as socially disadvantaged, with a vulnerable home learning environment (7.4%, $n = 142$); the final class features maternal mental health problems and vulnerable child socio-emotional adjustment (5.2%, $n = 99$). Each class had a unique and defining feature: for the vulnerable group (class 2) children presented with low use of early gestures, vocabulary and non-verbal cognition; class 3 was set apart from the other groups due to greater socio-economic disadvantage; while class 4 was the only group where more mothers presented with mental health problems.

To investigate whether the latent classes could predict language outcomes at 7-years we conducted binomial regression



Classes	AIC	BIC	Adjusted BIC	Entropy	Vuong-Lo-Mendell-Rubin	Lo-Mendell-Rubin Adjusted	Bootstrap LRT
2 Classes	61414.968	61664.937	61521.972	0.684	0.0000	0.0000	0.0000
3 Classes	60685.443	61035.4	60835.248	0.781	0.0000	0.0000	0.0000
4 Classes	60329.896	60779.839	60522.502	0.810	0.0034	0.0036	0.0000
5 Classes	60079.646	60629.577	60315.054	0.813	0.0880	0.0895	0.0000
6 Classes	59846.109	60496.027	60124.317	0.803	0.1124	0.1146	0.0000

FIGURE 3 | Latent class solutions.

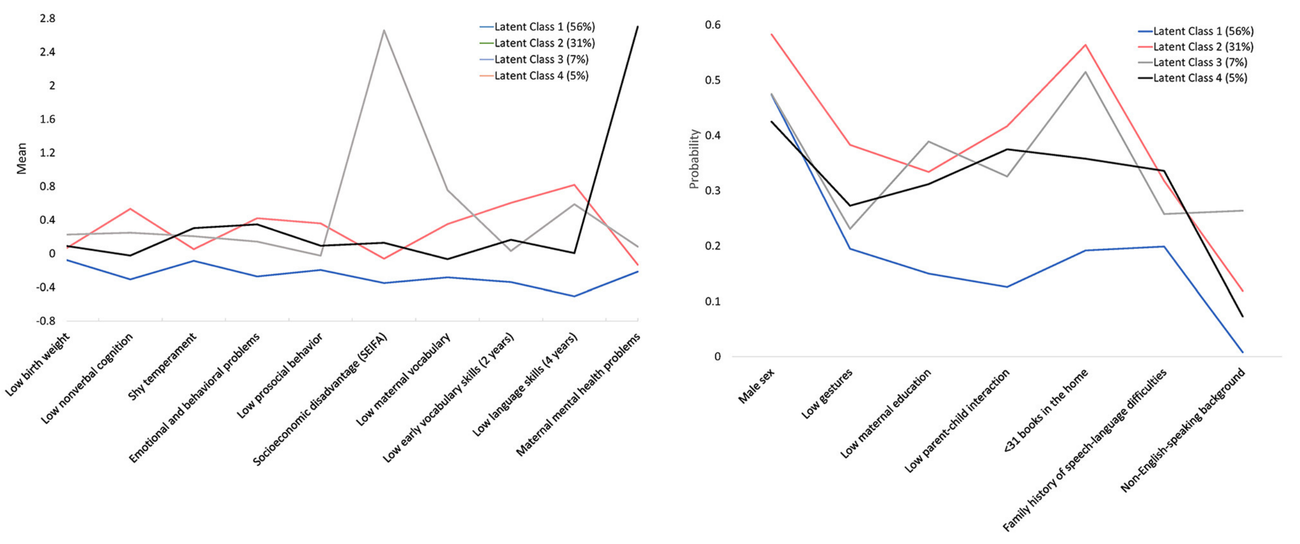


FIGURE 4 | The estimated means (continuous measures) and probabilities (binary measures) for latent classes.

analyses, including the 1,208 participants with a 7-year language outcome (see **Table 6**). The proportion of participants at 7-years with low language abilities was 10.5% ($n = 127$). The CELF core language score of the typical group ($n = 1{,}081$) was 100.6, with a standard deviation of 10.1 and a range of 82–135. In contrast the group with low language had a CELF core score of 71.9 (SD 10) and a range from 40 to 81. Compared to developmentally enabled children (class 1), the risk of low language at 7-years was 13 times greater for children in the developmentally vulnerable group with an insufficient home learning environment (class 2), eight times more likely for those children experiencing social disadvantage, low maternal education and insufficient home learning environment (class 3) and five times more likely for children whose mothers had more mental health problems (class 4).

DISCUSSION

This paper sought to determine the utility of developmental surveillance, as a way of identifying children at-risk of language difficulties. Using a cumulative risk approach, we wanted to inform the processes of monitoring language development over the first years of life. Moreover, identifying a clusters of influential early life factors, we aimed to provide guidance for tailored content for early interventions.

Cumulative risk categories revealed that the risk of low language outcomes at 7-years was significantly increased (from 5 up to 17 times more likely) when risks accumulated, with four, five, and six or more factors present. A broad range of early life factors were represented across all the cumulative categories including characteristics of the child, the family and the environmental context in which the child lives and learns, as well as maternal characteristics and parenting behaviors. These results confirmed one of the assumptions of Pennington's multiple deficit model, that the accumulation of factors increases the probability of developmental difficulties emerging. The findings also align with Hayiou-Thomas et al. (48) who found that cumulative risk was a core component of predicting language and reading difficulties in children at 12-years, using language and family history measures at 4-years. We included a broader set of early life factors, many measured in the first year of life, to investigate the "tipping point" for language vulnerability. From our analyses the accumulation of four or more risks was the critical point where children's risk of later language difficulties was significantly increased.

Several of the most frequently present factors are worth noting. Reflecting the home learning environment, the number of books in the home was consistently important in the current findings and replicates previous work with ELVS (45, 46), LSAC (i.e., frequency of reading to children) (71, 72) and clinical cohorts (73). In a bio-ecological framework this characteristic of parent behavior (i.e., number of books, frequency of reading) represents how the home learning environment can facilitate language learning strategies such as joint attention, labeling, expansion and responsive questioning. It is also the case that capturing this parent behavior during a child's first years is easier to measure than conducting parent-child observations during regular developmental monitoring. Furthermore, maternal resources, such as mental health and education level, influence parent-child interactions, specifically responsivity and reciprocity, features known to be important for language development (14, 74). This relationship has also been identified in studies using qualitative methods of clustering of risks to predict response to an early intervention (75). Early communication skills (e.g., gestures) and vocabulary are also key drivers of later language (2, 38) and need to be included in any subset of factors recommended for developmental surveillance.

All four developmental profiles in the latent class analysis were consistent with the "interaction" domain of the bio-developmental framework, where the foundations of healthy development (genes: "g") interact with sources of early adversity (environment: "e"). The profiles all represent a combination of the two interaction components, whether advancing outcomes for children with typical development (g) in supportive home learning environments (e) (class 1) or by weakening outcomes for vulnerable children, developmentally (g) and in their home learning environments (e) (class 2), or children who are growing up in disadvantaged circumstances with vulnerable home learning environments (e) where maternal education and language is low (g & e) (class 3). The fourth developmental profile is primarily driven by genetic factors, with vulnerable maternal mental health, child development (g) and low maternal language (g & e) (class 4). Attributing limited maternal resources to both a genetic and environmental source we relied on previous studies reported in the literature. The shared genetic and environmental influence of maternal resources was demonstrated in studies with twins (76). Pathways between parental language input and child language outcomes were determined by both substantial shared genetic influence and "child to parent" and "parent to child" relationship effects.

TABLE 5 | Distribution of risk factors across the four classes identified using Latent Class Analysis (N = 1,910)[a].

Risk factor	Class 1 (%)	Class 2 (%)	Class 3 (%)	Class 4 (%)
	(*N* = 1,073) 56.2%	**(*N* = 596) 31.2%**	**(*N* = 142) 7.4%**	**(*N* = 99) 5.2%**
Male sex	47.2	58.2	47.9	44.4
Low birth weight	3.3	5.2	7.3	4.1
Shy temperament	20.8	26.9	38.7	29.3
Emotional and behavioral problems	20.6	44.8	43.0	32.3
Low prosocial behavior	19.9	35.1	21.8	32.3
Low gestures	18.6	39.4	23.9	30.3
Low non-verbal cognition	10.2	41.4	27.5	18.4
Low early vocabulary skills at 2-years	16.1	54.0	31.3	36.6
Low language skills at 4-years	1.5	51.3	43.1	18.4
Non-English-speaking background	0.2	6.8	14.9	3.3
Socioeconomic disadvantage (SEIFA score)	8.5	23.7	100.0	26.3
Family history of speech-language difficulties	19.8	32.2	25.4	35.4
Low parent-child interaction	12.2	42.1	33.0	41.1
Low maternal education	14.2	35.4	38.7	30.6
Maternal mental health problems	25.1	30.6	34.7	100.0
Low maternal vocabulary	9.5	29.7	38.7	18.2
Number of books in the home (home learning environment)	18.5	57.9	51.7	36.6

[a]*Sample size ranges from 1,510 to 1,910.*

All four classes with these differing developmental profiles contained children with language scores within one standard deviation of the mean, considered within the typical range. We interpret this finding within a necessary but not sufficient framework, where risks can render development vulnerable but not for all children and not always at clinically diagnostic levels. For example, children growing up in socially disadvantaged circumstances do not all have developmental language difficulties, however, it is a known risk (77, 78). In our cluster model, the greatest variability in language scores was

TABLE 6 | Association between the four identified latent classes and risk of low language at age 7-years using binomial regression.

Latent class	CELF Core language score M (SD)	Risk ratio (95% CI)	p
Class 4 Maternal mental health problems	98.9 (12.6)	5.90 (2.60–13.40)	<0.001
Class 3 Socially disadvantaged	91.1 (17.9)	8.49 (4.07–17.71)	<0.001
Class 2 Developmentally vulnerable	86.5 (11.4)	13.67 (8.05–23.23)	<0.001
Class 1 Developmentally enabled	107.4 (9.8)	Reference group	

in class 3 (i.e., socially disadvantaged group, with a vulnerable home learning environment). It is important to note that a wide variety of language environments exist in families living with social disadvantage (i.e., not all socially disadvantaged children are exposed to vulnerable home learning environments). Of the classes considered "at risk" the highest mean language score was in class 4 (i.e., children from the maternal mental health and vulnerable child socio-emotional adjustment). We measured maternal mental health at 8-months and our findings suggest that the impact this factor has on language outcomes weakens as children get older. This is consistent with the findings of Taylor et al. (79) who found that maternal mental health distress was associated with higher rates of vocabulary growth between 4- and 8-years. It is possible, that children in this group had more fundamental issues with social-emotional development and adjustment, for which language, was one observable indicator, but not the primary source of vulnerability.

The developmental profiles in this study were comparable with previous work but demonstrated consistent gene-environment interactions unlike work from Christensen et al. (51) who demonstrated qualitatively different clusters of risks associated with vocabulary growth from 4- to 8-years. Six classes were included in the model representative of either environmental only (e.g., "working poor families"), genetic only (e.g., "developmental delay"), or interactions between both (e.g., "overwhelmed"). Similarly, developmental vulnerability at 5-years (47) and reading difficulties in later childhood (52) have been explained by latent class models with 5 and 4 classes, respectively. In both these models there was a mix of genetic only (e.g., child development risk) or environmental only (e.g., socioeconomic risk) classes, as well as classes derived from interactions between gene and environment characteristics (e.g., birth, sociodemographic and health behavior risks). Language skills measured at 7-years are complex including comprehension and expression of vocabulary, grammar, and semantic knowledge and are therefore, qualitatively different from vocabulary only and developmental vulnerability measured in previous studies. The socio-cultural and biological nature of language learning makes it particularly sensitive to both gene and environmental influences which may account for the consistent representation of both factors in the classes identified in our model.

Previous ELVS analyses at 4-, 7-, and 11-years have demonstrated associations between some early life factors and later language outcomes. However, due to large amounts of unexplained variance in language outcomes and the early instability in language profiles, the specific recommendations related to developmental surveillance and intervention were limited and under specified. Analysis of language growth curves from 4- to 7-years was more informative with respect to intervention levers. The inclusion in the present analyses of proximal and modifiable factors important for language outcomes together with early life factors provides compelling evidence to monitor children at-risk for later language difficulties based on a cumulative risk approach. In addition, the developmental profiles (classes) provide information about the content of early prevention and intervention strategies, which our findings suggest need to focus on the home learning environment and parent-child interactions. In the following sections we provide clinical implications and recommendations as they pertain to our findings.

Implications for Developmental Surveillance (Cumulative Risk)

Developmental surveillance provides the opportunity for flexible and continuous monitoring of development to meet the public health goal of detecting, preventing, and responding to specific disorders in the population. The present analyses suggest that surveillance based on just one to two risk factors will not be helpful in identifying young children with vulnerable language skills. Instead, through observing whether children have four or more accumulated risks, identification will be more accurate given the significant risk ratios associated with language outcomes at 7-years in our cohort. Importantly, most of the factors we considered are easily observed and reported by parents and/or non-specialist health and education staff who know the child. While observations of parent-child interaction are considered the gold standard measure, it is promising that parent report of interaction behaviors measured using the BPCIS identified those parents and children who may benefit from parent-child interaction intervention targeting responsive behaviors.

Specifically, a set of factors that include information about the child's early communication skills, the home learning environment, and parent-child interaction will be well-placed to identify children whose language development may be vulnerable and in need of ongoing monitoring. We recommend universal and regular monitoring throughout the preschool years and beginning in the first year of life, based on our current analyses and previous work (45). Leveraging off well-child health visits it is feasible that a set of factors could be included in a cumulative risk index for language development. While maternal and child health services provide an excellent opportunity to monitor during the very early years, from birth to 2-years, there is often a significant drop-off in families attending routine child health checks beyond this age and early childhood education and care contexts become more important for families and have wider uptake. Consequently, collaboration and seamless transitions from health monitoring to early childhood education services becomes vital to maintain contact with vulnerable children. Policy development that seeks to build collaborative partnerships and sharing of child data between early childhood health and education services is critical here.

Implications for Targeted Interventions (Latent Class Analysis)

The four developmental profiles represented in the classes identified are all characterized by interaction between genetic and environmental factors. Plomin (80) suggests that "genes are the major systematic force in children's development," going on to discuss how environmental factors influence development in ways that are not systematic but individual and contextualized; in this view the continuity of development is genetic and change is environmental (81). Consequently, intervention strategies focused on shifting developmental trajectories should first and foremost consider the environmental factors most likely to change and impact language abilities. Earlier ELVS findings and the current latent class model suggest that there are common goals for all children with vulnerable language abilities, specifically related to parent-child interaction and the home learning environment. Further, our findings speak to the need for individualized supports and strategies provided for sub-groups of at-risk children and families. Importantly, it is clear that across the developmental profiles, a key priority for intervention is the nurturing of parent-child relationships and responsive and reciprocal interactions. To achieve this outcome for families with differing developmental profiles will require tailored and personalized approaches to intervention.

Pleasingly, a recent systematic review and meta-analysis has demonstrated that parent-child interaction interventions that focus on promoting responsive parenting demonstrate greater effects on child cognitive development, parenting practices, and parent-child interactions when compared with interventions without a focus on responsive parenting (82). Universal early childhood services, such as maternal and child health, could utilize a targeted intervention approach following developmental surveillance. Our findings demonstrate that both child language and parent-child interaction measured through parent report can detect vulnerabilities and indicate the need for responsive parenting interventions. Observational tools, such as the Parental Responsiveness Rating Scale, provide an efficient, reliable method for practitioners to measure parental responsiveness during a brief, 5-min observation of parent-child interaction (83) and could be utilized in both identification and monitoring the impact of interventions.

Supplementing the common intervention goals, strategies that more precisely meet the needs of the different classes identified should include clinical supports and services for children in the vulnerable group who had delays in early communication skills. Prioritization of strategies that target language promotion and engagement of families in early childhood education and care settings is recommended for the third class where significant social disadvantage was a factor (84). It is important to note that there are broader social and structural inequalities which may make the provision of optimal home-learning environments challenging. Parent support programs, including engagement with health and/or early childhood agencies, need to provide a critical buffer for families where mental health problems are present.

Strengths and Limitations

ELVS is one of the few large prospective, community cohort studies with a focus on language skills to have collected data from the first year of life, and regularly through the preschool and primary school years. Language skills were measured at multiple time-points using gold standard measures. Using repeated surveys, we measured early life factors pertaining to the child, family, and environment. The robustness of this longitudinal data is rare for complex developmental disorders. These significant strengths have enabled ELVS to make a significant contribution to the literature on language development and disorder.

At the same time, we acknowledge that our sample at 7-years was not reflective of the original ELVS sample or the population more generally, both of which were more disadvantaged. Consequently, the latent classes and the impact of cumulative risk on language outcomes may be different with a more disadvantaged sample. The majority of caregivers in the ELVS sample are mothers and our findings should not be considered representative of father's behaviors or characteristics. Our data reflects the Australian societal characteristics from which it was collected; we accept that some features of which are more comparable to different country contexts than others. ELVS was not designed to assess specific or individual biological or genetic factors, so many of these are not included.

In a large study such as ELVS we did not have the resources to collect parental language input to children, *via* audio or video recordings, at multiple time-points. We acknowledge that this would have provided a rich source of data to further investigate parent-child interaction and is an important direction for future research. Finally, despite our analysis including an extensive set of early life factors, as for any study they are not a complete list of all the possible risk and protective factors may have resulted in different findings to those presented here.

CONCLUSION

The cumulative and cluster risk analyses demonstrate the potential to use developmental surveillance to identify children within the first year/s of life who are at increased risk for language difficulties. Next steps will require consultation with practitioners to determine feasibility of this approach. Building a cumulative risk index, our findings demonstrate a "tipping point" when children accumulate four or more risks from a range of early life factors. Many of these risks can be monitored efficiently and repeatedly through existing maternal and child health services. Importantly, parent-child interaction and the home learning environment emerged as common features of clusters of risks. We recommend they be adopted as the common focus for early intervention and universal language promotion programs that target all children and families in a community. Developing policy to implement these recommendations does not require wholesale restructuring of existing services but the careful allocation of resources and training to ensure universal and frequent developmental surveillance is available to all young children and their families regardless of their circumstances.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because data is shared following application to the Chief Investigators of the ELVS study. Requests to access the datasets should be directed to Dr. Penny Levickis, penny.levickis@unimelb.edu.au.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Royal Children's Hospital (Melbourne). Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

PE, CM, and SR conceived and designed the study. PL, EW, and BG analyzed the data. PE, PL, EW, EB, and BG drafted the manuscript. PE, PL, CM, EW, EB, RW, BG, and SR revised the manuscript critically for important intellectual content. All authors contributed to the article and approved the submitted version.

FUNDING

The Early Language in Victoria Study was funded by the Australian National Health and Medical Research Council (NHMRC grant numbers 237106, 436958, and 1041947). SR was supported by a Practitioner Fellowship (#1041892). Research at the Murdoch Children's Research Institute was supported by the Victorian Government's Operational Infrastructure Support Program.

ACKNOWLEDGMENTS

The authors thank all ELVS participating families and researchers.

REFERENCES

1. Bates E, Thal D, Finlay B, Clancy B. Early language development and its neural correlates. In: Segalowitz SJ, Rapin I, editors. *Handbook of Neuropsychology*. 2nd ed. Amsterdam (2002). p. 109–76.
2. Reilly S, Wake M, Ukoumunne OC, Bavin E, Prior M, Cini E, et al. Predicting language outcomes at 4 years of age: findings from Early Language in Victoria Study. *Pediatrics.* (2010) 126:e1530–7. doi: 10.1542/peds.2010-0254
3. Shonkoff JP, Phillips DA. *From Neurons to Neighborhoods: The Science of Early Childhood Development.* Washington, DC: National Research Council (US) and Institute of Medicine (US) (2000).
4. Wallace IF, Berkman ND, Watson LR, Coyne-Beasley T, Wood CT, Cullen K, et al. Screening for speech and language delay in children 5 years old and younger: a systematic review. *Pediatrics.* (2015) 136:e448–62. doi: 10.1542/peds.2014-3889
5. CDC Foundation. *What Is Public Health?* Atlanta GA: Center for Disease Control. Available from: https://www.cdcfoundation.org/what-public-health (accessed November 5, 2021).
6. Guralnick MJ. Why early intervention works: a systems perspective. *Infants Young Children.* (2011) 24:6–28. doi: 10.1097/IYC.0b013e3182002cfe
7. American Academy of Pediatrics CoCwD. Identifying infants and young children with developmental disorders in the medical home: an algorithm for developmental surveillance and screening. *Pediatrics.* (2006) 118:405–20. doi: 10.1542/peds.2006-1231
8. Wallace IF. *Universal Screening of Young children for Developmental Disorders: Unpacking the Controversies*. Research Triangle Park, NC: RTI Press (2018).
9. Nelson HD, Nygren P, Walker M, Panoscha R. Screening for speech and language delay in preschool children: systematic evidence review for the US Preventive Services Task Force. *Pediatrics.* (2006) 117:e298–319. doi: 10.1542/peds.2005-1467
10. Leblanc JC, Williams R. Developmental surveillance of young children in clinical settings: time to step out or step up? *Can Med Assoc J.* (2017) 189:E709–E10. doi: 10.1503/cmaj.161182
11. Chapman RS. Children's language learning: an interactionist perspective. *J Child Psychol Psychiatry*. (2000) 41:33–54. doi: 10.1017/S0021963099004953
12. Bruner J. *Child's Talk: Learning to Use Language*. New York, NY: W.W. Norton (1983).
13. Vygotsky LS. *Mind in Society: The Development of Higher Psychological Processes*. Cambridge: Harvard University Press (1978).
14. Rowe ML, Snow CE. Analyzing input quality along three dimensions: interactive, linguistic, and conceptual. *J Child Lang.* (2020) 47:5–21. doi: 10.1017/S0305000919000655
15. Johnson C, Beitchman JH, Brownlie EB. Twenty-year follow-up of children with and without speech-language impairments: family, educational, occupational, and quality of life outcomes. *Am J Speech Lang Pathol.* (2010) 19:51–65. doi: 10.1044/1058-0360(2009/08-0083)
16. Silva PA, McGee R, Williams SM. Developmental language delay from three to seven years and its signficance for low intelligence and reading difficulties at age seven. *Dev Med Child Neurol.* (1983) 25:783–93. doi: 10.1111/j.1469-8749.1983.tb13847.x
17. Beitchman JH, Brownlie EB, Bao L. Age 31 mental health outcomes of childhood language and speech disorders. *J Am Acad Child Adolesc Psychiatry.* (2014) 53:1102–10 e8. doi: 10.1016/j.jaac.2014.07.006
18. Beitchman JH, Nair R, Clegg M, Patel PG. Prevalence of speech and language disorders in 5-year-old kindergarten children in the Ottawa-Carleton region. *J Speech Hearing Disord.* (1986) 51:98–110. doi: 10.1044/jshd.5102.98
19. Tomblin JB, Records NL, Buckwalter P, Zhang X, Smith E, O'Brien M. Prevalence of specific language impairment in kindergarten children. *J Speech Lang Hearing Res.* (1997) 40:1245–60. doi: 10.1044/jslhr.4006.1245
20. Tomblin JB, Hardy JC, Hein HA. Predicting poor-communication status in preschool children using risk factors present at birth. *J Speech Hearing Res.* (1991) 34:1096–105. doi: 10.1044/jshr.3405.1096
21. Tomblin JB, Smith E, Zhang X. Epidemiology of specific language impairment: prenatal and perinatal risk factors. *J Commun Disord.* (1997) 30:325–48. doi: 10.1016/S0021-9924(97)00015-4
22. Zubrick SR, Taylor CL, Rice ML, Slegers DW. Late language emergence at 24 months: an epidemiological study of prevalence, predictors and covariates. *J Speech Lang Hearing Res.* (2007) 50:1562–92. doi: 10.1044/1092-4388(2007/106)
23. Laing GJ, Law J, Levin A, Logan S. Evaluation of a structured test and a parent led method for screening for speech and language

problems: prospective population based study. *BMJ.* (2002) 325:1152. doi: 10.1136/bmj.325.7373.1152
24. Law J, Boyle J, Harris F, Harkness A, Nye C. The feasibility of universal screening for primary speech and language delay: findings from a systematic review of the literature. *Dev Med Child Neurol.* (2000) 42:190–200. doi: 10.1017/S0012162200000335
25. Ward S. An investigation into the effectiveness of an early intervention method for delayed language development in young children. *Int J Lang Commun Disord.* (1999) 34:243–64. doi: 10.1080/136828299247405
26. Klee T, Carson DK, Gavin WJ, Hall L, Kent A, Reece S. Concurrent and predictive validity of an early language screening program. *J Speech Lang Hearing Res.* (1998) 41:627–41. doi: 10.1044/jslhr.4103.627
27. Klee T, Pearce K, Carson DK. Improving the positive predictive value of screening for developmental language disorder. *J Speech Lang Hearing Res.* (2000) 43:821–33. doi: 10.1044/jslhr.4304.821
28. de Koning HJ, de Ridder-Sluiter JG, van Agt HM, Reep-van den Bergh CMM, van der Stege HA, Korfage IJ, et al. A cluster-randomised trial of screening for language disorders in toddlers. *J Med Screen.* (2004) 11:109–16. doi: 10.1258/0969141041732229
29. Reilly S, Cook F, Bavin EL, Bretherton L, Cahir P, Eadie P, et al. Cohort profile: the early language in victoria study (ELVS). *Int J Epidemiol.* (2018) 47:11–20. doi: 10.1093/ije/dyx079
30. Bronfenbrenner U. *The Ecology of Human Development*. Cambridge, MA: Harvard University Press (1979).
31. Wetherby AM, Prizant BM. *Communication and Symbolic Behavior Scales Developmental Profile: Infant/Toddler Checklist*. Baltimore, MD: Paul H. Brookes Publishing Co. (2001).
32. Fenson L, Marchman VA, Thal D, Dale P, Reznick J, Bates E. *The MacArthur-Bates Communicative Development Inventories: User's Guide and Technical Manual*. 2nd ed. Baltimore, MD: Paul Brookes Publishing Company (2007).
33. Wiig EH, Secord WA, Semel E. *Clinical Evaluation of Language Fundamentals-Preschool: Australian Standardised Edition*. 2nd ed. Sydney, NSW: Harcourt Assessment (2006).
34. Semel E, Wiig EH, Secord WA. *Clinical Evaluation of Language Fundamentals - Fourth Edition, Australian Standardised Edition*. Marrickville, NSW: PsychCorp (2006).
35. Eadie P, Bavin EL, Bretherton L, Cook F, Gold L, Mensah F, et al. Predictors in infancy for language and academic outcomes at 11 years. *Pediatrics.* (2021) 147:e20201712. doi: 10.1542/peds.2020-1712
36. McKean C, Reilly S, Bavin EL, Bretherton L, Cini E, Conway L, et al. Language outcomes at 7 years: early predictors and co-occurring difficulties. *Pediatrics.* (2017) 139:e20161684. doi: 10.1542/peds.2016-1684
37. Reilly S, Eadie P, Bavin EL, Wake M, Prior M, Williams J, et al. Growth of infant communication between 8 and 12 months: a population study. *J Paediatr Child Health.* (2006) 42:764–70. doi: 10.1111/j.1440-1754.2006.00974.x
38. Reilly S, Wake M, Bavin EL, Prior M, Williams J, Bretherton L, et al. Predicting language at 2 years of age: a prospective community study. *Pediatrics.* (2007) 120:e1441–9. doi: 10.1542/peds.2007-0045
39. Bavin E, Bretherton L. The early language in victoria study: late talkers, predictors, and outcomes. In: Rescorla L, Dale P, editors. *Late Talkers: Language Development, Interventions, and Outcomes*. Baltimore, MD: Paul H. Brookes Publishing (2013). p. 3–22.
40. Zambrana IM, Pons F, Eadie P, Ystrom E. Trajectories of language delay from age 3 to 5: persistence, recovery and late onset. *Int J Lang Commun Disord.* (2014) 49:304–16. doi: 10.1111/1460-6984.12073
41. Poll GH, Miller CA. Late talking, typical talking, and weak language skills at middle childhood. *Learn Individual Differ.* (2013) 26:177–84. doi: 10.1016/j.lindif.2013.01.008
42. Eadie P, Nguyen C, Carlin J, Bavin E, Bretherton L, Reilly S. Stability of language performance at 4 and 5 years: measurement and participant variability. *Int J Lang Commun Disord.* (2014) 49:215–27. doi: 10.1111/1460-6984.12065
43. Ukoumunne OC, Wake M, Carlin J, Bavin EL, Lum J, Skeat J, et al. Profiles of language development in pre-school children: a longitudinal latent class analysis of data from the Early Language in Victoria Study. *Child Care Health Dev.* (2012) 38:341–9. doi: 10.1111/j.1365-2214.2011.01234.x
44. McKean C, Wraith D, Eadie P, Cook F, Mensah F, Reilly S. Subgroups in language trajectories from 4 to 11 years: the nature and predictors of stable, improving and decreasing language trajectory groups. *J Child Psychol Psychiatry.* (2017) 58:1081–91. doi: 10.1111/jcpp.12790
45. McKean C, Law J, Mensah F, Cini E, Eadie P, Frazer K, et al. Predicting meaningful differences in school-entry language skills from child and family factors measured at 12 months of age. *Int J Early Childhood.* (2016) 48:329–51. doi: 10.1007/s13158-016-0174-0
46. McKean C, Mensah FK, Eadie P, Bavin EL, Bretherton L, Cini E, et al. Levers for language growth: characteristics and predictors of language trajectories between 4 and 7 years. *PLoS ONE.* (2015) 10:e0134251. doi: 10.1371/journal.pone.0134251
47. Taylor CL, Christensen D, Stafford J, Venn A, Preen D, Zubrick SR. Associations between clusters of early life risk factors and developmental vulnerability at age 5: a retrospective cohort study using population-wide linkage of administrative data in Tasmania, Australia. *BMJ Open.* (2020) 10:e033795. doi: 10.1136/bmjopen-2019-033795
48. Hayiou-Thomas ME, Smith-Woolley E, Dale PS. Breadth versus depth: cumulative risk model and continuous measure prediction of poor language and reading outcomes at 12. *Dev Sci.* (2020) 24:e12998. doi: 10.1111/desc.12998
49. Pennington BF, Bishop DV. Relations among speech, language, and reading disorders. *Annu Rev Psychol.* (2009) 60:283–306.
50. Ring J, Black JL. The multiple deficit model of dyslexia: what does it mean for identification and intervention? *Ann Dyslexia.* (2018) 68:104–25. doi: 10.1007/s11881-018-0157-y doi: 10.1146/annurev.psych.60.110707.163548
51. Christensen D, Taylor CL, Zubrick SR. Patterns of multiple risk exposures for low receptive vocabulary growth 4-8 years in the longitudinal study of Australian children. *PLoS ONE.* (2017) 12:e0168804. doi: 10.1371/journal.pone.0168804
52. Taylor CL, Zubrick SR, Christensen D. Multiple risk exposures for reading achievement in childhood and adolescence. *J Epidemiol Commun Health.* (2019) 73:427–34. doi: 10.1136/jech-2018-211323
53. Shonkoff JP. Building a new biodevelopmental framework to guide the future of early childhood policy. *Child Dev.* (2010) 81:357–67. doi: 10.1111/j.1467-8624.2009.01399.x
54. Australian Bureau of Statistics. *Socio-Economic Indexes for Areas.* Canberra, ACT: Australian Bureau of Statistics (2001).
55. Kessler R, Mroczek D. *Final Version of our Non-Specific Psychological Distress Scale (Memorandum).* Ann Arbour, MI: Institute for Social Research (1994).
56. Raven JC. *Mill Hill Vocabulary Scale*. Oxford, United Kindgdom: JC Raven Ltd. (1997).
57. Mullan K, Daraganova G. Children's reading in Australia: the home and family context. In: *12th Australian Institute of Family Studies Conference*. Melbourne, VIC (2012).
58. Yu M, Daraganova G. *Children's Early Home Learning Environment and Learning Outcomes in the Early Years of School*. Melbourne, VIC: Australian Institute of Family Studies (2015).
59. Glascoe FP. The Brigance Infant and Toddler Screen: standardization and validation. *J Dev Behav Pediatrics.* (2002) 23:145–50. doi: 10.1097/00004703-200206000-00003
60. Glascoe FP, Leew S. Parenting behaviors, perceptions, and psychosocial risk: impacts on young children's development. *Pediatrics.* (2010) 125:313–9. doi: 10.1542/peds.2008-3129
61. Prior M, Sanson A, Oberklaid F. The Australian temperament project. In: Kohnstamm GA, Bates JE, Rothbart MK, editors. *Temperament in Childhood*. Chichester: Wiley (1989). p. 537–54.
62. Goodman R. The strengths and difficulties questionnaire: a research note. *J Child Psychol Psychiatry.* (1997) 38:581–6. doi: 10.1111/j.1469-7610.1997.tb01545.x
63. Toseeb U, Pickles A, Durkin K, Botting N, Conti-Ramsden G. Prosociality from early adolescence to young adulthood: a longitudinal study of individuals with a history of language impairment. *Res Dev Disabil.* (2017) 62:148–59. doi: 10.1016/j.ridd.2017.01.018
64. Kaufman AS, Kaufman NL. *Kaufman Brief Intelligence Test - Second Edition (K-BIT 2).* San Antonio TX: PsychCorp (2004).

65. Lewis V, Boucher J, Lupton L, Watson S. Relationships between symbolic play, functional play, verbal and non-verbal ability in young children. *Int J Lang Commun Disord.* (2000) 35:117–27. doi: 10.1080/136828200247287
66. Muthén LK, Muthén BO. *Mplus User's Guide*. 5th ed. Los Angeles CA: Muthén & Muthén (2007).
67. McLachlan G, Peel D. *Finite Mixture Models.* New York, NY: John Wiley & Sons, Inc. (2000).
68. Gudicha DW, Schmittmann VD, Tekle FB, Vermunt JK. Power analysis for the likelihood-ratio test in latent Markov models: shortcutting the Bootstrap P-value-based method. *Multivariate Behav Res.* (2016) 51:649–60. doi: 10.1080/00273171.2016.1203280
69. Asparouhov T, Muthén BO. Using Mplus TECH11 and TECH14 to test the number of latent classes. *Mplus web notes.* (2012) 14:1–17. Available online at: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.310.2341&rep=rep1&type=pdf
70. Nylund KL, Asparouhov T, Muthén BO. Deciding on the number of classes in latent class analysis and growth mixture modeling: a Monte Carlo simulation study. *Struct Equation Model.* (2007) 14:535–69. doi: 10.1080/10705510701575396
71. Farrant BM, Zubrick SR. Early vocabulary development: the importance of joint attention and parent-child book reading. *First Lang.* (2011) 32:343–64. doi: 10.1177/0142723711422626
72. Brown MI, Wang C, McLeod S. Reading with 1-2 year olds impacts academic achievement at 8-11 years. *Early Childhood Res Q.* (2022) 58:198–207. doi: 10.1016/j.ecresq.2021.09.008
73. Torppa M, Vasalampi K, Eklund K, Niemi P. Long-term effects of the home literacy environment on reading development: familial risk for dyslexia as a moderator. *J Exp Child Psychol.* (2022) 21:e105314. doi: 10.1016/j.jecp.2021.105314
74. Levickis P, Reilly S, Girolametto L, Ukoumunne O, Wake M. Maternal behaviors promoting language acquisition in slow-to-talk toddlers: prospective community-based study. *J Dev Behav Pediatrics.* (2014) 35:274–81. doi: 10.1097/DBP.0000000000000056
75. Short K, Eadie P, Kemp L. Influential factor combinations leading to language outcomes following a home visiting intervention: a qualitative comparative analysis (QCA). *Int J Lang Commun Disord.* (2020) 55:936–954. doi: 10.1111/1460-6984.12573
76. Dale PS, Tosto MG, Hayiou-Thomas ME, Plomin R. Why does parental language input style predict child language development? A twin study of gene-environment correlation. *J Commun Disord.* (2015) 57:106–17. doi: 10.1016/j.jcomdis.2015.07.004
77. Law J, McBean K, Rush R. Communication skills in a population of primary school-aged children raised in an area of pronounced social disadvantage. *Int J Lang Commun Disord.* (2011) 46:657–64. doi: 10.1111/j.1460-6984.2011.00036.x
78. Nelson KE, Welsh JA, Trup EMV, Greenberg MT. Language delays of impoverished preschool children in relation to early academic and emotion recognition skills. *First Lang.* (2011) 31:164–94. doi: 10.1177/0142723710391887
79. Taylor CL, Christensen D, Lawrence D, Mitrou F, Zubrick SR. Risk factors for children's receptive vocabulary development from four to eight years in the longitudinal study of Australian children. *PLoS ONE.* (2013) 8:e73046. doi: 10.1371/journal.pone.0073046
80. Plomin R. *Blueprint: How DNA Makes Us Who We Are*. Cambridge, MA: The MIT Press (2018).
81. Asbury K, Plomin R. *G Is for Genes: The Impact of Genetics on Education and Achievement*. Oxford: John Wiley & Sons (2013).
82. Jeong J, Franchett EE, Ramos De Oliveira CV, Rehmani K, Yousafzai AK. Parenting interventions to promote early child development in the first three years of life: a global systematic review and meta-analysis. *PLoS Med.* (2021) 18:e1003602. doi: 10.1371/journal.pmed.1003602
83. Levickis P, McKean C, Walls E, Law J. Training community health nurses to measure parent-child interaction: a mixed-methods study. *Euro J Public Health.* (2020) 30:445–50. doi: 10.1093/eurpub/ckz155
84. Frizelle P, Mullane E, O'shea A, Ceroni A, Dahly D, Horgan A, et al. Happy talk: a pilot effectiveness study of a targeted-selective speech-language and communication intervention for children from areas of social disadvantage. *Int J Lang Commun Disord.* (2021) 56:954–74. doi: 10.1111/1460-6984.12648

Author Disclaimer: The contents of the published material herein are the sole responsibility of the authors and do not reflect the views of the NHMRC.

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.

frontiers in Pediatrics

SYSTEMATIC REVIEW
published: 23 February 2022
doi: 10.3389/fped.2022.801220

Systematic Review and Meta-Analysis of Screening Tools for Language Disorder

*Kevin K. H. So and Carol K. S. To**

Academic Unit of Human Communication, Development, and Information Sciences, Faculty of Education, The University of Hong Kong, Hong Kong, Hong Kong SAR, China

Language disorder is one of the most prevalent developmental disorders and is associated with long-term sequelae. However, routine screening is still controversial and is not universally part of early childhood health surveillance. Evidence concerning the detection accuracy, benefits, and harms of screening for language disorders remains inadequate, as shown in a previous review. In October 2020, a systematic review was conducted to investigate the accuracy of available screening tools and the potential sources of variability. A literature search was conducted using CINAHL Plus, ComDisCome, PsycInfo, PsycArticles, ERIC, PubMed, Web of Science, and Scopus. Studies describing, developing, or validating screening tools for language disorder under the age of 6 were included. QUADAS-2 was used to evaluate risk of bias in individual studies. Meta-analyses were performed on the reported accuracy of the screening tools examined. The performance of the screening tools was explored by plotting hierarchical summary receiver operating characteristic (HSROC) curves. The effects of the proxy used in defining language disorders, the test administrators, the screening-diagnosis interval and age of screening on screening accuracy were investigated by meta-regression. Of the 2,366 articles located, 47 studies involving 67 screening tools were included. About one-third of the tests (35.4%) achieved at least fair accuracy, while only a small proportion (13.8%) achieved good accuracy. HSROC curves revealed a remarkable variation in sensitivity and specificity for the three major types of screening, which used the child's actual language ability, clinical markers, and both as the proxy, respectively. None of these three types of screening tools achieved good accuracy. Meta-regression showed that tools using the child's actual language as the proxy demonstrated better sensitivity than that of clinical markers. Tools using long screening-diagnosis intervals had a lower sensitivity than those using short screening-diagnosis intervals. Parent report showed a level of accuracy comparable to that of those administered by trained examiners. Screening tools used under and above 4yo appeared to have similar sensitivity and specificity. In conclusion, there are still gaps between the available screening tools for language disorders and the adoption of these tools in population screening. Future tool development can focus on maximizing accuracy and identifying metrics that are sensitive to the dynamic nature of language development.

Systematic Review Registration: https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=210505, PROSPERO: CRD42020210505.

Keywords: surveillance, screening, language disorder, PRISMA review, meta-analysis, summary receiver-operating characteristics, meta-regression

OPEN ACCESS

Edited by:
Daniel Holzinger,
Hospitaller Brothers of Saint John of God Linz, Austria

Reviewed by:
Karin Wiefferink,
Dutch Foundation for the Deaf and Hearing Impaired Child (NSDSK), Netherlands
Steffi Sachse,
Heidelberg University of Education, Germany

****Correspondence:***
Carol K. S. To
tokitsum@hku.hk

Specialty section:
This article was submitted to Children and Health, a section of the journal Frontiers in Pediatrics

Received: *25 October 2021*
Accepted: *13 January 2022*
Published: *23 February 2022*

Citation:
So KKH and To CKS (2022) Systematic Review and Meta-Analysis of Screening Tools for Language Disorder. Front. Pediatr. 10:801220. doi: 10.3389/fped.2022.801220

INTRODUCTION

Language disorder refers to persistent language problems that can negatively affect social and educational aspects of an individual's life (1). It is prevalent and estimated to affect around 7.6% of the population (2). Children with language disorder may experience difficulties in comprehension and/or in the use of expressive languages (3). Persistent developmental language disorder not only has a negative impact on communication but is also associated with disturbance in various areas such as behavioral problems (4), socio-emotional problems (5), and academic underachievement (6).

Early identification of persistent language disorder is challenging. There are substantial variabilities in the trajectories of early language development (7, 8). Some children display consistently low language, some appear to resolve the language difficulties when they grow older, and some demonstrated apparently typical early development but develop late-emerging language disorder. This dynamic nature of early language development has introduced difficulties in the identification process in practice (9). Therefore, rather than a one-off assessment, late talkers under 2 years old are recommended to be reassessed later. Referral to evaluation may not be not based on positive results in universal screening, but mainly concerns from caregivers, the presence of extreme deviation in development, or the manifestation of behavioral or psychiatric disturbances under 5 years old (9). Those who have language problems in the absence of the above conditions are likely to be referred for evaluation after 5 years old. Only then will they usually receive diagnostic assessment.

Ideally, screening should identify at-risk children early enough to provide intervention and avoid or minimize adverse consequences for them, their families, and society, improving the well-being of the children and the health outcomes of the population at a reasonable cost. Despite the high prevalence and big impact of language disorder, universal screening for language disorder is not practiced in every child health surveillance. Screening in the early developmental stages is controversial (10). While early identification has been advocated to support early intervention, there are concerns about the net cost and benefits of these early screening exercises. For example, the US Preventive Task Force reviewed evidence concerning screening for speech and language delay and concluded that there was inadequate evidence regarding the accuracy, benefits, and harms of screening. The Task Force therefore did not support routine screening in asymptomatic children (11). This has raised concerns in the professional community who believe in the benefits of routine screening (12). However, it is undeniable that another contributing factor for the recommendation of the Task Force was that screening tools for language disorder vary greatly in design and construct resulting in the variability in identification accuracy.

Previous reviews of screening tools for early language disorders have shown that these tools make use of different proxies for defining language issues, including a child's actual language ability, clinical markers such as non-word repetition, or both (13). Screening tools have been developed for children at different ages [e.g., toddlers (14) and preschoolers (15)] given the higher stability of language status at a later time point (16, 17). Screening tools also differ in the format of administration. For example, some tools are in the form of a parent-report questionnaire while some have to be administered by trained examiners via direct assessment or observations. Besides the test design, methodological variations have also been noted in primary validation studies, such as the validation sample, the reference standards (i.e., the gold standard for language disorder), and the screening-diagnosis interval. These variations might eventually lead to different levels of screening accuracy, which has been pointed out in previous systematic reviews (10, 13).

These variations have been examined in terms of the screening accuracy (13). Parent-report instruments and trained-examiner screeners have been found to be comparable in screening accuracy. In longitudinal studies in which language disorder status has been validated at various time points, accuracy appears to be lower for longer-term prediction than for concurrent prediction. Although the reviews have provided a comprehensive overview regarding the variations in different language screening tools, the analyses have mainly been based on qualitative and descriptive data. In the current study we performed a systematic review of all currently available screening tools for early language disorders that have been validated against a reference standard. We report on the variations noted in terms of (1) the type of proxy used in defining language disorders, (2) the type of test administrators, (3) the screening-diagnosis intervals and (4) age of screening. Second, we conducted a meta-analysis of the diagnostic accuracy of the screening tools and examined the contributions of the above four factors to accuracy.

METHODS

The protocol for the current systematic review was registered at PROSPERO, an international prospective register of systematic reviews (Registration ID: CRD42020210505, record can be found on https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=210505). Due to COVID-19, the registration was published with basic automatic checks in eligibility by the PROSPERO team. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses for Diagnostic Test Accuracy (PRISMA-DTA) (18) checklist was used as a guide for the reporting of this review.

Search Strategy

A systematic search of the literature was conducted in 2020 October based on the following databases: CINAHL Plus, ComDisDome, PsycINFO, PsycArticles, ERIC, PubMed, Web of Science, and Scopus. The major search terms were as follows: Child* OR Preschool* AND "Language disorder" * OR "language impairment*" OR "language delay" AND Screening OR identif*. To be as exhaustive as possible, the earliest studies available in the databases and those up to October 2020 were retrieved and screened. **Appendix A Table A1** showed the detailed search strategies in each database. Articles from the previous reviews were also retrieved.

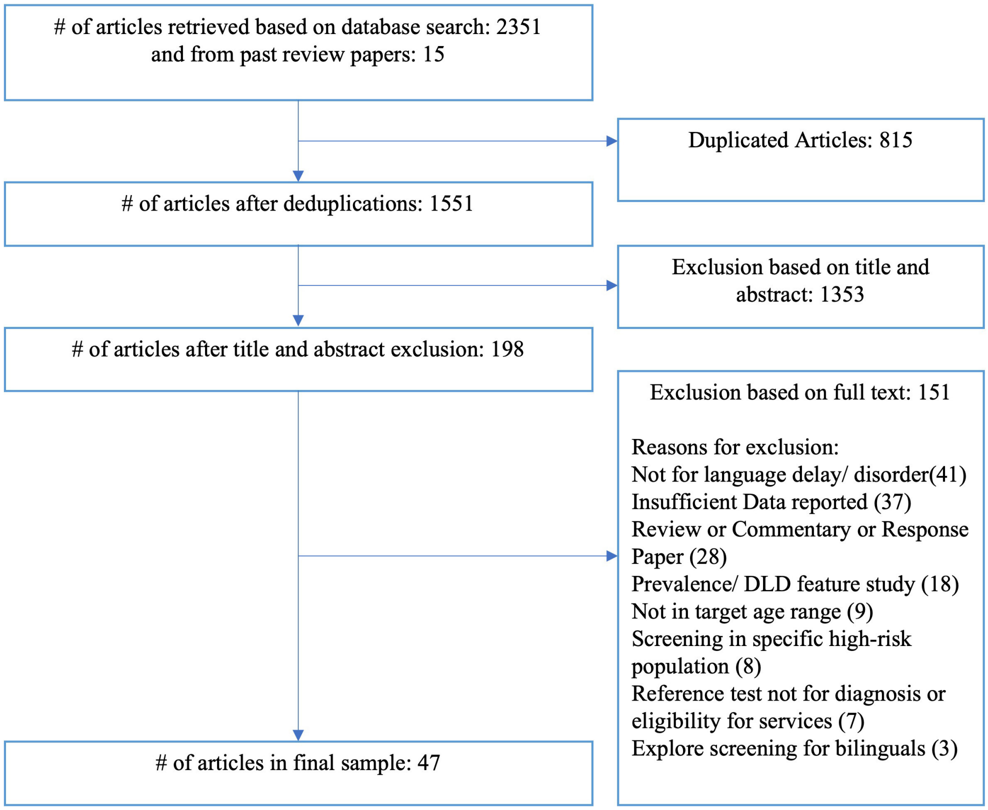


FIGURE 1 | Flow-chart for the inclusion and exclusion of articles in literature search.

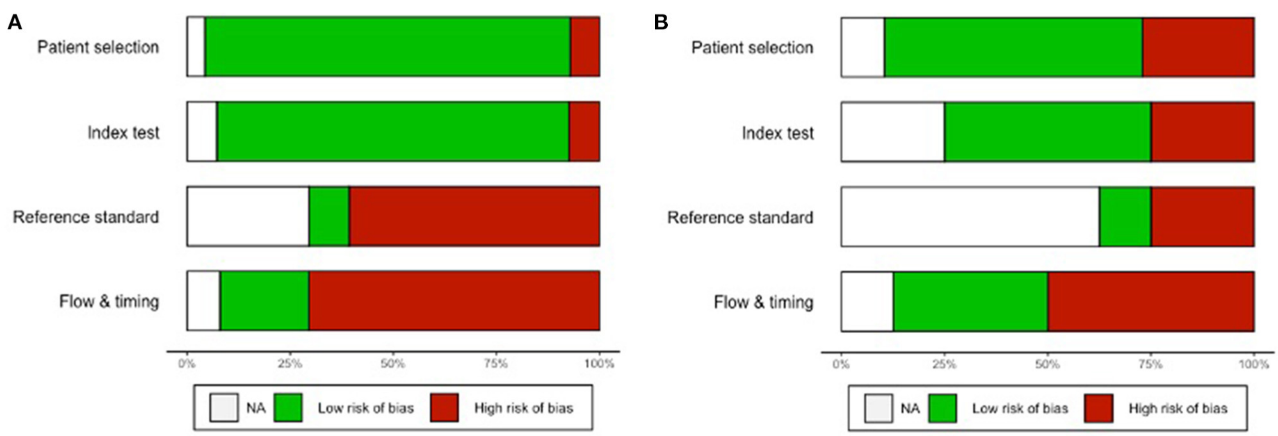


FIGURE 2 | (A) Weighted and **(B)** unweighted overall risk-of-bias as assessed using QUADAS-2.

Inclusion and Exclusion Criteria

The relevance of the titles, abstracts, and then the full texts were determined for eligibility. Cross-sectional or prospective studies validating screening tools or comparing different screening tools for language disorders were included in the review. The focus was on screening tools validated with children aged 6 or under from the general population or those with referral, regardless of the administration format of the tools, or how language disorder was defined in the studied. Studies that did not report adequate data on the screening results, and in which accuracy data cannot be deduced from the data reported, were excluded from the review (see **Appendix A Table A2** for details).

Data Extraction

Data was extracted by the first author using a standard data extraction form. The principal diagnostic accuracy measures extracted were test sensitivity and specificity. The number of people being true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) was also extracted. Sensitivity and specificity were calculated based on 2 by 2 contingency tables in the event of discrepancy between the text description and the data reported. The data extraction process was repeated after the first extraction to improve accuracy. Screening tools with both sensitivity and specificity exceeding 0.90 were regarded as good and those with both measures exceeding 0.80 but below 0.90 were regarded as fair (19).

Quality Assessment

Quality assessment of included articles was conducted by the first author using QUADAS-2 by Whiting, Rutjes (20). QUADAS-2 can assist in assessing risk of bias (ROB) in diagnostic test accuracy studies with signaling questions concerning four major domains. The ROB in patient selection, patient flow, index tests, or the screening tools in the current review, and the reference standard tests were evaluated. Ratings of ROB for individual studies were illustrated using a traffic light plot. A summary ROB figure weighted with sample size was generated using the R package “robvis” (21). Due to the large discrepancy in the sample size across studies, an unweighted summary plot was also generated to show the ROB of the included studies.

Data Analysis

The overall accuracy of the tools was compared using descriptive statistics. Because sensitivity and specificity are correlated, increasing either one of them by varying the cut-off of test positivity would usually result in a decrease in the other. Therefore, a bivariate approach was used to jointly model sensitivity and specificity (22) in generating hierarchical summary receiver-operating characteristic (HSROC) curves to assess the overall accuracy of screening by proxy and by screening-diagnosis intervals. HSROC is a more robust method accounting for both within and between study variabilities (23).

Three factors that could be associated with screening accuracy, chosen *a priori*, were included in the meta-analysis: proxy used, test administrators, and screening-diagnosis interval. Effect of screening age on accuracy was also evaluated. The effect of each variable was evaluated using a separate regression model. The variables of proxy used were categorical, with the categories being “child’s actual language,” “performance in clinical markers,” and “using both actual language and performance in clinical markers.” Test administrator was also a categorical variable with the categories being “parent” and “trained-examiners.” The variable of screening-diagnosis interval was dichotomously defined—intervals within 6 months were categorized as evaluating concurrent validity, whereas intervals of more than 6 months were categorized as evaluating predictive validity. The variable of screening age was also dichotomously defined with age 4 as the cut-off– those screened for children under the age of 4yo and those for children above 4yo. This categorization was primarily based on the age range of the sample, or the target screening age reported by the authors. Studies with age range that span across age 4 were excluded from the analysis. Considering the different thresholds used across studies and the correlated nature of sensitivity and specificity, meta-regression was conducted using a bivariate random effect model based on Reitsma et al. (22).

For studies examining multiple index tests and/or multiple cut-offs using the same population, only one screening test per category per study was included in the HSROC and meta-regression models. The test or cut-off with the highest Youden’s index was included in the meta-analytical models. Youden’s Index, *J*, was defined as

$$J = Sensitivity + Specificity - 1$$

All data analyses were conducted with RStudio Version 1.4.1106 using the package *mada* (24). Sensitivity analysis was carried out to exclude studies with a very high ROB (with 2 or more indicating a high risk in rating) to assess its influence on the results.

RESULTS

A total of 2351 articles, including 815 duplicates, were located using the search strategies, and an additional 15 articles were identified from previous review articles. After the inclusion and exclusion criteria were applied, a final sample of 47 studies were identified for inclusion in the review. **Figure 1** shows the number of articles included and excluded at each stage of the literature search.

Risk of Bias

The weighted overall ROB assessment for the 47 studies is shown in **Figure 2A**, and the individual rating for each study is shown in **Appendix B**. Overall, half of the data was exposed to a high ROB in the administration and interpretation of the reference standard test, while almost two-thirds of the data had a high ROB in the flow and timing of the study. As indicated by the unweighted overall ROB summary plot in **Figure 2B**, half of the 47 studies were unclear about whether the administration and interpretation of the reference standard test would introduce bias. This was mainly attributable to a lack of reporting of the reference standard test performance. About half of the studies had a high ROB in the flow and timing of the study. This usually arose from a highly variable or lengthy follow-up period.

Types and Characteristics of Current Screening Tools for Language Disorder

A total of 67 different index tests (or indices) were evaluated in the 47 included articles. The tests were either individual tests *per se* or part of a larger developmental test. The majority (50/67, 74.6%) of the screening tools examined children’s actual language. Thirty of these index tests involved parents or caregivers as the main informants. Some of these screening tools were in the form of a questionnaire with Yes-No questions regarding children’s prelinguistic skills, receptive language, or expressive language based on parent’s observations. Some used a vocabulary checklist (e.g., CDI, LDS) in which parents checked off the vocabulary their child can was able to comprehend and/or produce. Some tools also asked parents to report

TABLE 1 | Studies involving tools based on a child's actual language ability.

References	Agent	Index test	Reference standard test(s)	Sc. age (months)[a]	N	SN	SP	Accuracy[b]	Included in meta-analysis
Allen and Bliss (25)	Trained personnel	The Northwestern Syntax Screening Test (26)	Sequenced inventory of communication development (27)	36–47	182	0.92	0.48	Below fair	✓
Blaxley et al. (28)	Trained personnel	Bankson Language Screening Test (29)	Developmental sentence scoring (30)	48–72	90	0.46	0.94	Below fair	✓
Burden et al. (31)	Parents/caregivers	The Parent Language Checklist and The Developmental Profile II (32)	Action Picture Test (33), Bus Story test (34), self-developed test on receptive and phonological ability	36–39	425	0.87	0.45	Below fair	✓
Carscadden et al. (35)	Parents/caregivers	Speech and Language Pathology Early Screening Instrument (35)	Receptive Expressive Emergent Language Test – 3rd Edition (36)	17–23	53	0.91	0.95	Good	✓
Chaffee et al. (37)	Parents/caregivers	Minnesota Child Development Inventory – Comprehension Conceptual Language	Reynell Developmental Language Scales – revised (38)	24–87 M = 49	152	0.76	0.63	Below fair	✓
		Minnesota Child Development Inventory – Expressive Language (39)				0.89	0.45	Below fair	×
Dias et al. (40)	Parents/caregivers	Screening Tool by ASHA (41)	ABFW test (42)	0–60	962	0.83	0.99	Fair	✓
Dixon et al. (43)	Trained personnel	The Hackney Early Language Screening Test (43)	Reynell Developmental Language Scales (44), Lowe and Costello Symbolic Play Test (45)	30	40	0.94	0.95	Good	✓
Gray et al. (46)	Trained personnel	Expressive One-word Picture Vocabulary Test (47)	Referred by speech-language pathologist	48–60	62	0.71	0.71	Below fair	×
		Peabody Picture Vocabulary Test – III (48)				0.74	0.71	Below fair	×
		Receptive One-word Picture Vocabulary Test (49)				0.77	0.77	Below fair	✓
		Expressive Vocabulary Test (50)				0.71	0.68	Below fair	×
Guiberson (14)	Parents/caregivers	Parent reported vocabulary	Bilingual early childhood assessment team identification, parent report of concern, Spanish Preschool Language Scale – 4th Edition (51)	24–35	62	0.86	0.88	Fair	✓

(Continued)

TABLE 1 | Continued

References	Agent	Index test	Reference standard test(s)	Sc. age (months)[a]	*N*	SN	SP	Accuracy[b]	Included in meta-analysis
		Parent report of mean length of child's three longest utterances				0.46	0.93	Below fair	×
Guiberson and Rodriguez (52)	Parents/caregivers	Pilot Inventories III, translated version of MacArthur- Bates Communicative Development Inventory-III (53)	Spanish Preschool Language Scale – 4th Edition (51)	36–62 M = 45.5	48	0.82	0.81	Fair	✓
		Ages and Stages Questionnaire – communication subscales (54)				0.59	0.92	Below fair	×
Guiberson et al. (55)	Parents/caregivers	Reported children's three longest utterances	Parent concern, enrollment in speech-language intervention services, Spanish Preschool Language Scale – 4th Edition (51)	24–35 M = 29.4	45	0.91	0.86	Fair	✓
		Ages and Stages Questionnaire – communication subscales (56)				0.56	0.95	Below fair	×
		The Inventarios del Desarrollo de Habilidades Communicatives Palabras u Enunciado (57)				0.87	0.86	Fair	×
Guiberson et al. (58)	Parents/caregivers	Vocabulary score	SLP assessment, parental concern, Spanish Preschool Language Scale – 4th Edition (51)	37–69 M = 53.7	82	0.79	0.77	Below fair	✓
		Language questions				0.74	0.69	Below fair	×
Heilmann et al. (59)	Parents/caregivers	MacArthur- Bates Communicative Development Inventory – Words and Sentences (60)	Preschool Language Scale – 3rd Edition (61), language sampling	24 M = 23.8	100	0.68	0.98	Below fair	✓
Klee et al. (62)	Parents/caregivers	The Language Development Survey (63)	Mullen Scales of Early Learning (64), language sampling, parent interview, direct observation	24–26 M=24.7	64	0.91	0.87	Fair	×[c]

(Continued)

TABLE 1 | Continued

References	Agent	Index test	Reference standard test(s)	Sc. age (months)[a]	*N*	SN	SP	Accuracy[b]	Included in meta-analysis
Klee et al. (65)	Parents/caregivers	The Language Development Survey (63)	Mullen Scales of Early Learning (64), language sampling, language sampling, parent interview, direct observation	24–26 M = 24.7	64	0.91	0.96	Good	✓
Laing et al. (66)	Trained Personnel	Structured Screening Test	Reynell Developmental Language Scales – III (67)	30–36 M=32	282	0.66	0.89	Below fair	✓
Law (68)	Trained personnel	Structured Screening Test	Reynell Developmental Language Scales (2nd revision) (44)	30	189	0.86	0.76	Below fair	✓
Levett and Muir (69)	Trained personnel	Levett-Muir Language Screening Test (69)	Reynell Developmental Language Scales (revised) (70), Goldman-Fristoe Test of Articulation (71), Language Assessment and Remediation Procedure (72)	34.9–39.6	42	1	1	Good	✓
Visser-Bochane et al. (73)	Parents/caregivers	Early Language Screen (73)	LLC (74), SLC (75), LLP (76), SWP, SSP (77), LS-CCS (78), CCC-PCS (79)	12–72	124	0.79	0.86	Below fair	✓
Visser-Bochane et al. (80)	Trained personnel	The Dutch well child language screening protocol (80)	SLC (75), SWP, SSP (77)	26	265	0.62	0.93	Below fair	✓
Mattsson et al. (81)	Parents/caregivers and trained personnel	Questionnaire and Direct Observation by nurse	Clinical Examination by SLP	28–32 M = 30	105	0.81	0.87	Fair	✓
McGinty (82)	Parents/caregivers and trained personnel	The Mayo Early Language Screening Test (83)	Reynell Developmental Language Scales (44), Edinburgh Articulation Test (84)	18–60	200	0.84	0.7	Below fair	✓
Nair et al. (85)	Trained personnel	The Language Evaluation Scale Trivandrum For 0–3 Years (85)	Receptive-Expressive Emergent Language Scale (86)	0–36	643	0.96	0.78	Below fair	✓
Nayeb et al. (87)	Trained personnel	Nurse screening	Clinical Examination by SLP	29–31	100	1	0.85	Fair	✓
Puglisi et al. (15)	Trained personnel	Screening for Identification of Oral Language Difficulties by Preschool Teachers (15)	Expressive Vocabulary Test (88), Test for Reception of Grammar Version 2 (89), The Brazilian Children's Test of Pseudoword Repetition (90),	51–65 M = 57	100	0.86	0.95	Fair	✓
Rescorla (63)	Parents/caregivers	The Language Development Survey (63)	Reynell Developmental Language Scales (38)	23.7–34.4 M = 25.9	81	0.76	0.89	Below fair	✓
Rescorla and Alley (91)	Parents/caregivers	The Language Development Survey (63)	Reynell Developmental Language Scales (44)	23.7–34.4 M = 25.9	66	0.89	0.77	Below fair	✓

(Continued)

TABLE 1 | Continued

References	Agent	Index test	Reference standard test(s)	Sc. age (months)[a]	*N*	SN	SP	Accuracy[b]	Included in meta-analysis
Sachse and Von Suchodoletz (92)	Parents/caregivers	German version of the CDI, Toddler Form-2 (93)	Language Test for 2-Year-Old Children (94)	24–26	117	0.93	0.87	Fair	✓
Stokes (95)	Trained personnel	Nurse screen	Language sampling, Reynell Developmental Language Scales (70)	34–40	366	0.77	0.97	Below fair	✓
	Parent/caregivers	Parent Questionnaire				0.75	0.95	Below fair	×
van Agt et al. (96)	Parents/caregivers	Van Wiechen (96)	Specialists' judgement	26–58 M = 39	8,877	0.71	0.89	Below fair	✓
		General Language Screen (97)				0.81	0.78	Below fair	×
		Language Screening Instrument – Parent Form (98)				0.86	0.73	Below fair	×
	Trained personnel	Language Screening Instrument – Child Test (98)				0.54	0.88	Below fair	×
Walker et al. (99)	Parents/caregivers	Early Language Milestone Scale (100)	Sequenced Inventory of Communication Development (27)	0–36	77	0.77	0.85	Below fair	✓
Wetherby et al. (101)	Parents/caregivers	Communication And Symbolic Behavior Scales – Developmental Profile, Infant-Toddler Checklist (102)	Behavior Sample	12–24 M = 14.5	151	0.89	0.74	Below fair	✓

For tests that were validated against multiple cut-offs, only the one with highest Youden's index was shown; Sc. Age, screening age; MA, Meta-analysis; ASHA, American Speech-Language and Hearing Association; ABFW, Andrade CRF, Befi-Lopes DM, Fernandes FDM, Wertzner HF. Teste de Language Infantil nas Áreas de Fonologia, Vocabulário, Fluência e Pragmática. 2[nd] ed. Barueri: Pró-Fono, 2011; LLC, Lexilist Comprehension; SLC, Schlichting test for Language Comprehension; LLP, Lexilist Production; SWP, Schlichting test for Word Production; SSP, Schlichting test for Sentence Production; LS-CCS – Language Standard – Communication Schlichting test for Language Composite Score; CCC-PCS, CCC-2-NL-Pragmatic Composite Score.

[a]*Age of screening is reported in range or mean in the form of X_1-X_2 and $M=X_3$; In case range or mean is not reported, the intended age for screening of the tool will be reports as X_4.*

[b]*Based on Plante and Vance (19), Fair = over 0.8 in both sensitivity and specificity; Good = over 0.9 in both sensitivity and specificity.*

[c]*Not included because the sample was identical to Klee et al. (65).*

their child's longest utterances according to their observation and generated indices. The other 20 index tests on language areas were administered by trained examiners such as nurses, pediatricians, health visitors or speech language pathologists (SLPs). These screening tools were constructed as checklists, observational evaluations, or direct assessments, tapping into children's developmental milestones, their word combinations and/or their comprehension, expression, and/or articulation. Some of these direct assessments involved the use of objects or pictures as testing stimuli for children.

A small proportion (3/67, 4.48%) of tests evaluated clinical markers performance including non-word repetitions and sentence repetitions rather than children's actual structural language skills or communication skills. About nine percent (6/67, 8.96%) screened for both language abilities and clinical markers. Both types of tests required trained examiners to administer them. The tests usually made use of a sentence repetition task and one test also included non-word repetition. Another nine percent (6/67, 8.96%) utilized indices from language sampling, such as percentage of grammatical utterances (PGU), mean length of utterances in words (MLU3-W), and number of different words (NDW) as proxies. These indices represented a child's syntactic, semantic, or morphological performance. The smallest proportion (2/67, 2.99%) of the tests elicited parental concerns about their children being screened for language disorder. One asked parents to rate their concern using a visual analog scale, while the other involved interviews with the parents by a trained examiner.

Sixty-five of the 67 screening tools had reported concurrent validity. **Tables 1–5** summarize the characteristics of these 65 studies by the proxy used. Nine studies investigated the predictive validity of screening tools. **Table 6** summarizes the studies. All the studies used child's actual language ability as the proxy.

Screening Accuracy

Two of the 67 screening tools only reported predictive validity. Of the 65 screening tools that reported concurrent validity, about one-third (23/65, 35.4%) achieved at least fair accuracy and a smaller proportion (9/65, 13.8%) achieved good accuracy. The nine tools which achieved good accuracy include (i) Non-word Repetition, (ii) Speech and Language Pathology Early Screening Instrument (35), (iii) The Hackney Early Language Screening Test (43), (iv) The Language Development Survey (63), (v) Levett-Muir Language Screening Test (69), (vi) The Grammar and Phonology Screening (GAPS) Test (105), (vii) Tamiz de Problemas de Lenguaje (113), (viii) The Screening Kit of Language Development (117) and (ix) Short Language Measures (120).

Screening Performance by Proxy and Screening-Diagnosis Interval

Screening tools based on children's actual language ability had a sensitivity ranging from 0.46 to 1 (median = 0.81) and a specificity of 0.45 to 1 (median = 0.86). About 30% of the studies showed that their tools achieved at least fair accuracy, while 8.89% achieved good accuracy. Screening tools using clinical markers had a sensitivity ranging from 0.3 to 1 (median =

TABLE 2 | Studies involving tools based on clinical marker.

References	Agent	Index test	Reference standard test(s)	Sc. age[a] (months)	*N*	SN	SP	Accuracy[b]	Included in meta-analysis
Guiberson et al. (58)	Trained personnel	Non-word Repetition	SLP assessment, parental concern, Spanish Preschool Language Scale – 4th Edition (51)	37–69 M = 53.7	82	0.74	0.75	Below fair	✓
Kapalkova et al. (103)	Trained personnel	Non-word repetition	Clinical judgment and qualitative assessment	51–66	32	0.94	1	Good	✓
Nash et al. (104)	Trained personnel	The Grammar and Phonology Screening (GAPS) Test (105)	Clinical Evaluation of Language Fundamentals – Preschool, 2nd Edition (106)	36–72 M = 62.3	106	0.3	0.91	Below fair	✓
Sturner et al. (107)	Trained personnel	The Sentence Repetition Screening Task (108)	Illinois Test of Psycholinguistic Abilities (109), Bankson Language Screening Test (29)	54–66 Med = 60	323	0.62	0.91	Below fair	✓
van der Lely et al. (110)	Trained personnel	The Grammar and Phonology Screening (GAPS) Test (105)	Assessment by SLP and educational psychologist	43–80	41	1	1	Good	✓

For tests that were validated against multiple cut-offs, only the one with highest Youden's index was shown; Sc. Age, screening age.
[a]*Age of screening is reported in range, mean or median in the form of X_1-X_2, M=X_3 and Med=X_4, respectively.*
[b]*Based on Plante and Vance (19), Fair = over 0.8 in both sensitivity and specificity; Good = over 0.9 in both sensitivity and specificity.*

TABLE 3 | Studies involving tools based on both language ability and clinical marker.

Study	Agent	Index test	Reference standard test(s)	Sc. age[a] (months)	*N*	SN	SP	Accuracy[b]	Included in meta-analysis
Allen and Bliss (25)	Trained personnel	The Fluharty Preschool Screening Test (111)	Sequenced Inventory of Communication Development (112)	36–47	182	0.6	0.81	Below fair	✓
Benavides et al. (113)	Trained personnel	Tamiz de Problemas de Lenguaje (113)	Clinical Evaluation of Language Fundamentals-5th edition, Spanish Version (114)	48–72	200	0.94	0.92	Good	✓
Blaxley et al. (28)	Trained personnel	The Fluharty Preschool Screening Test (115)	Developmental Sentence Scoring (116)	48–72	90	0.36	0.96	Below fair	✓
Bliss and Allen (117)	Trained personnel	The Screening Kit of Language Development (118)	Sequenced Inventory of Communication Development (112), clinical judgment by SLP	30–48	100	1	0.93	Good	✓
Lavesson et al. (119)	Trained personnel	Language tasks and non-word repetition (119)	SLP judgment based on test results	46–53 M = 48.5	328	0.84	0.96	Fair	✓
Matov et al. (120)	Trained personnel	Short Language Measures (121)	Clinical Evaluation of Language Fundamentals-4 (122)	63.6	126	0.94	0.93	Good	✓
Wright and Levin (123)	Trained personnel	Preschool Articulation and Language Screening (123)	SLP judgement based on test results	26–81	152	0.71	0.94	Below fair	✓

For tests that were validated against multiple cut-offs, only the one with highest Youden's index was shown; Sc. Age, screening age.

[a] *Age of screening is reported in range or mean in the form of X_1-X_2 and M=X_3; In case range or mean is not reported, the intended age for screening of the tool will be reports as X_4.*

[b] *Based on Plante and Vance (19), Fair = over 0.8 in both sensitivity and specificity; Good = over 0.9 in both sensitivity and specificity.*

TABLE 4 | Studies involving tools based on language sampling.

References	Agent	Index test	Reference standard test(s)	Sc. age[a] (months)	*N*	SN	SP	Accuracy[b]	Included in meta-analysis
Eisenberg and Guo (124)	Trained personnel	Percentage Grammatical Utterances	LI2: Previously diagnosed LI3: Parent rating, Structured Photographic Expressive Language Test – Preschool 2nd Edition (125)	36–47	34	1	0.88	Fair	✓
		Percentage Sentence Point			34	1	0.82	Fair	×
		Percentage Verb Tense Usage (126)			34	1	0.82	Fair	×
Guiberson et al. (58)	Trained personnel	Ungrammaticality Index	SLP assessment, parental concern, Spanish Preschool Language Scale – 4th Edition (51)	37–69 M = 53.7	82	0.59	0.67	Below fair	×
	Trained personnel	Mean Length of Utterances in Words				0.65	0.92	Below fair	✓
Guiberson (14)	Parents/caregivers	Number of Different Words	Bilingual early childhood assessment team identification, parent report of concern, Spanish Preschool Language Scale – 4th Edition (51)	24–35	62	0.73	0.83	Below fair	✓

For tests that were validated against multiple cut-offs, only the one with highest Youden's index was shown; Sc. Age, screening age; LI2, language impairment at age 2; LI3, language impairment at age 3.
[a]*Age of screening is reported in range or mean in the form of X_1-X_2 and M=X_3; In case range or mean is not reported, the intended age for screening of the tool will be reports as X_4.*
[b]*Based on Plante and Vance (19), Fair = over 0.8 in both sensitivity and specificity; Good = over 0.9 in both sensitivity and specificity.*

TABLE 5 | Studies involving tools based on parental concern.

References	Agent	Index test	Reference standard test(s)	Sc. age[a] (months)	*N*	SN	SP	Accuracy[b]	Included in meta-analysis
Laing et al. (66)	Parents/caregivers	Parent led method	Reynell Developmental Language Scales – III (67)	30–36 M = 32	176	0.79	0.74	Below fair	✓
van Agt et al. (96)	Parents/caregivers	Visual analog scale to evaluate child's language development	Specialists' judgement	26–58 M = 39	8,877	0.76	0.81	Below fair	✓

For tests that were validated against multiple cut-offs, only the one with highest Youden's index was shown; Sc. Age, screening age.
[a]Age of screening is reported in range or mean in the form of X_1-X_2 and M=X_3; In case range or mean is not reported, the intended age for screening of the tool will be reports as X_4.
[b]Based on Plante and Vance (19), Fair = over 0.8 in both sensitivity and specificity; Good = over 0.9 in both sensitivity and specificity.

0.71) and a specificity of 0.45 to 1 (median = 0.91). Two of the five studies[1] (40%) evaluating screening tools based on clinical markers showed their tools had good sensitivity and good specificity, but the other three studies showed a sensitivity and a specificity below fair. Concerning screening tools based on both actual language ability and clinical marker performance, the sensitivity ranged from 0.36 to 1 (median = 0.84), and the specificity ranged from 0.81 to 0.96 (median=0.93) and above half of these studies (4/7[2], 57.1%) achieved at least fair performance in both sensitivity and specificity, and 3 of the 7 studies achieved good performance. Screening tools based on indices from language sampling had sensitivity ranging from 0.59 to 1 (median = 0.865) and specificity ranging from 0.67 to 0.92 (median = 0.825). Half of these six screening tools achieved fair accuracy, but none achieved good accuracy. None of the two screening tools based on parental concern achieved at least fair screening accuracy.

Fifteen of the 65 studies also reported predictive validity, with a sensitivity ranging from 0.32 to 0.94 (median = 0.81) and a specificity ranging from 0.61 to 0.93 (median = 0.85). Three of the tools (20%) achieved at least fair accuracy in both sensitivity and specificity, but none of them were considered to have good accuracy.

Test Performance Based on HSROC

Three HSROC curves were generated for screening tools based on language ability, clinical markers, both language ability and clinical markers, and those assessing concurrent validity. Two HSROC curves were generated for screening tools administered by trained examiners and parents/ caregivers, respectively. Two HSROC curves were generated for screening under and above the age of 4, respectively. A separate HSROC curve was generated for screening tools assessing predictive validity. Screening based on indices from language sampling (n = 3) or parental concern (n = 2) were excluded from the HSROC analysis due to the small number of primary studies.

Figure 3 shows the overall performance of screening tools based on language ability, clinical markers and both. Visual inspection of the plotted points and confidence region revealed considerable variation in accuracy in all three major types of screening tools. The summary estimates and confidence regions indicated that the overall performance of screening tools based on language ability achieved fair specificity (<0.2 in false positive rate) but fair-to-poor sensitivity. Screening tools based on clinical markers showed considerable variation in both sensitivity and specificity in that both measures ranged from good-to-poor. Screening tools based on both language ability and clinical markers achieved good-to-fair specificity, but fair-to-poor sensitivity. **Figure 4** shows the overall performance of

[1]The total number here refers to the number of studies: there were five studies evaluating tools based on clinical markers, but there were in total three different tests; hence number is different from that in types and characteristics of current screening tools for language disorder.

[2]The total number here refers to the number of studies: there were seven studies evaluating tools based on both actual language and clinical markers, but there were in total six different tests; hence number is different from that in types and characteristics of current screening tools for language disorder.

TABLE 6 | Studies assessing predictive validity of screening tools.

References	Agent	Index test	Sc. age (months)	Sc-V int. (months)	F/U age (months)	Reference standard test(s)	*N*	SN	SP	Accuracy[a]	MA included
Bruce et al. (127)	Parents/caregivers and trained personnel	Direct assessment through play and parent questionnaire	18–22	NA	54	NELLI (128)[b], The Test for Reception of Grammar (129)	43	0.6	0.85	Below fair	✓
Frisk et al. (130)	Trained personnel	Early Screening Profiles (131)	54	NA	60	Preschool Language Scale – 4th Edition (132)	110	0.86	0.81	Fair	✓
	Parents/caregivers	Ages and Stages Questionnaire (54)				Bracken Basic Concepts Scale	110	0.84	0.66	Below fair	×
	Trained personnel	Battelle Developmental Inventory Screening Test (133)				Preschool (134) Language Scale – 4th Edition (132)	110	0.68	0.86	Below fair	×
	Trained personnel	Brigance Preschool Screen (135)				Preschool Language Scale – 4th Edition (132)	110	0.91	0.78	Below fair	×
Jessup et al. (136)	Trained personnel	Kindergarten Development Check (137)	48–54	8–12	NA	Clinical Evaluation of Language Fundamentals-4 (122)	286	0.5	0.93	Below fair	✓
Klee et al. (62)	Parents/caregivers	The Language Development Survey (63)	24	NA	36–40	Mullen Scales of Early Learning (64), language sampling, parent interview, direct observation	36	0.67	0.9	Below fair	✓
Pesco and O'Neill (138)	Parents/caregivers	Language Use Inventory (139)	24–47	14.54–54.76	NA	DELV- NR (140) CELF-2 (141), Children's Communication Checklist – 2nd Edition (142)	236	0.81	0.93	Fair	✓
Sachse and Von Suchodoletz (92)	Parent/caregivers	German Version of The CDI, Toddler Form-2 (93)	24–26	12	NA	Language Test For 3–5-Year-Old Children (94)	102	0.94	0.61	Below Fair	×
	Trained personnel	Language Test for 2-Year-Old Children (94)	24–26	12	NA	Language Test For 3–5-Year-Old Children (94)	102	0.94	0.64	Below Fair	✓
Visser-Bochane et al. (80)	Trained personnel	The Dutch well-child language screening protocol (80)	$M = 26$	12	NA	SLC (75), SWP, SSP (77)	123	0.82	0.74	Below Fair	✓
Westerlund et al. (143)	Parents/caregivers	The Swedish Communication Screening at 18 Months of Age (144, 145)	18	NA	36	LO-3 (146, 147)	891	0.5	0.9	Below Fair	✓
	Trained personnel	Traditional Methods	18	NA	36	LO-3 (146, 147)	1,189	0.32	0.91	Below Fair	×
Wetherby et al. (101)	Parents/caregivers	Communication And Symbolic Behavior Scales – Developmental Profile Infant-Toddler Checklist (102)	12–24	$M = 14.5$	NA	Mullen Scales of Early Learning (148), Preschool Language Scale – 3rd Edition (61)	246	0.81	0.79	Below Fair	×

(Continued)

TABLE 6 | Continued

References	Agent	Index test	Sc. age (months)	Sc-V int. (months)	F/U age (months)	Reference standard test(s)	*N*	SN	SP	Accuracy[a]	MA included
	Trained personnel	Behavioral Sample	12–24	*M* = 18.2	NA	Mullen Scales Of Early Learning (148), Preschool Language Scale – 3rd Edition (61)	90	0.84	0.85	Fair	✓

For tests that were validated against multiple cut-offs, only the one with highest Youden's index was shown; Sc. Age, screening age; Sc-V int., Screening-validation Interval; F/U age, age at follow-up; DELV-NR, Diagnostic Evaluation of Language Variation – Norm Referenced; CELF-2, Clinical Evaluation of Language Fundamentals – Preschool, 2nd Edition; LO-3, Language Observation at 3 years of age.

[a]Based on Plante and Vance (19), Fair = over 0.8 in both sensitivity and specificity; Good = over 0.9 in both sensitivity and specificity.

[b]Spraklig snabbscreening av forskolebarn 3–6 arunderlag for diagnostisering av art och grad av sprakstorning, Stora Fonemtestet. Pedagogisk, Grammatiktest. Pedagogisk.

*[c]Based on **Table 5** in the paper, description in the discussion differed from the figures in the table.*

screening tools administered by parents/caregivers or trained examiners. Visual inspection revealed that both types of screening tools achieved fair-to-poor sensitivity and good-to-fair specificity. **Figure 5** shows the overall performance of screening for children under and above 4yo, respectively. Visual inspection revealed screening under 4yo achieved good-to-poor sensitivity and specificity, while screening above 4yo achieved good-to-poor sensitivity and good-to-fair specificity. **Figure 6** shows the performance of the screening tools evaluating predictive validity. These screening tools achieved fair-to-poor sensitivity and specificity.

Meta-Regression Investigating Effects of Screening Proxy, Test Administrator, Screening-Diagnosis Interval, and Age of Screening

The effects of screening proxy, test administrator, screening-diagnosis interval and age of screening on screening accuracy were investigated using bivariate meta-regression. **Table 7** summarizes the results. Screening tools with <6-month screening-diagnosis interval (i.e., concurrent validity) were associated with higher sensitivity when compared to those with longer than a 6-month interval (i.e., predictive validity). Tools using language ability as the proxy showed a marginally significantly higher sensitivity than those based on clinical markers. Screening tools based on language ability and those based on both language ability and clinical markers appeared to show a similar degree of sensitivity. For tools assessing concurrent validity, screening under the age of 4 had a higher sensitivity with marginal statistical significance but showed similar specificity with screening above the 4yo. As for tools assessing predictive validity, screening under and above 4yo appeared to show similar sensitivity and specificity. Similarly, screening tools relying on parent report and those conducted by trained examiners appeared to show a similar sensitivity. Despite the large variability in specificity, none of the factors in the meta-regression model explained this variability.

Results of sensitivity analysis after excluding studies with high ROB are illustrated in **Table 8**. The observed higher sensitivity for screening tools using actual language as proxy compared with those using clinical markers became statistically significant. The difference in sensitivity between screening tools assessing concurrent validity and those assessing predictive validity appeared to be larger than before the removal of the high ROB studies. However, the observed marginal difference between screening under and above 4yo became non-significant after the exclusion of high-risk studies. Similar to the results without excluding studies with high ROB, none of the included factors in sensitivity analysis explained variation in specificity.

DISCUSSION

The present review shows that currently available screening tools for language disorders during preschool years varies widely in their design and screening performance. Large variability in screening accuracy across different tools was a major issue

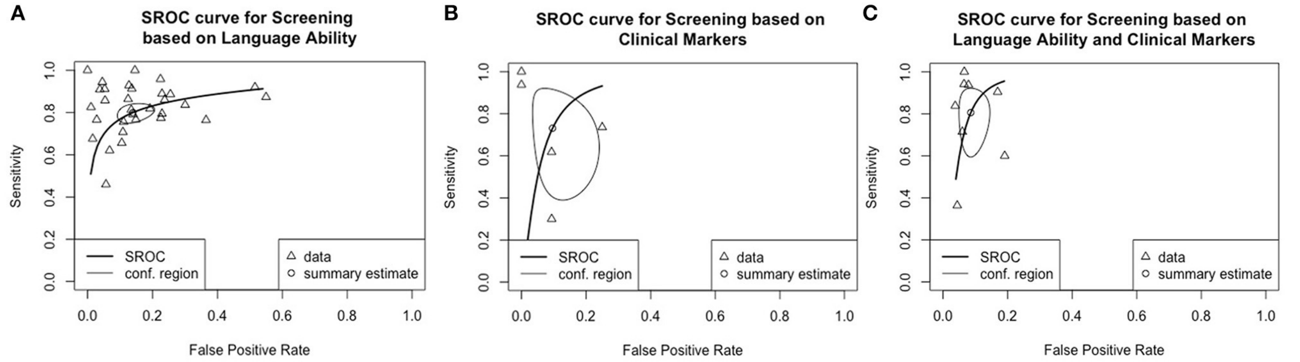


FIGURE 3 | Summary receiver operating characteristics curves for screening tools based on **(A)**, language ability, **(B)** clinical markers, and **(C)** language & clinical markers.

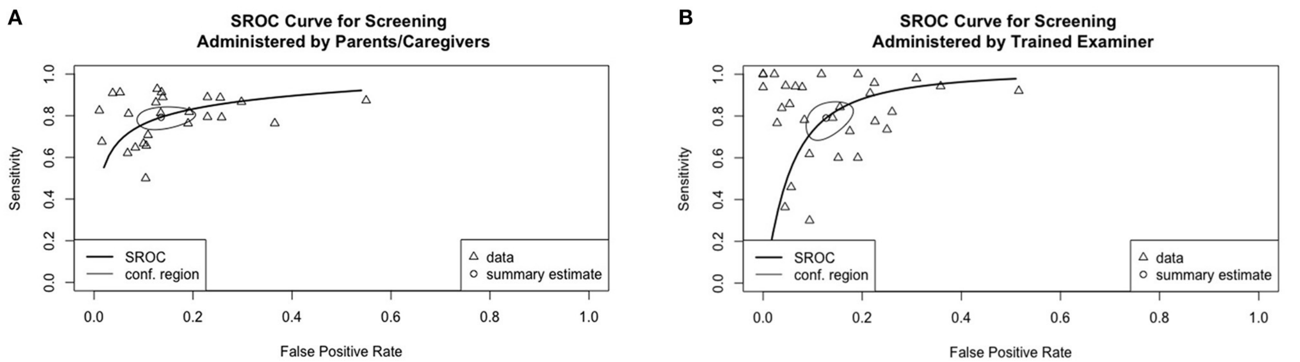


FIGURE 4 | Summary receiver operating characteristics curves for screening tools administered by **(A)** parents/caregivers and **(B)** trained examiners.

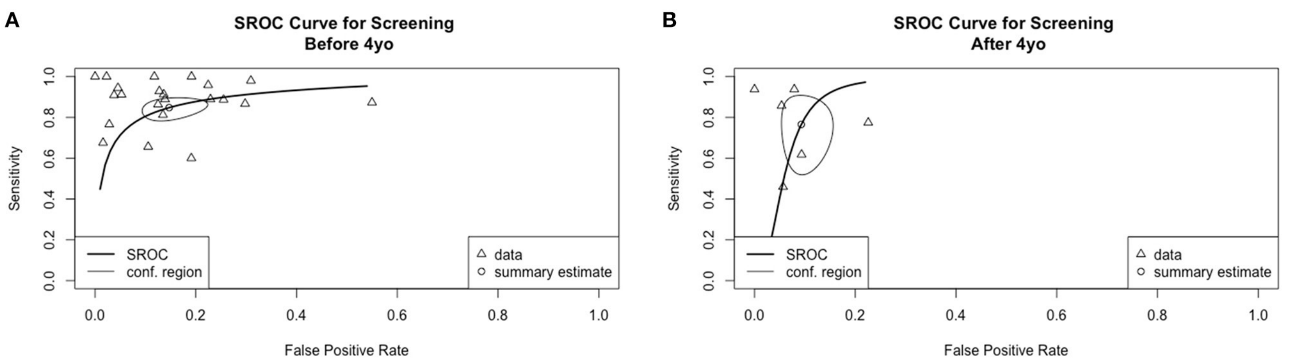


FIGURE 5 | Summary receiver operating characteristics curves for screening **(A)** under 4-year-old and **(B)** above 4-year-old.

in screening for language disorder. The present review also revealed that the variations arose from the choices of proxy and screening-diagnosis interval.

Screening tools based on children's actual language ability were shown to have higher sensitivity than tools based on clinical markers. The fact that screening tools based on clinical markers did not prove to be sensitive may be related to the mixed findings from primary studies. Notably, one of the primary studies using non-word repetition and sentence repetition tasks showed perfect accuracy in classifying all children with and without language disorder (110). The findings, however, could not be replicated in another study, using exactly the same test, which identified only 3 of the 10 children with language disorder (104). The difference highlighted the large variability in the performance of non-word and sentence repetition even among children with language disorders, in addition to the inconsistent difference found between children with and without language disorder (149). Another plausible explanation for the relatively higher sensitivity of using child's actual language skills lies in the resemblance between the items used for screening based on the child's actual language and the diagnostic tests used as the reference standard. Differences in task design and test item selection across studies may have further increased the inconsistencies (149). Therefore, in future tool development or refinement, great care should be taken in the choice of screening proxy. More systematic studies directly comparing how different proxies and factors affect screening accuracy are warranted.

There was no evidence that other factors related to tool design, such as the test administrators of the screening tools, explained variability in accuracy. In line with a previous review (13), parent-report screening appeared to perform similarly to screening administered by trained examiners. This seemingly comparable accuracy supports parent-report instruments as a viable tool for screening, in addition to their apparent advantage of lower cost of administration. Primary studies directly comparing both types of screening in the same population may provide stronger evidence concerning the choice of administrators.

As predicted, long term prediction was harder to achieve than estimating concurrent status. Meta-analysis revealed that screening tools reporting predictive validity showed a significantly lower sensitivity than that of tools reporting concurrent validity, which was also speculated in the previous review (13). One possible explanation lies in the diverse developmental trajectories of language development in the preschool years. Some of the children who perform poorly in early screening may recover spontaneously at a later time point, while some who appeared to be on the right track at the beginning may develop language difficulties later on (7). Current screening tools might not be able to capture this dynamic change in language development in the preschool years, resulting in lower predictive validity than expected. Hence, language disorder screening should concentrate on identifying or introducing new proxies or metrics that are sensitive to the dynamic nature of language development. Vocabulary *growth* estimates, for example, might be more sensitive to long-term outcomes than a single point estimation (150). Although the current review has shown that different proxies has been used in screening language disorder, there is a limited number of studies examining how proxies other than children's actual language ability perform in terms of predictive validity. It would be useful to investigate the interaction between the proxy used and the screening-diagnosis interval in future studies.

Age of screening was expected to be affected by the varying developmental trajectories. Screening at an earlier age might have lower accuracy than screening at a later age when language development becomes more stable. This expected difference was not found in the current meta-analysis. However, it is worth noting that screening tools used at different ages not only differed in the age of screening, but also other domains. In the meta-analysis, over half (55%, 16/29) of the screening under 4 relied on parent reports and used tools such as vocabulary checklists and reported utterances while none of the screening above 4 (0/8) were based on parent reports. Inquiry about the effect of screening age on screening accuracy is crucial as it has direct

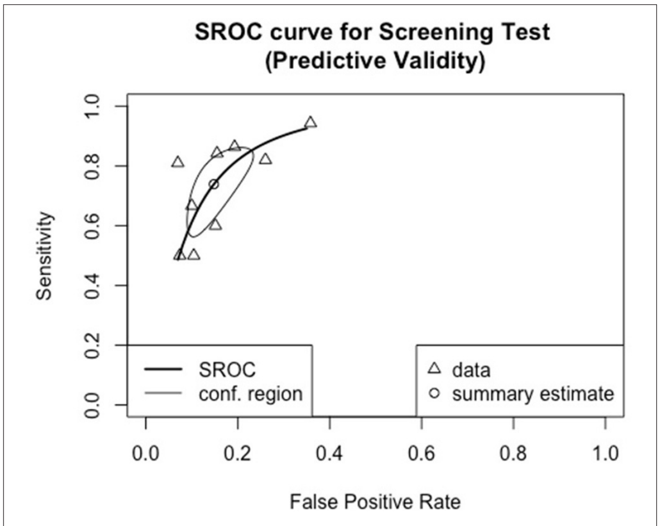


FIGURE 6 | Summary receiver operating characteristic curve for screening tools reporting predictive validity.

implication on the optimal time of screening. Future studies that compare the screening accuracy at different ages with the method of assessment being kept constant (e.g., using the same screening tool) may reveal a clearer picture.

Overall, only a small proportion of all the available screening tools achieved good accuracy in identifying both children with and without language disorder. Yet, there is still insufficient evidence to recommend any screening tool, especially given the presence of ROB in some studies. Besides, the limited number of valid tools may explain partly why screening for language disorder has not yet been adopted as a routine surveillance exercise in primary care, in that the use of any one type of screening tools may result in a considerable amount of over-identification and missing cases, which can lead to long term social consequences (19). As shown in the current review, in the future development of screening tools, the screening proxy should be carefully chosen in order to maximize test sensitivity. However, as tools that have good accuracy are limited, there remains room for discussion on whether future test development should aim at maximizing sensitivity even at the expense of specificity. The cost of over-identifying a false-positive child for a more in-depth assessment might be less than that of under-identifying a true-positive child and depriving the child of further follow-ups (104). If this is the case, the cut-off for test positivity can be adjusted. The more stringent the criteria used in screening, the higher the sensitivity the test yield but with the trade-off of a decrease in specificity. However, the decision should be made by fully acknowledging the harms and benefits, which has not been addressed in the current review. While an increase in sensitivity by adjusting the cut-off might lead to the benefit of better follow-ups, the accompanying increase in false positive rate might lead to the harms of stigmatization and unnecessary procedures. Given the highly variable developmental trajectories in asymptomatic children, another direction for future studies could be to evaluate

TABLE 7 | Bivariate meta-regression on studies-related factors on sensitivity and false-positive rate.

Factor	Transformed sensitivity				Transformed false positive rate			
	Coeff.	95% CI		*p*-value	Coeff.	95% CI		*p*-value
		LL	UL			LL	UL	
Types (L vs. Cm)	0.657	−0.055	1.370	0.070[#]	0.325	−0.774	1.423	0.562
Types (L vs. Mx)	−0.300	−0.855	0.255	0.290	0.435	−0.330	1.201	0.265
Types (Mx vs. Cm)	0.885	−0.244	2.015	0.124	−0.094	−0.958	0.770	0.832
Time (P vs. C)	−0.528	−1.018	−0.037	0.035*	−0.016	−0.726	0.695	0.965
Sc. AgeC (<4yo vs. ≥ 4*yo*)	1.676	−0.115	1.467	0.094[#]	0.560	−0.292	1.412	0.198
Sc. AgeP (<4yo vs. ≥ 4*yo*)	1.061	−1.115	3.238	0.339	0.663	−0.737	2.064	0.353
Informant (TP∧ vs. Pa)	−0.003	−0.525	0.519	0.992	−0.031	−0.836	0.773	0.939

First group in the bracket as the reference; L, language only; Cm, clinical markers; Mx, both language and clinical markers; P, predictive validity; C, concurrent validity; Pa, parent; TE, trained personnel; ScAgeC, Screening Age (for studies evaluating concurrent validity); ScAgeC, Screening Age (for studies evaluating predictive validity).
[#]*$p < 0.1$; *$p < 0.05$.*

TABLE 8 | Bivariate meta-regression of study-related factors on sensitivity and false-positive rate excluding high ROB studies.

Factor	Transformed sensitivity				Transformed false positive rate			
	Coeff.	95% CI		*p*-value	Coeff.	95% CI		*p*-value
		LL	UL			LL	UL	
Types (L∧ vs. Cm)	0.960	0.291	1.629	0.005**	−0.020	−1.295	1.256	0.976
Types (L∧ vs. Mx)	−0.173	−0.784	0.439	0.580	0.157	−0.753	1.067	0.735
Types (Mx∧ vs. Cm)[a]	-	-	-	-	-	-	-	-
Time (P∧ vs. C)	−0.819	−1.377	−0.262	0.004*	−0.104	−1.009	0.801	0.822
Sc. Age C(<4yo vs. ≥ 4*yo*)	0.234	−0.926	1.394	0.692	0.520	−0.388	1.428	0.262
Sc. Age P(<4yo vs. ≥ 4*yo*)[a]	-	-	-	-	-	-	-	-
Informant (TE∧ vs. Pa)	0.149	−0.514	0.812	0.660	0.160	−0.870	1.189	0.761

First group in the bracket as the reference; L, language only; Cm, clinical markers; Mx, both language and clinical markers; P, predictive validity; C, concurrent validity; Pa, parent; TE, trained examiner; ScAgeC, Screening Age (for studies evaluating concurrent validity); ScAgeC, Screening Age (for studies evaluating predictive validity).
[a]*Too few studies after exclusion for a valid analysis.*
[#]*$p < 0.1$. *$p < 0.05$.*

the viability of targeted screening in a higher-risk population and compare it with universal screening.

This is the first study to use meta-analytical techniques specifically to evaluate the heterogeneity in screening accuracy of tools for identifying children with language disorder. Nonetheless, there were several limitations of the study. One limitation was related to the variability and validity of the gold standard in that the reference standard tests. Different countries or regions use different localized standardized or non-standardized tools and criteria to define language disorder. There is no one consensual or true gold standard. More importantly, the significance and sensitivity and specificity of the procedures used to identify children with language disorders in those reference tests were not examined. Some reference tests may employ arbitrary cut-offs (e.g.,−1.25 SD) to define language disorders while some researchers advocate children's well-being as the outcome, such that when children's lives are negatively impacted by their language skills, they are considered as having language disorders (151). This lack of consensus might further explain the diverse results or lack of agreement in replication studies.

Another limitation of the study was that nearly all the included studies had at least some ROB. This was mainly due to many unreported aspects in the studies. It is suggested that future validation studies on screening tools should follow reporting guidelines such as STARD (152). A third limitation was that the rating of ROB only involved one rater, and more raters may minimize potential bias. Lastly, not all included screening tools were analyzed in the meta-analysis. Some studies evaluated multiple screening tools at a number of cut-offs or times of assessment. Only one data point per study was included in the meta-analysis and the data used in meta-analysis were chosen based on Youden's index. This selection would inevitably inflate the accuracy shown in the meta-analysis. With the emergence of new methods for meta-analysis for diagnostic studies, more sophisticated methods for handling this complexity of data structure may be employed in future reviews.

This review shows that current screening tools for developmental language disorder vary largely in accuracy, with only some achieving good accuracy. Meta-analytical data identified some sources for heterogeneity. Future development of screening tools should aim at improving overall screening accuracy by carefully choosing the proxy or designing items for screening. More importantly, metrics that are more sensitive to persistent language disorder should be sought. To fully inform surveillance for early language development, future research in the field can also consider broader aspects, such as the harms and benefits of screening as there is still a dearth of evidence in this respect.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found at: Reference lists of the article.

AUTHOR CONTRIBUTIONS

KS and CT conceived and designed the study, wrote the paper, conducted the format and tables, reviewed, and edited the manuscript. KS performed the statistical analysis. All authors have approved the final manuscript for submission.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fped.2022.801220/full#supplementary-material

REFERENCES

1. Bishop DV, Snowling MJ, Thompson PA, Greenhalgh T, Consortium C, Adams C, et al. Phase 2 of CATALISE: A multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *J Child Psychol Psychiatry*. (2017) 58:1068–80. doi: 10.1111/jcpp.12721
2. Norbury CF, Gooch D, Wray C, Baird G, Charman T, Simonoff E, et al. The impact of nonverbal ability on prevalence and clinical presentation of language disorder: evidence from a population study. *J Child Psychol Psychiatry*. (2016) 57:1247–57. doi: 10.1111/jcpp.12573
3. American Speech-Language-Hearing Association. *Preschool Language Disorders*. Available online at: https://www.asha.org/public/speech/disorders/preschool-languagedisorders/
4. Beitchman JH, Wilson B, Brownlie E, Walters H, Inglis A, Lancee W. Long-term consistency in speech/language profiles: II. Behavioral, emotional, and social outcomes. *J Am Acad Child Adolesc Psychiatry*. (1996) 35:815–25. doi: 10.1097/00004583-199606000-00022
5. Brownlie E, Bao, Bao L, Beitchman J. Childhood language disorder and social anxiety in early adulthood. *J Abnormal Child Psychol*. (2016) 44:1061–70. doi: 10.1007/s10802-015-0097-5
6. Beitchman JH, Wilson B, Brownlie EB, Walters H, Lancee W. Long-term consistency in speech/language profiles: I. Developmental and academic outcomes. *J Am Acad Child Adolesc Psychiatry*. (1996) 35:804–14. doi: 10.1097/00004583-199606000-00021
7. Zambrana IM, Pons F, Eadie P, Ystrom E. Trajectories of language delay from age 3 to 5: Persistence, recovery and late onset. *Int J Lang Commun Disord*. (2014) 49:304–16. doi: 10.1111/1460-6984.12073
8. Armstrong R, Scott JG, Whitehouse AJ, Copland DA, Mcmahon KL, Arnott W. Late talkers and later language outcomes: Predicting the different language trajectories. *Int J Speech-Lang Pathol*. (2017) 19:237–50. doi: 10.1080/17549507.2017.1296191
9. Bishop DV, Snowling MJ, Thompson PA, Greenhalgh T, Consortium C. CATALISE: A multinational and multidisciplinary Delphi consensus study. Identifying language impairments in children. *PLoS ONE*. (2016) 11:e0158753. doi: 10.1371/journal.pone.0158753
10. Sim F, Thompson L, Marryat L, Ramparsad N, Wilson P. Predictive validity of preschool screening tools for language and behavioural difficulties: A PRISMA systematic review. *PLoS ONE*. (2019) 14:e0211409. doi: 10.1371/journal.pone.0211409
11. Siu AL. Screening for speech and language delay and disorders in children aged 5 years or younger: US Preventive Services Task Force recommendation statement. *Pediatrics*. (2015) 136:e474–81. doi: 10.1542/peds.2015-1711
12. Wallace IF. *Universal Screening of Young Children for Developmental Disorders: Unpacking the Controversies*. Occasional Paper. RTI Press Publication OP-0048-1802. RTI International (2018). doi: 10.3768/rtipress.2018.op.0048.1802
13. Wallace IF, Berkman ND, Watson LR, Coyne-Beasley T, Wood CT, Cullen K, et al. Screening for speech and language delay in children 5 years old and younger: a systematic review. *Pediatrics*. (2015) 136:e448–62. doi: 10.1542/peds.2014-3889
14. Guiberson M. Telehealth measures screening for developmental language disorders in Spanish-speaking toddlers. *Telemed E-Health*. (2016) 22:739–45. doi: 10.1089/tmj.2015.0247
15. Puglisi ML, Blasi HF, Snowling MJ. Screening for the identification of oral language difficulties in Brazilian preschoolers: a validation study. *Lang Speech Hear Serv Schools*. (2020) 51:852–65. doi: 10.1044/2020_LSHSS-19-00083
16. Bornstein MH, Hahn CS, Putnick DL, Suwalsky JT. Stability of core language skill from early childhood to adolescence: A latent variable approach. *Child Dev*. (2014) 85:1346–56. doi: 10.1111/cdev.12192
17. Tomblin JB, Zhang X, Buckwalter P, O'Brien M. The stability of primary language disorder. *J Speech Lang Hear Res*. (2003) 46:1283–96. doi: 10.1044/1092-4388(2003/100)
18. Salameh J-P, Bossuyt PM, McGrath TA, Thombs BD, Hyde CJ, Macaskill P, et al. Preferred reporting items for systematic review and meta-analysis of diagnostic test accuracy studies (PRISMA-DTA): explanation, elaboration, and checklist. *BMJ*. (2020) 370:m2632. doi: 10.1136/bmj.m2632
19. Plante E, Vance R. Selection of preschool language tests: A data-based approach. *Lang Speech Hear Serv Schools*. (1994) 25:15–24. doi: 10.1044/0161-1461.2501.15
20. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Internal Med*. (2011) 155:529–36. doi: 10.7326/0003-4819-155-8-201110180-00009
21. McGuinness LA, Higgins JP. Risk-of-bias VISualization (robvis): An R package and Shiny web app for visualizing risk-of-bias assessments. *Res Synthesis Methods*. (2021) 12:55–61. doi: 10.1002/jrsm.1411
22. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol*. (2005) 58:982–90. doi: 10.1016/j.jclinepi.2005.02.022
23. Macaskill P, Gatsonis C, Deeks J, Harbord R, Takwoingi Y. *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy*. London: The Cochrane Collaboration (2010).
24. Doebler P, Holling H. *mada: Meta-Analysis of Diagnostic Accuracy*. R package version 0.5.10. (2020). Available online at: https://CRAN.R-project.org/package=mada
25. Allen DV, Bliss LS. Concurrent validity of two language screening tests. *J Commun Disord*. (1987) 20:305–17. doi: 10.1016/0021-9924(87)90012-8
26. Lee LL. *Northwestern Syntax Screening Test (NSST)*. Press Evanston, IL: Northwestern University Press Evanston (1971).
27. Hedrick DL, Prather EM, Tobin AR. *Sequenced Inventory of Communication Development*. Washington, DC: University of Washington Press Seattle (1984).
28. Blaxley L, et al. Two language screening tests compared with developmental sentence scoring. *Lang Speech Hear Serv Schools*. (1983) 14:38–46. doi: 10.1044/0161-1461.1401.38

29. Bankson NW. *Bankson Language Screening Test*. Baltimore, MD: University Park Press (1977).
30. Lee LL. *Developmental Sentence Analysis: A Grammatical Assessment Procedure for Speech and Language Clinicians*. Evanston, IL: Northwestern University Press (1974).
31. Burden V, Stott CM, Forge J, Goodyer I, Burden V, Stott CM, et al. The Cambridge Language and Speech Project (CLASP). I. Detection of language difficulties at 36 to 39 months. *Dev Med Child Neurol.* (1996) 38:613–31. doi: 10.1111/j.1469-8749.1996.tb12126.x
32. Alpern G, Boll T, Shearer M. Developmental profile II. *J Read.* (1980) 18:287–91.
33. Renfrew CE. Persistence of the open syllable in defective articulation. *J Speech Hear Disord.* (1966) 31:370–3. doi: 10.1044/jshd.3104.370
34. Renfrew C. *The Bus Story: A Test of Continuous Speech*. Oxford: England (1969).
35. Carscadden J, Corsiatto P, Ericson L, Illchuk R, Esopenko C, Sterner E, et al. A pilot study to evaluate a new early screening instrument for speech and language delays. *Canad J Speech-Lang Pathol Audiol.* (2010) 34:87–95.
36. Bzoch KR, League R, Brown V. *Receptive-Expressive Emergent Language Test*. Austin, TX: Pro-Ed (2003).
37. Chaffee CA, Cunningham CE, Secord-Gilbert M, Elbard H, Richards J. Screening effectiveness of the Minnesota child development inventory expressive and receptive language scales: sensitivity, specificity, and predictive value. *Psychol Assessment.* (1990) 2:80–5. doi: 10.1037/1040-3590.2.1.80
38. Reynell J. *Reynell Developmental Language Scales, Revised (Windsor: NFER)*. Windsor: NFER Publishing Company (1977).
39. Ireton HR, Thwing EJ. *Minnesota Child Development Inventory*. Minneapolis, MN: Behavior Science Systems (1972).
40. Dias DC, Rondon-Melo S, Molini-Avejonas DR. Sensitivity and specificity of a low-cost screening protocol for identifying children at risk for language disorders. *Clinics.* (2020) 75:e1426. doi: 10.6061/clinics/2020/e1426
41. American Speech-Language and Hearing Association. How does your child hear and talk? (2006). Available online at: http://www.asha.org/public/speech/development/chart/
42. Andrade CRF, Befi-Lopes DM, Fernandes FDM, Wertzner HF. *Teste de Language Infantil nas Áreas de Fonologia, Vocabulário, Fluência e Pragmática*. 2nd ed. Barueri: Pró-Fono (2011).
43. Dixon J, Kot A, Law J. Early language screening in City and Hackney: work in progress. *Child Care Health Dev.* (1988) 14:213–29. doi: 10.1111/j.1365-2214.1988.tb00576.x
44. Reynell J, Huntley M. *Reynell Developmental Language Scales, 2nd revision*. Windsor: NFER-Nelson (1985).
45. Lowe M, Costello A. *The Symbolic Play Test*. Windsor: NFER-Nelson (1976).
46. Gray S, Plante E, Vance R, Henrichsen M. The diagnostic accuracy of four vocabulary tests administered to preschool-age children. *Lang Speech Hear Serv Schools.* (1999) 30:196–206. doi: 10.1044/0161-1461.3002.196
47. Gardner MF. *EOWPVT-R: Expressive One-Word Picture Vocabulary Test, Revised*. Novato, CA: Academic Therapy Publications (1990).
48. Dunn LM, Dunn LM. *Peabody Picture Vocabulary Test: PPVT-IIIB*. Circle Pines, MN: American Guidance Service Circle Pines (1997). doi: 10.1037/t15145-000
49. Gardner MF. *Receptive One-Word Picture Vocabulary Test*. Novato, CA: Academic Therapy Publications (1985).
50. Williams K. *Expressive Vocabulary Test (EVT)*. Circle Pines, MN: American Guidance Service Inc. (1997).
51. Zimmerman IL, Steiner VG. *Preschool Language Scale 4th edition, Spanish*. Antonio, TX: The Psychological Corporation (2004).
52. Guiberson M, Rodriguez BL. Measurement properties and classification accuracy of two Spanish parent surveys of language development for preschool-age children. *Am J Speech-Lang Pathol.* (2010) 19:225–37. doi: 10.1044/1058-0360(2010/09-0058)
53. Guiberson M. Concurrent validity of a parent survey measuring communication skills of Spanish speaking preschoolers with and without delayed language. *Perspect Commun Disord Sci.* (2008) 15:73–81. doi: 10.1044/cds15.3.73
54. Squires J, Potter l, Bricker D. *The ASQ User's Guide (2nd ed.)*. Baltimore, MD: Paul H. Brookes (1999).
55. Guiberson M, Rodríguez BL, Dale PS. Classification accuracy of brief parent report measures of language development in Spanish-speaking toddlers. *Lang Speech Hear Serv Sch.* (2011) 42:536–49. doi: 10.1044/0161-1461(2011/10-0076)
56. Squires J, Bricker D, Mounts L, Potter L, Nickel R, Twombly E, et al. *Ages and Stages Questionnaire*. Baltimore, MD: Paul H. Baltimore, MD: Brookes (1999).
57. Jackson-Maldonado D, Thal DJ, Fenson L. MacArthur *Inventarios del Desarrollo de Habilidades Comunicativas: User's Guide and Technical Manual*. Baltimore, MD: Brookes Pub (2003).
58. Guiberson M, Rodriguez BL, Zajacova A. Accuracy of telehealth-administered measures to screen language in Spanish-speaking preschoolers. *Telemed E-Health.* (2015) 21:714–20. doi: 10.1089/tmj.2014.0190
59. Heilmann J, Weismer SE, Evans J, Hollar C. Utility of the MacArthur-bates communicative development inventory in identifying language abilities of late-talking and typically developing toddlers. *Am J Speech-Lang Pathol.* (2005) 14:40–51. doi: 10.1044/1058-0360(2005/006)
60. Board CA. *The MacArthur-Bates Communicative Development Inventory: Words and Sentences*. Baltimore, MD: Paul H Brookes (1992).
61. Zimmerman I, Steiner V, Pond R. *The Preschool Language Scale-3*. Columbus: Merrill (1992).
62. Klee T, Carson DK, Gavin WJ, Hall L, Kent A, Reece S. Concurrent and predictive validity of an early language screening program. *J Speech Lang Hear Res.* (1998) 41:627–41. doi: 10.1044/jslhr.4103.627
63. Rescorla L. The language development survey: A screening tool for delayed language in toddlers. *J Speech Hear Disord.* (1989) 54:587–99. doi: 10.1044/jshd.5404.587
64. Mullen EM. *Infant MSEL Manual: Infant Mullen Scales of Early Learning*. Cranston, RI: TOTAL Child (1989).
65. Klee T, Pearce K, Carson DK. Improving the positive predictive value of screening for developmental language disorder. *J Speech Lang Hear Res.* (2000) 43:821–33. doi: 10.1044/jslhr.4304.821
66. Laing GJ, Law J, Levin A, Logan S. Evaluation of a structured test and a parent led method for screening for speech and language problems: prospective population based study. *BMJ.* (2002) 325:1152. doi: 10.1136/bmj.325.7373.1152
67. Edwards S, Fletcher P, Garman M, Hughes A, Letts C, Sinka I. *The Reynell Developmental Language Scales III, NFER*. NELSON Publishing, The University of Reading Edition (1997).
68. Law J. Early language screening in City and Hackney: The concurrent validity of a measure designed for use with 2½-year-olds. *Child.* (1994) 20:295–308. doi: 10.1111/j.1365-2214.1994.tb00392.x
69. Levett L, Muir J. Which three year olds need speech therapy? Uses of the Levett-Muir language screening test. *Health Visitor.* (1983) 56:454–6.
70. Reynell J, Reynell Huntley M. *Reynell Developmental Language Scales (Revised)*. Windsor: NFER Publishing Company Ltd. (1977).
71. Goldman R. *Goldman-Fristoe Test of Articulation*. Circle Pines, MI: American Guidance Service. Inc. (1969).
72. Crystal D, Fletcher P, Garman M. *The Grammatical Analysis of Language Disability: A Procedure for Assessment and Remediation*. Hodder Education (1976).
73. Visser-Bochane MI, van der Schans CP, Krijnen WP, Reijneveld SA, Luinge MR. Validation of the early language scale. *Eur J Pediatrics.* (2020) 180:63–71. doi: 10.1007/s00431-020-03702-8
74. Schlichting JEPT, Spelberg L. *Lexilijst Begrip [Lexilist Comprehension]*. Amsterdam: Harcourt Test Publishers (2007).
75. Schlichting L, Spelberg HL. *Schlichting test voor taalbegrip [Schlichting test for language comprehension]*. Houten: Bohn Stafleu van Loghum (2010).
76. Schlichting JEPT. Lexilijst Nederlands. (2002). doi: 10.1240/sav_gbm_2002_h_000262
77. Schlichting JEPT, Spelberg HC. *Schlichting Test voor Taalproductie-II: voor Nederland en Vlaanderen*. Houten: Bohn Stafleu van Loghum (2012). doi: 10.1007/978-90-313-9842-3
78. Slofstra-Bremer C, van der Meulen S, Lutje Spelberg H. *De Taalstandaard [The Language standard]*. Amsterdam: Pearson (2006).

79. Geurts H. *Handleiding CCC-2-NL*. Amsterdam: Harcourt Test Publishers (2004).
80. Visser-Bochane M, Luinge M, Dieleman L, van der Schans C, Reijneveld S. The Dutch well child language screening protocol for 2-year-old children was valid for detecting current and later language problems. *Acta Paediatrica.* (2020) 110:2825–32. doi: 10.1111/apa.15447
81. Mattsson CM, Mårild S, Pehrsson NG. Evaluation of a language-screening programme for 2.5-year-olds at Child Health Centres in Sweden. *Acta Paediatr.* (2001) 90:339–44. doi: 10.1080/080352501300067776
82. McGinty C. An investigation into aspects of the Mayo early language screening test. *Child Care Health Dev.* (2000) 26:111–28. doi: 10.1046/j.1365-2214.2000.00176.x
83. Garvey-Cecchetti B, Heslin C, Laundon O, McGinty C, O'Malley L, Dowd P, et al. *The Mayo Early Language Screening Test.* Western Health Board: Mayo Speech and Language Therapy Department (1993).
84. Anthony A, Bogle D, Ingram T, McIsaac M. *1971: Edinburgh Articulation Test*. Edinburgh: Churchill Livingstone (1971).
85. Nair MKC, Harikumaran Nair GS, Mini AO, Indulekha S, Letha S, Russell PS. Development and validation of language evaluation scale Trivandrum for children aged 0-3 years - LEST (0-3). *Indian Pediatrics.* (2013) 50:463–7. doi: 10.1007/s13312-013-0154-5
86. Bzoch K, League R. *Receptive-Expressive Emergent Language Scale*. Gainesville, FL: Tree of Life Press. Inc. (1971).
87. Nayeb L, Lagerbere D, Westerlund M, Sarkadi A, Lucas S, Eriksson M. Modifying a language screening tool for three-year-old children identified severe language disorders six months earlier. *Acta Paediatrica.* (2019) 108:1642–8. doi: 10.1111/apa.14790
88. Andrade Cd, Befi-Lopes DM, Fernandes FDM, Wertzner HF. *ABFW: teste de linguagem infantil nas áreas de fonologia, vocabulário, fluência e pragmática*. São Paulo: Pró-Fono. (2004).
89. Bischop D. *The Test for Reception of Grammar, Version 2 (TROG-2).* Oxford: Pearson (2009).
90. Santos FH, Bueno OFA. Validation of the Brazilian Children's Test of Pseudoword Repetition in Portuguese speakers aged 4 to 10 years. *Brazil J Med Biol Res.* (2003) 36:1533–47. doi: 10.1590/S0100-879X2003001100012
91. Rescorla L, Alley A. Validation of the language development survey (LDS): A parent report tool for identifying language delay in toddlers. *J Speech Lang Hear Res.* (2001) 44:434–45. doi: 10.1044/1092-4388(2001/035)
92. Sachse S, Von Suchodoletz W. Early identification of language delay by direct language assessment or parent report? *J Dev Behav Pediatrics.* (2008) 29:34–41. doi: 10.1097/DBP.0b013e318146902a
93. Grimm H, Doil H. *Elternfragebögen für die Früherkennung von Risikokindern*. ELFRA: Hogrefe, Verlag für Psychologie (2000).
94. Grimm H. *Sprachentwicklungstest für sweijährige slindes-SETK-2 und für dreibis fünfjährige Kinder-SETK 3-5*. Göttingen: Hogrefe (2000).
95. Stokes SF. Secondary prevention of paediatric language disability: a comparison of parents and nurses as screening agents. *Eur J Disord Commun.* (1997) 32:139–58. doi: 10.1111/j.1460-6984.1997.tb01628.x
96. van Agt HM, van der Stege HA, de Ridder-Sluiter JG, de Koning HJ. Detecting language problems: accuracy of five language screening instruments in preschool children. *Dev Med Child Neurol.* (2007) 49:117–22. doi: 10.1111/j.1469-8749.2007.00117.x
97. Stott CM, Merricks MJ, Bolton PF, Goodyer IM. Screening for speech and language disorders: The reliability, validity and accuracy of the General Language Screen. *Int J Lang Commun Disord.* (2002) 37:133–51. doi: 10.1080/13682820110116785
98. Brouwers-de Jong E, Burgmeijer R, Laurent de Angulo M. Ontwikkelingsonderzoek op het consultatiebureau: handboek bij het vernieuwde Van Wiechenonderzoek. (1996).
99. Walker D, Gugenheim S, Downs MP, Northern JL. Early Language Milestone Scale and language screening of young children. *Pediatrics.* (1989) 83:284–8. doi: 10.1542/peds.83.2.284
100. Coplan J. *ELM Scale: The Early Language Milestone Scale*. Pro-Ed (1983).
101. Wetherby AM, Goldstein H, Cleary J, Allen L, Kublin K. Early identification of children with communication disorders: Concurrent and predictive validity of the CSBS Developmental Profile. *Infants Young Child.* (2003) 16:161–74. doi: 10.1097/00001163-200304000-00008
102. Wetherby AM, Prizant BM. *Communication and Symbolic Behavior Scales: Developmental Profile*. Baltimore, MD: Paul H Brookes Publishing Co. (2002). doi: 10.1037/t11529-000
103. Kapalkova S, Polisenska K, Vicenova Z. Non-word repetition performance in Slovak-speaking children with and without SLI: novel scoring methods. *Int J Lang Commun Disord.* (2013) 48:78–89. doi: 10.1111/j.1460-6984.2012.00189.x
104. Nash H, Leavett R, Childs H. Evaluating the GAPS test as a screener for language impairment in young children. *Int J Lang Commun Disord.* (2011) 46:675–85. doi: 10.1111/j.1460-6984.2011.00038.x
105. Van der Lely H, Gardner H, McClelland AGR, Froud KE. *Grammar and Phonology Screening Test: (GAPS)* London: DLDCN (2007).
106. Wiig E, Secord W, Semel E. *Clinical Evaluation of Language Fundamentals-Preschool*. Second *UK edition*. London: Harcourt Assessment (2006).
107. Sturner RA, Funk SG, Green JA. Preschool speech and language screening: further validation of the sentence repetition screening test. *J Dev Behav Pediatr.* (1996) 17:405–13. doi: 10.1097/00004703-199612000-00006
108. Sturner R, Kunze L, Funk S, Green J. Elicited imitation: its effectiveness for speech and language screening. *Dev Med Child Neurol.* (1993) 35:715–26. doi: 10.1111/j.1469-8749.1993.tb11717.x
109. Kirk SA, Kirk WD, McCarthy JJ. *Illinois Test of Psycholinguistic Abilities*. Champaign, IL: University of Illinois press (1968).
110. van der Lely HKJ, Payne E, McClelland A. An investigation to validate the grammar and phonology screening (GAPS) test to identify children with specific language impairment. *PLoS ONE.* (2011) 6:e022432. doi: 10.1371/journal.pone.0022432
111. Fluharty NB. The design and standardization of a speech and language screening test for use with preschool children. *J Speech Hear Disord.* (1974) 39:75–88. doi: 10.1044/jshd.3901.75
112. Hedrick DL, Prather EM, Tobin AR. *Sequenced Inventory of Communication Development*. Seattle, WA: University of Washington Press (1975).
113. Benavides AA, Kapantzoglou M, Murata C. Two grammatical tasks for screening language abilities in Spanish-speaking children. *Am J Speech-Lang Pathol.* (2018) 27:690–705. doi: 10.1044/2017_AJSLP-17-0052
114. Zimmerman IL, Steiner VG, Pond RE. *Preschool Language Scale-Fifth Edition Spanish Screening Test (PLS-5 Spanish Screening Test). [Measurement assessment]*. San Antonio, TX: Pearson (2011) doi: 10.1037/t15141-000
115. Fluharty NB. *Fluharty Preschool Speech and Language Screening Test: Teaching Resources*. Austin, TX: Pro-Ed (1978).
116. Lee LL, Koenigsknecht RA, Mulhern ST. *Developmental Sentence Scoring*. Evanston, IL: North Western University (1974).
117. Bliss LS, Allen DV. *Screening Kit of Language Development*. Baltimore, MD: University Park Press (1983).
118. Bliss LS, Allen DV. Screening kit of language development: A preschool language screening instrument. *J Commun Disord.* (1984) 17:133–41. doi: 10.1016/0021-9924(84)90019-4
119. Lavesson A, Lovden M, Hansson K. Development of a language screening instrument for Swedish 4-year-olds. *Int J Lang Commun Disord.* (2018) 53:605–14. doi: 10.1111/1460-6984.12374
120. Matov J, Mensah F, Cook F, Reilly S. Investigation of the language tasks to include in a short-language measure for children in the early school years. *Int J Lang Commun Disord.* (2018) 53:735–47. doi: 10.1111/1460-6984.12378
121. Matov J, Mensah F, Cook F, Reilly S, Dowell R. The development and validation of the Short Language Measure (SLaM): A brief measure of general language ability for children in their first year at school. *Int J Lang Commun Disord.* (2020) 55:345–58. doi: 10.1111/1460-6984.12522
122. Semel E, Wiig EH, Secord W. *Clinical Evaluation of Language Fundamentals-Fourth Edition, Australian Standardised Edition*. Sydney, NSW: PsychCorp (2006).
123. Wright R, Levin B. A preschool articulation and language screening for the identification of speech disorders. *Final Rep.* (1971).
124. Eisenberg SL, Guo L-Y. Differentiating children with and without language impairment based on grammaticality. *Lang Speech Hear Serv Schools.* (2013) 44:20–31. doi: 10.1044/0161-1461(2012/11-0089)

125. Dawson JI, Stout C, Eyer J, Tattersall PJ, Fonkalsrud J, Croley K. *SPELT-P 2: Structured Photographic Expressive Language Test*. Preschool: Janelle Publications (2005).
126. Eisenberg SL, Guo L, Germezi M. How grammatical are three-year-olds? *Lang Speech Hear Serv Schools.* (2012) 43:36–52. doi: 10.1044/0161-1461(2011/10-0093)
127. Bruce B, Kornfalt R, Radeborg K, Hansson K, Nettelbladt U. Identifying children at risk for language impairment: screening of communication at 18 months. *Acta Paediatrica.* (2003) 92:1090–5. doi: 10.1080/08035250310004414
128. Holmberg E, Nelli SB. *Neurolingvistisk undersökningsmodell fr språkstörda barn.* Utbildningsproduktion AB (kommer inom kort att ges ut i ny, något omarbetad upplaga av Pedagogisk Design) (1986).
129. Bishop D. *TROG svensk manual [svensk översättning och bearbetning: Eva Holmberg och Eva Lundälv]*. Göteborg: SIH Läromedel (Originalarbete publicerat 1983). (1998).
130. Frisk V, Montgomery L, Boychyn E, Young R, Vanryn E, McLachlan D, et al. Why screening canadian preschoolers for language delays is more difficult than it should be. *Infants Young Child.* (2009) 22:290–308. doi: 10.1097/IYC.0b013e3181bc4db6
131. Harrison PL. *AGS Early Screening Profiles*. Circle Pines, MN: American Guidance Service (1990).
132. Zimmerman I, Steiner V, Pond R. *Preschool Language Scale-Fourth Edition (PLS-4).* San Antonio, TX: The Psychological Corporation (2002) doi: 10.1037/t15140-000
133. Newborg J, Stock J, Wnek L, Guidabaldi J, Svinicki J. *Battelle Developmental Inventory*. Itasca: Riverside (1988).
134. Bracken BA. *Bracken Basic Concept Scale-Revised*. San Antonio, TX: Psychological Corporation (1998).
135. Gascoe FP. *Technical Report for Brigance Screens*. Cheltenham, VIC: Hawker Brownlow Education (1998).
136. Jessup B, Ward E, Cahill L, Keating D. Teacher identification of speech and language impairment in kindergarten students using the Kindergarten Development Check. *Int J Speech-Lang Pathol.* (2008) 10:449–59. doi: 10.1080/17549500802056151
137. Office for Educational Review. *Revised Kindergarten Development Check*. Hobart: Office for Educational Review (2003).
138. Pesco D, O'Neill DK. Predicting later language outcomes from the language use inventory. *J Speech Lang Hear Res.* (2012) 55:421–34. doi: 10.1044/1092-4388(2011/10-0273)
139. O'Neill D. *Language Use Inventory: An Assessment of Young Children's Pragmatic Language Development for 18-to 47-Month-Old Children*. Waterloo, ON: Knowledge in Development (2009).
140. Seymour HN, Roeper T, De Villiers JG, De Villiers PA. *Diagnostic Evaluation of Language Variation, Norm Referenced*. Pearson: DELV (2005).
141. Wiig E, Secord W, Semel E. *Clinical Evaluation of Language Fundamentals (Preschool 2nd edition ed.)*. San Antonio, TX: The Psychological Corporation: Harcourt Assessment Inc. (2004).
142. Bishop DV. *CCC-2: Children's Communication Checklist-2*. Pearson (2006).
143. Westerlund M, Berglund E, Eriksson M. 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.* (2006) 49:237–47. doi: 10.1044/1092-4388(2006/020)
144. Berglund E, Eriksson M. Communicative development in Swedish children 16-28 months old: The Swedish early communicative development inventory-words and sentences. *Scand J Psychol.* (2000) 41:133–44. doi: 10.1111/1467-9450.00181
145. Eriksson M, Berglund E. Swedish early communicative development inventories: Words and gestures. First *Lang.* (1999) 19:55–90. doi: 10.1177/014272379901905503
146. Westerlund M, Sundelin C. Can severe language disability be identified in three-year-olds? Evaluation of a routine screening procedure. *Acta Paediatrica.* (2000) 89:94–100. doi: 10.1111/j.1651-2227.2000.tb01195.x
147. Westerlund M, Sundelin C. Screening for developmental language disability in 3-year-old children. Experiences from a field study in a Swedish municipality. *Child.* (2000) 26:91–110. doi: 10.1046/j.1365-2214.2000.00171.x
148. Mullen EM. *Mullen Scales of Early Learning*. Circle Pines, MN: AGS (1995).
149. Ahufinger N, Berglund-Barraza A, Cruz-Santos A, Ferinu L, Andreu L, Sanz-Torrent M, et al. Consistency of a nonword repetition task to discriminate children with and without developmental language disorder in Catalan-Spanish and European Portuguese speaking children. *Children.* (2021) 8:85. doi: 10.3390/children8020085
150. Rowe ML, Raudenbush SW, Goldin-Meadow S. The pace of vocabulary growth helps predict later vocabulary skill. *Child Dev.* (2012) 83:508–25. doi: 10.1111/j.1467-8624.2011.01710.x
151. Nippold MA, Tomblin JB. *Understanding Individual Differences in Language Development Across the School Years*. New York, NY: Psychology Press (2014). doi: 10.4324/9781315796987
152. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *Clin Chem.* (2015) 61:1446–52. doi: 10.1373/clinchem.2015.246280

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.

frontiers in Pediatrics

ORIGINAL RESEARCH
published: 08 March 2022
doi: 10.3389/fped.2022.814415

Identifying Language Disorder Within a Migration Context: Development and Performance of a Pre-school Screening Tool for Children With German as a Second Language

Daniel Holzinger[1,2,3], Christoph Weber[2,4] and Magdalena Jezek[1,2]*

[1] Institute of Neurology of Senses and Language, Hospital of St. John of God, Linz, Austria, [2] Research Institute for Developmental Medicine, Johannes Kepler University, Linz, Austria, [3] Institute of Linguistics, University of Graz, Graz, Austria, [4] Department for Inclusive Education, University of Education Upper Austria, Linz, Austria

OPEN ACCESS

Edited by:
Sheena Reilly,
Griffith University, Australia

Reviewed by:
Carol Kit Sum To,
The University of Hong Kong, Hong Kong SAR, China
Annette Fox-Boyer,
University of Lübeck, Germany

****Correspondence:***
Daniel Holzinger
daniel.holzinger@jku.at

Specialty section:
This article was submitted to Children and Health, a section of the journal Frontiers in Pediatrics

Received: *13 November 2021*
Accepted: *04 February 2022*
Published: *08 March 2022*

Citation:
Holzinger D, Weber C and Jezek M (2022) Identifying Language Disorder Within a Migration Context: Development and Performance of a Pre-school Screening Tool for Children With German as a Second Language. Front. Pediatr. 10:814415. doi: 10.3389/fped.2022.814415

Background: There is a lack of accurate and practicable instruments for identifying language disorders in multilingual children in pre-school settings.

Objective: To develop a language screening instrument for pre-school children who are growing up with German as their second language.

Design: After the development and initial validation of a language screening tool, the new instrument (LOGiK-S) was administered to three cohorts of children (2014, 2015, 2017) with a non-German first language attending a variety of public pre-schools in Upper Austria. The screening instrument measures expressive and receptive grammatical skills in German. The final validation study included the results for 270 children for the screening measure and reference tests. A combination of a standardized comprehensive language test of grammatical skills developed for children acquiring German as a second language and a test of expressive vocabulary with the use of specific cutoffs for second language learners was applied as the gold standard for identifying language disorders.

Results: The LOGiK-S screening of expressive grammar demonstrated excellent accuracy (AUC.953). The screening subscale of receptive grammar did not improve the prediction of language disorders. Using an optimized cutoff yielded a fail rate of 17%, excellent sensitivity (0.940), and specificity (0.936). Time economy and acceptance of the screening by children and screeners were mostly rated as high.

Conclusion: The LOGiK-S language screening instrument assessing expressive German grammar development using bilingual norms is a valid and feasible instrument for the identification of language disorders in second language learners of German at the pre-school age.

Keywords: language screening, multilingual, migration, pre-school, German as a second language

INTRODUCTION

With prevalence rates of about 10%, language disorders (LDs) can be considered the most frequent developmental problem in children under the age of 7 (1–4). Prevalence estimations vary because of the lack of a generally accepted definition of LD. The term developmental language disorder (DLD) was endorsed in a consensus document by Bishop et al. (5) as referring to language difficulties characterized by a lack of known biomedical etiology, functional impairment, and poor prognosis. Therefore, LD remains a diagnosis to be made by experienced clinicians able to assess different dimensions of language, the degree of impairment caused by the language difficulties, and the probability of persistence. A population study on LD in England by Norbury et al. (6) resulted in a prevalence of DLD of 7.58%. In addition, 2.34% of LDs were associated with an intellectual disability or a medical diagnosis, adding up to about 10% of LDs in total. The authors classified a child as language disordered when language performance was at least 1.5 standard deviations below the norm on at least two of five language domains. Other researchers (1, 4, 7) defined a specific LD by scores of at least −1.25 standard deviations in at least two language domains. Problems often associated with pre-school LD include increased rates of behavioral, social, and emotional difficulties (8, 9), poor academic outcomes (10), and higher risk of unemployment (11).

The prevalence of LD is expected to be the same in children growing up monolingually or multi-lingually. Multilingual children growing up in an environment with a sufficient quantity and quality of language input are no more likely to develop LD than their monolingual peers (12).

Previous research has highlighted the effectiveness of early parent-facilitated and child-directed language intervention (13–16). As a consequence, suitable and practical screening instruments are needed for the early identification of language difficulties. As the population of young children growing up bilingually grows in Europe increases, there is a pressing need for reliable measures identifying what is typical or not in their language development.

In Upper Austria, the context of this study, the proportion of children with a first language (L1) other than German has been increasing continuously in recent years. For example, the share a of children from non-German speaking countries increased in primary education within 5 years from 16% (2012–2013) to 20% (2017–2018). In 2018–2019, one out of four children attending a pre-school had a L1 other than German. Notably, this figure is much higher in urbanized areas (17). In Upper Austria, first languages are predominantly. Bosnian/Croatian/Serbian (30%) and Turkish (20%). The remaining languages include languages such as Romanian or Arabic. In Austria, public pre-schools, with German being almost exclusively the only language of instruction, can be attended from the age of 3 up to the age of 6.

There are a number of challenges involved in the development and validation of language screening tools for children who grow up in a bilingual context (18). First, the group of bilingual children is extremely heterogeneous in relation to the length of exposure to the second language, quality and quantity of input in both languages in their families and institutional settings (e.g., pre-schools), or the family's socioeconomic status and parental education level. Second, in many cases, no instruments are available for assessing children's linguistic skills in their first language (19). When instruments are available, the examiners are faced with the problem of being unfamiliar with the diversity of first languages of the children to be screened. Third, tests developed for a particular language targeting monolingual children do not apply equally to bilingual children using this language as their L1 outside their home country. In a migration context language is in a state of constant change due to contact phenomena and does not necessarily overlap in all linguistic aspects with "the same" language in a non-migration context (20, 21). In addition, L1 attrition phenomena have been described in situations with early acquisition of an L2 and a literacy acquisition restricted to the second language (22, 23). Fourth, different profiles of language difficulties in children with LD (e.g., morpho-syntactic, semantic, phonological) complicate the time-efficient and reliable identification of increased risk of LD (24).

The systematic review by Sim et al. (9) compared pre-school screening tools. It concluded that language screening instruments could improve the rate of early identification of developmental language difficulties if incorporated into routine child-health surveillance. Therefore, a nationwide language screening program including specific instruments for pre-school children growing up bilingually is essential, especially as high percentages of children with developmental difficulties are not being detected prior to school entry (9).

As a consequence of the complexity of language screening in a multilingual context, a variety of approaches have been explored. Although generally claimed, the assessment of the L1 is usually not feasible. Another option is the use of instruments to assess the acquisition of the majority language by use of bilingual norms with specific cutoffs (25). For the acquisition of German as a second language, the LiSe-DaZ [Linguistische Sprachstandserhebung Deutsch als Zweitsprache (26)] is the only available standardized language test that provides specific norms for German as an L2 taking the length of German language exposure into account. However, the LiSe-DaZ is a comprehensive language assessment rather than an instrument that can be applied for universal screening. Finally, tools constructed according to linguistic principles that can be applied across individual languages (e.g., non-word and sentence repetition) have been proposed and shown to be useful for the identification of children with increased risk of LD in bilingual contexts (27).

The aim of the present research was to develop and evaluate a screening instrument for the identification of LD in pre-school children learning German as their second language in terms of screening accuracy and feasibility within a community pre-school setting in Austria. The new screening tool assesses the acquisition of German grammar. We report the results of two studies. Study 1 was a pilot study focusing on the screening development and initial validation of the screening instrument. The aim of Study 2 was the final validation of the screening instrument by the additional use of a comprehensive reference test developed for learners of German as an L2.

STUDY 1 (PILOT STUDY)

Methods

Participants

In 2012, all children growing up with a language other than German attending 1 of 13 public pre-schools well-distributed over the central and less urbanized areas of Upper Austria were invited to participate in the pilot study (Study 1). After the exclusion of children with German as their dominant language and those with a length of German language exposure below 1 year [following (28, 29)], the final sample consisted of 112 children (49.1% girls) with a mean age of 57.4 months (SD = 4 months) and a mean length of exposure to German of 18.9 months (SD = 5.7 months). Note that the length of exposure is limited as children can be enrolled in pre-school at the earliest at the age of 3 years and the study focuses on children in their penultimate pre-school year. The most frequent first languages spoken by the participants were Bosnian/Croatian/Serbian (29.5%), Turkish (15.2%), Albanian (9.8%), Czech (7.1%), Arabic (7.1%), and Romanian (6.3%).

Procedures

The screening procedures were carried out by clinical linguists from the Institute of Neurology of Senses and Language and by trained students of speech-language therapy from the University for Health Professions (Fachhochschule für Gesundheitsberufe) in Linz. Before the direct screening of a child, the examiners completed a structured interview with the parents on sociodemographic factors, language use in the family, the child's dominant language(s), time of exposure to German, and pre-school attendance. After the language screening, the results were reported to the parents and the pre-school teachers. Within a maximum of 90 days, the children were tested again using standardized reference tests. The tests were administered by language experts from the Institute of Neurology of Senses and Language who were blinded to the screening results.

Screening Measures

As LD in German, whether acquired as first or second language, manifests itself at pre-school age particularly in morphosyntax, such as subject-verb-agreement (30), verbal inflection (31), and elimination of function words (19), LOGiK-S was used to assess the following grammatical dimensions and structures:

i *Expressive grammar* (EG) was assessed by sentence completion supported by illustrations and included verb position, verb inflection, subordinate clauses, perfect forms, determiners, comparatives, noun plurals, prepositions, questions (open and closed, wh-questions), and passive structures.

ii *Receptive grammar* (RG) includes the comprehension of morpho-syntactic structures, such as intransitive clauses, prepositional phrases, coordination, pronouns, and embedded and subordinate clauses. Comprehension of the grammatical structures was assessed by having the children point at the appropriate illustration from a selection of four.

In the pilot study, the screening of RG included 20 items, and the EG subscales comprised 27 items. After exclusion of items with very low and high difficulty and low items-scale correlations, and considering the input of a group of screeners involved in the pilot study, a set of 10 items for the RG subscale and a set of 17 items for the EG subscale were used. The EG subscale showed good reliability (Cronbach's $\alpha = 0.82$). In constrast, the reliability of the RG subscale was relatively poor (Cronbach's $\alpha = 0.61$).

Reference Tests

Following other studies on LDs (1, 4, 7), a child was classified as having an LD when performance in the second language was below −1.25 SDs in at least two language domains, applying bilingual norms, and when the experienced clinicians performing the diagnosis had identified serious indications of LD in the L1 from parent interviews. This goldstandard used was the best available at the time of planning the study. We used three standardized tests to assess EG, RG, and expressive vocabulary.

(1) EG skills were assessed by the plural and case marking subtests of the PDSS [(32) Patholinguistische Diagnostik bei Sprachentwicklungsstörungen] as well as the subtests for comparatives and perfect tense of the ETS 4-8 [(33); Entwicklungstest Sprache für Kinder von 4 bis 8 Jahren]. The manuals only provides *t*-values for monolingual German-speaking children. However, relying on these *t*-values would have resulted in high rates of children with atypical results (t-values ≤ 37.5) for the four subscales (between 50 and 70%). Therefore, we used principal component analysis (PCA) to extract a composite score based on all the subscales. The PCA yielded one component with an eigenvalue of 3.2 (80% explained variance). The loadings ranged from 0.88 to 0.92. The internal consistency (Cronbach's α) was high at 0.90. We saved the component score (i.e., z-score with M = 0 and SD = 1). Children were classified as atypical in EG if they scored in the bottom 10% (1.25 SDs) of the component score.

(2) The TROG-D [German version of the Test the Reception of Grammar (34)] assesses the understanding of German grammar. Similar to the PDSS and ETS 4–8, the TROG-D only provides norm values for German-speaking monolingual children. Applying these norms to German language learners would again result in high rates (55%) of children with atypical results ($t \leq 37.5$). Therefore, we again used the sample percentiles to identify the bottom 10% of the TROG-D scores.

(3) The AWST-R [Revised Active Vocabulary Test for 3–5 year-old children, Aktiver Wortschatztest für 3- bis 5-Jährige, Revision (35)] is a standardized picture-naming test for the age range from 3.0 to 5.5 years. The items are ordered by increasing difficulty. To reduce the length of the assessment, we only used the first of the two picture folders (35 items) for the assessment of expressive vocabulary. As the AWST-R again lacks norm values for the reduced version of 35 items, we again estimated norm values based on the study data. However, because the AWST-R was applied in Study 1 and Study 2, we used pooled data from both studies to estimate norm values. The samples were pooled to achieve a larger ($n = 400$) and more representative database for calculating

norm values. In short, we applied a continuous norming approach using three age groups (48–50, 51–56, and 57–62 months). Continuous norming was conducted using the Cnormj package (36) in jamovi 1.6 (37).

A teacher questionnaire was used to collect child sociodemographic information, length of pre-school attendance and the teacher's assessment of the children's German language level as compared to their peers learning German as a second language.

Following our definition of atypical scores (≤1.25 SD), in at least two of the reference tests, 11 children (9.8%) were classified as LD in the pilot study. Notably, pre-school teachers estimated the language development of eight children classified as LD to be significantly worse than that of their peers (2 children's language development was estimated as slightly worse; $\chi^2_{(2)} = 18.480$, $p < 0.001$, Cramers V = 0.412). The LiSe-DaZ [Linguistic Language Assessment—German as a Second Language (26)] used as reference test in Study 2 was not available when Study 1 was planned.

Statistical Analyses

First, we reported descriptive statistics for the subscales. In a second step, we applied receiver operator characteristic (ROC) analyses to evaluate the diagnostic accuracy of the subscales. Following Swets (38), AUCs ≥ 0.9 are regarded as excellent, AUCs ≥ 0.8 and <0.9 as good, AUCs ≥ 0.7 and <0.8 as fair, and tests with AUCs < 0.7 as poor. We used the bootstrapped test for paired ROC curves—as implemented in the pROC package (39) in R—to compare the AUCs between the subtests. In the next step, logistic regression was applied to investigate whether both subscales independently contribute to the prediction of LD. Finally, we determined an optimal cutoff score using the R-OptimalCutpoints package (40) and estimated the following diagnostic accuracy statistics: sensitivity (Se), specificity (Sp), positive predictive values (PPV), negative predictive values (NPV), and diagnostic likelihood ratios for positive and negative screening results (DLR+ and DLR–, respectively). Following Plante and Vance (41), Se and Sp ≥ 0.90 indicate good diagnostic accuracy, and Se and Sp ≥ 0.80 are regarded as fair. Values below 0.80 indicate an unacceptably high rate of misclassification. DLR+ and DLR– are alternative measures of diagnostic accuracy and have the advantage that—unlike predictive values—they do not depend on the prevalence of the disorder under investigation (42) DLR+ indicates the multiplicative change in the pre-screening odds of having an LD given a positive screening result (i.e., post-screening odds = DLR+ × pre-screening odds) and DLR– is the change in the pre-screening odds of having an LD given a negative screening result (post-screening odds = DLR– × pre-screening odds). DLR+ values ≥ 10 and DLR– ≤ 0.1 indicate large changes in pre-screening odds, DLR+ ≤ 10 and > 5, and DLR– > 0.1 and ≤0.2 indicate moderate changes, DLR+ ≤ 5 and >2, and DLR– > 0.2 and ≤0.5 indicate small changes. DLR+ < 2 and DLR– > 0.5 are rarely important (43). The logistic regression and descriptive analyses were conducted using Jamovi 1.6 (37).

The whole study project (Study 1 and Study 2) was approved by the hospital's ethics commission "Ethikkommission Barmherzige Schwestern und Barmherzige Brüder." All parents gave their written consent to their children's participation in the study.

Results

Descriptive Statistics

The distribution of the screening subscales is depicted in **Figure 1**. The mean of the RG subscale (M = 5.60, SD = 2.52) is above the theoretical mean of 5, indicating the relative ease of the receptive grammar items. In contrast, the mean of the EG subscale (M = 4.85, SD = 3.84) is clearly below the theoretical scale mean of 7.5, indicating that the items of the EG subscale are more difficult. Moreover, the distribution of the EG subscale appears left-censored, indicating that children with a very low EG proficiency all score at the minimum of the EG scale. The correlations of screening variables and reference tests are provided as supplement.

Reliabilty

The EG subscale showed good internal consistency (Cronbach's α = 0.82). In contrast, the internal consistency of the RG subscale was relatively poor (Cronbach's α = 0.61).

Criterion Validity

Both subscales moderately correlate with LD. The point-biserial correlation (r_{pb}) is −0.317 ($p < 0.001$) for RG and −0.384 ($p < 0.001$) for EG. The AUC is fair for RG [0.793, DeLong 95% confidence interval (CI) = (0.623, 0.786)] and excellent for EG [0.912, DeLong 95% CI = (0.857–0.967)]. However, a bootstrapped test for paired ROC curves shows that the AUCs for EG and RG do not differ significantly (D = −1.401, $p > 0.05$). Next, we applied logistic regression to evaluate the independent contribution of RG and EG to LD. Results reveal a significant effect of EG only [b = −1.130, $p < 0.05$; OR = 0.323; 95% CI = (0.136, 0.770)], whereas the effect for RG was not significant [b = −0.164, $p > 0.05$; OR = 0.849; 95% CI = (0.595, 1.212)]. Thus, RG was not found to contribute independently to the prediction of LD.

Cutoff Estimation

Subsequently, we focused on the selection of suitable cutoff values. Due to the non-significant contribution of RG to the prediction of LD, we focused only on EG. Using the "SpEqualSe" criterion (i.e., specificity equals sensitivity) in the Optimal Cutoff Package (40), a cutoff value of 1 turned out to be the most efficient. This cutoff results in acceptable diagnostic accuracy statistics. Sensitivity was high at 0.910 [95% CI = (0.587, 0.998)], specificity was 0.818 [95% CI = (0.728, 0.889)], PPV was 0.357 [95% CI = (0.248, 0.960)], NPV was 0.988 [95% CI = (0.920, 0.993)], DLR+ was 5.000 [95% CI = (3.164, 7.903)], and DLR- was 0.111 [95% CI = (0.017, 0.722)]. Other cutoff values seemed inappropriate because a cutoff of 2 would have resulted in a sensitivity of 1, but a low specificity of 0.707, and a cutoff of 0 would have yielded a low sensitivity of 0.636.

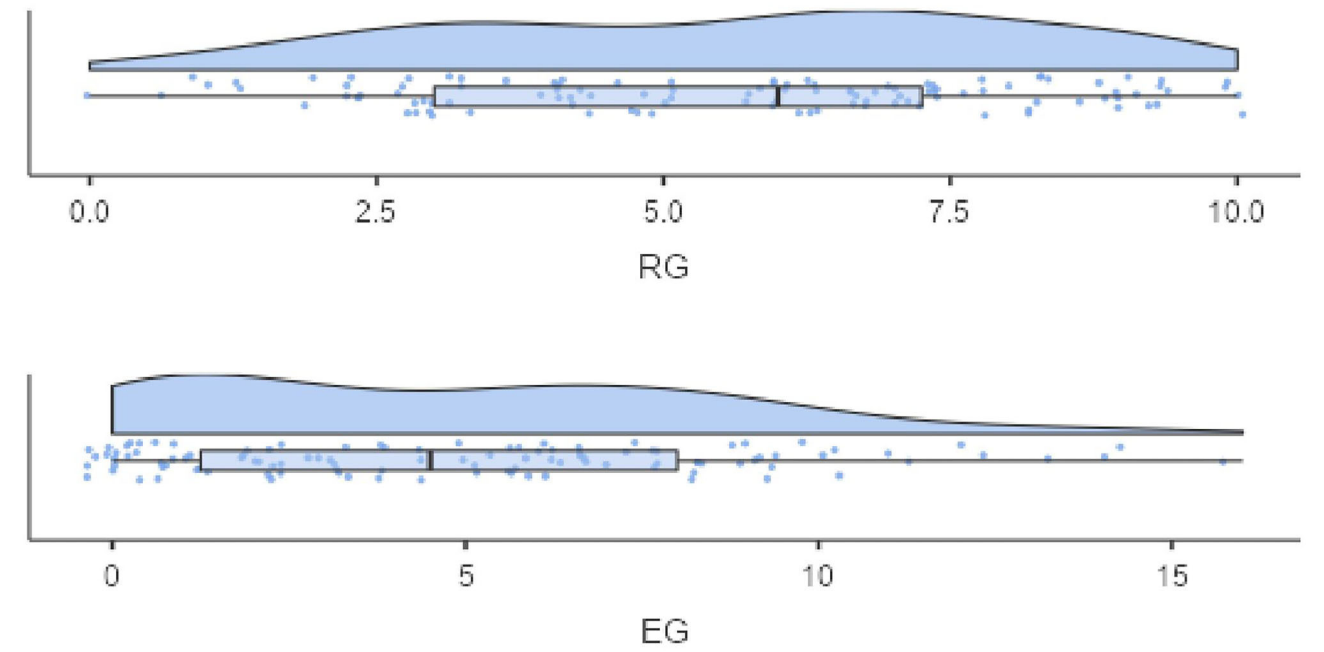


FIGURE 1 | Distribution of the screening subscales—Study 1.

TABLE 1 | Sample description (family and child characteristics).

	Sample A	Sample B	Sample C	Total
Survey year	2014	2015	2017	
Number of pre-schools	12	10	7	27
Sample size	62	96	112	270
Age (months) M (SD)[a]	57.1 (3.65)	59.1 (3.74)	58.9 (3.43)	58.5 (3.67)
Female participants %	34%	56%	55%	50%
Length of institutionalized exposure to the German language (months) M (SD)[b]	17.6 (4.51)	22.9 (7.96)	21.1 (5.66)	20.9 (6.65)

[a] *Tukey Post-hoc test indicates that Sample A is significantly younger than Samples B and C.*
[b] *Games-Howell Post-hoc test indicates that length of exposure in Sample A is significantly lower than in Samples B and C.*

STUDY 2 (VALIDATION STUDY)

Methods

Participants

A total of 443 children in their penultimate year of pre-school were recruited, with parental consent, from 27 public pre-schools in the central area of Upper Austria. For practical reasons, the selected pre-schools were mostly located in the urban central area of Upper Austria, which is characterized by a high proportion of non-German-speaking children (17). Data were collected over a period of 3 years due to limited human resources in the research team and to avoid overburdening the collaborating pre-schools. Participation was voluntary at the pre-school level, and there was no selection of the children. Speech and language therapists from Upper Austria responsible for language screenings in the pre-schools were trained to administer the new measure. They performed the screening in three different test periods (Sample A: 2014, Sample B: 2015, and Sample C: 2017), but did not differ in terms of recruitment (except for the 2017 cohort, which included only children from pre-schools located in the city of Linz). According to parent reports, all the included children had a dominant first language other than German and were therefore acquiring German as a second language (L2). As the new screening tool was intended to identify children with any LD (specific and non-specific) children with additional developmental difficulties (such as hearing loss, cognitive delay, autism-spectrum-disorder) were included in the study sample. Fifty children were excluded because they had <12 months of institutionalized exposure to German. Another 73 children were excluded because of missing data on length of exposure. In addition, 50 children were excluded due to incomplete data for screening or reference tests. Time of exposure was operationalized as the institutionalized contact time (i.e., number of months children were attending pre-schools) because most children are first significantly exposed to German when they enter pre-school. In addition, it was not possible to obtain reliable parent information, and the inclusion of valid parent information on language exposure in the study [e.g., using parent diaries or interviews (44, 45)] was not considered feasible for developing a measure intended for universal screening.

Finally, 270 children were included in Study 2 (mean age = 58.5 months, SD = 3.67; 50% females) (**Table 1**). The children had on average 20.9 months (SD = 6.65) of institutionalized exposure to German. The distribution of first languages was as follows: The main groups were Bosnian/Croatian/Serbian (23.9%), Albanian (14.1%), Turkish (13.2%), Arabic (6.2%), Romanian (5.8%), and Czech (4.5%). This distribution broadly reflects the proportion of the language groups in the Austrian population of pre-schoolers. Between the cohorts, ages varied between 57.1 and 59.1 months, rate of female participants from 34 to 56%, and length of exposure to German in pre-school from 17.6 to 22.9 months. All the differences reached significance levels [age: $F_{(2,267)} = 6.684$, $p < 0.001$, $\eta^2 = 0.048$; exposure to the German language: $F_{\text{Welch}(2,164.92)} = 16.91$, $p < 0.001$, $\eta^2 = 0.089$; sex: $\chi^2_{(2)} = 8.83$, $p < 0.05$, Cramers V = 0.181) demonstrating the diversity of the samples. Children of only two out a total of 27 pre-schools were included in two samples.

Procedures

As in Study 1, the screening procedures were carried out by clinical linguists from the Institute for Neurology of Senses and Language and by trained students of speech-language therapy from the University for Health Professions (Fachhochschule für Gesundheitsberufe) in Linz. After the direct assessment of a child, the results were reported to the parent and the pre-school teachers. Within a maximum of 90 days, the children were tested again using standardized reference tests. The tests were again administered by language experts from the Institute of Neurology of Senses and Language who were blinded to the screening results.

Measures

Screening Measure

The same two screening subscales (EG and RG) were used in Study 2.

Reference Tests

In Study 2, we again used the AWST-R to assess expressive vocabulary, and we also used the LiSe-DaZ, a standardized test for assessing German EG and RG with norms for learners of German as L2 (3–7.11 years), accounting for time of German language exposure. In a systematic review of a variety of pre-school language screening instruments and tests in German, the LiSe-DaZ stood out from the other measures by its good differentiation of tasks and its orientation to a model of language acquisition. In the overall evaluation, the test achieved a "very good" result (46). Following Hamann and Abed Ibrahim (27), the classification of LD was used for children who scored a *t*-value of below 38 (i.e., the 10th percentile) in at least two out of nine subtests and below the 10% percentile in the AWST-R (expressive vocabulary test). Based on this classification, 6.7% ($n = 18$) of the children are regarded as having an LD. Supporting the validity of the LD-classification, there is a strong correlation (Phi = 0.538, $p < 0.001$) between LD and a clinical assessment (LD yes/no) made by clinical linguists for the 2017 sample.This assessment was made directly after the administraton of the reference tests including observations of spontaneous language production and interaction, but before scoring. This information is only available for sample C).

Feasibility

A short questionnaire (10 items) was developed for screeners to assess time economy, acceptance by children and staff, and practicability in the pre-school setting.

Statistical Analyses

We used the same statistical analyses as in Study 2, with two extensions. First, we used confirmatory factor analysis (CFA) for binary items to evaluate the construct validity of the screening scales. The CFA was conducted using weighted least squares estimation (WLSMV) in Mplus 8 (47). Model fit was evaluated following the guidelines proposed by Schermelleh-Engel et al. (48). A good fit is indicated by $\chi^2/\text{df} \leq 2$, CFI ≥ 0.97, RMSEA ≤ 0.05, and the left boundary of the 90% CI of the RMSEA equals 0. An acceptable fit is indicated by $\chi^2/\text{df} \leq 3$, CFI ≥ 0.95, RMSEA ≤ 0.08, and a 90% CI close to the RMSEA. As SRMR has been shown to over-reject models for binary indicators (49), we do not report this fit index. Second, we also conducted tests for unpaired ROC curves. We compared ROC curves between subsamples (age groups, sex and length of exposure to the German language). Significant differences between subsamples indicate variations in diagnostic accuracy and limit the generalizability of the screening results (50). In short, we used a bootstrapped test for unpaired ROC curves to compare the AUC of groups for age, sex, and length of exposure to the German language. Additionally, we applied the Venkatraman permutation test (51) that, instead of AUCs, compares actual ROC curves. Notably, if two ROC curves do not differ significantly, cutoff values would result in the same sensitivity and specificity for the subsamples, indicating that a single cutoff would be appropriate for both subsamples.

Results

Descriptive Statistics

Figure 2 shows the distribution of the screening subscales. Similar to Study 1, the mean of RG (M = 6.38, SD = 2.13) is above the theoretical scale mean of 5. The EG mean (M = 7.86, SD = 4.86) is near the theoretical scale mean of 7.5. However, the EG subscale is again left-censored, indicating that children with low proficiency in EG accumulate at the lower end of the scale. The correlations of screening variables and reference tests are provided as supplement.

Reliability

Again, similar to study 1, EG showed good reliability (Cronbach's $\alpha = 0.88$), and RG repeatedly turned out to have low internal consistency (Cronbach's $\alpha = 0.63$).

Construct Validity

First, we estimated separate single-factor models for RG (M0a) and EG (M0b). Second, we tested a two-dimensional model (RG and EG, M1) against a unidimensional model, where all items

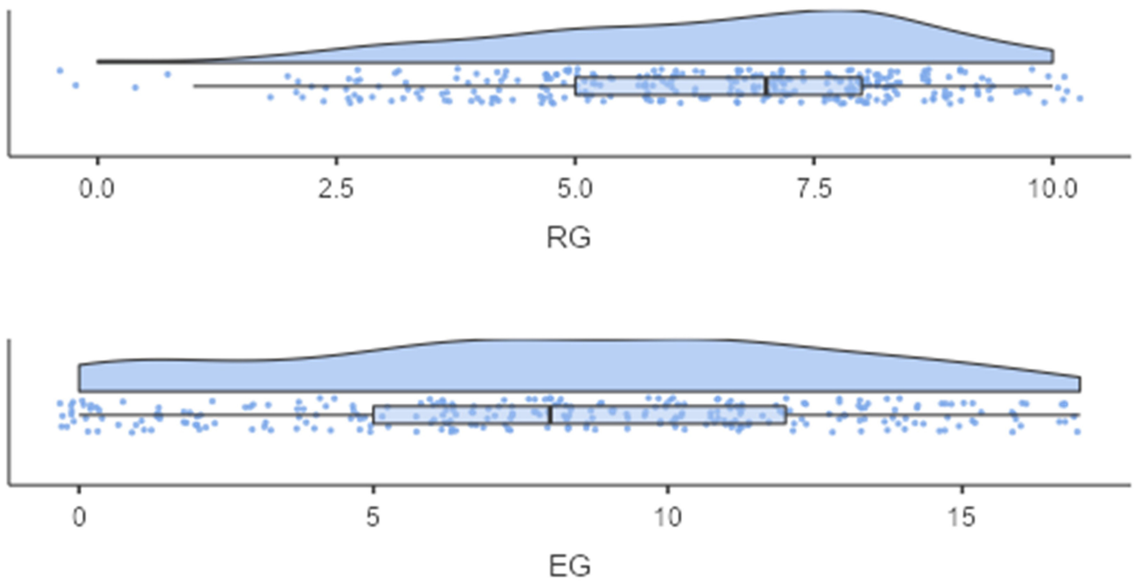


FIGURE 2 | Distribution of the screening subscales—Study 2.

TABLE 2 | CFA-model fit.

	χ^2 (*df*)	$\Delta\chi^2$	RMSEA	CFI
M0a: receptive grammar RG (10 Items)	57.052 (35), $p < 0.05$		0.047 (0.023; 0.068)	0.925
M0b: expressive grammar EG (17 Items)	349.575 (119), $p < 0.001$		0.083 (0.073; 0.093)	0.926
M1: 2 Factors (RG and EG)	513.911 (323), $p < 0.001$	–	0.045 (0.038, 0.053)	0.950
M2: 1 Factor (27 Items)	524.707 (324), $p < 0.001$	8.404 (1), $p < 0.01$	0.046 (0.039; 0.054)	0.947

load on a single latent variable (i.e., general grammar). **Table 2** shows fit indices for the estimated models. The results indicate an acceptable fit for models M0a and M0b. The highly significant ($p < 0.001$) standardized loadings range from 0.34 to 0.79 (median loading = 0.53) for RG and from 0.55 to 0.87 for EG (median loading = 0.70). Furthermore, M1 shows a better fit than M2, supporting the assumption that RG and EG are distinct but highly correlated (latent correlation = 0.87, $p < 0.001$) latent variables.

Criterion Validity

Table 3 shows the means for children with and without LDs on the screening subscales. In addition, the r_{pb} and AUC are reported. As in Study 1, EG shows an excellent AUC of 0.953 [DeLong 95% CI = (0.904, 1.000)], whereas the AUC for RG is good (0.814). A bootstrapped test for paired ROC curves shows that EG outperforms RG (D = −2.523, $p < 0.05$).

A logistic regression shows that—as in Study 1—only EG significantly predicts LD [b = −0.867, $p < 0.001$; OR = 0.420, 95% CI = (0.267, 0.662)]. The additional effect of RG is insignificant [b = −0.036, $p > 0.05$; OR = 0.964, 95% CI = (0.710, 1.309)], indicating that RG does not have an incremental utility in the prediction of LD. Therefore, the EG subscale seems sufficient as a screening tool.

In the next step, we compared AUC and ROC curves between age groups, sex, and groups defined by the length of institutionalized exposure. **Table 4** shows the results. Most notably, AUCs are excellent for all subsamples (>0.90), and we found no significant difference between subsamples. Therefore, these results highlight the generalizability of the diagnostic accuracy across groups and indicate that there is no need for group-specific cutoff values.

Cutoff Estimation

Finally, we again used the "SpEqualSe" criterion (i.e., specificity equals sensitivity) in the Optimal Cutoff Package (40) to determine an optimal cutoff value. The results show that a cutoff value of 1 is the most efficient. This cutoff results in good diagnostic accuracy statistics. Sensitivity and specificity are high at 0.940 [95% CI = (0.727, 0.999)] and 0.936 [95% CI = (0.898, 0.963)], respectively. PPV is 0.515 [95% CI = (0.390, 0.978)], NPV is 0.996 [95% CI = (0.973, 0.998)]. DLR+ and DLR– indicate high confidence in ruling in and ruling out, respectively, a LD. DLR+ is 14.698 [95% CI = (9.031, 23.921)] and DLR– is 0.059 [95% CI = (0.009, 0.399)].

Feasibility

The feasibility questionnaire was completed by 42 out of 46 participating speech-language therapists (91.3%) who administered the new screening measure. The assessment of practicabiltiy and acceptance of the screening measure did

TABLE 3 | Descriptive statistics and AUC for the subtests.

	LD (n = 18)	No LD (n = 252)	r_{pb}	AUC	95% CI (DeLong)	Comparison[a]
(1) Screening RG	2.585 (6.482)	6.679 (1.913)	−0.353***	0.814	[0.689–0.940]	
(2) Screening EG	0.667 (1.878)	8.639 (4.411)	−0.423***	0.953	[0.904–1.000]	D = −2.523, p = 0.012

Comparison is based on a bootstrapped test for unpaired ROC curves. r_{pb}, point-biserial correlation.
*[a]Comparison is based on a bootstrapped test for unpaired ROC curves. ***$p < .001$.*

TABLE 4 | Tests for unpaired ROC curves.

	AUC	95%-CI (DeLong)	Comparisons[a]
A – comparing age groups (median split)			
(1) <59 months	0.968	[0.943–0.994]	
(2) ≥59 months	0.954	[0.886–1.000]	E = 0.004, p = 0.875/D = 0.406, p = 0.685
B – comparing sex			
(1) boys	0.980	[0.960–0.999]	
(2) girls	0.918	[0.805, 1.000]	E = 0.008, p = 0.140/D = 1.115, p = 0.265
C – comparing LoiE-groups			
(1) Screening Total (12–18 months inst. German language contact)	0.959	[0.931–0.987]	
(2) Screening Total (19+ month inst.German language contact)	0.944	[0.843–1.000]	E = 0.005, p = 0.261/D = 0.296, p = 0.767

[a]The first test statistic E refers to the Venkatraman test for paired ROC curves. The second test statistic D refers to a bootstrapped test for paired ROC curves. LoiE, Length of individual exposure to German.

not differentiate between the new instrument for multilingual children and a version for monolingual German children that had been implemented before, as both versions of LOGiK-S are very similar (materials, procedures). Only administration time was collected specifically for the screening of children with German as their second language. Screening time included the whole procedure including expressive and RG and an additional phonology scale. The results of the phonology scale were not used to contribute to the decision of LD or typical development. Speech-language therapists reported an average screening time of 11.9 min (SD = 4.39; range from 5 to 20 min), demonstrating excellent time economy. The feasibility of LOGiK-S within a regular pre-school setting was considered very good and good by almost all the speech-language therapists, and the efficiency of the new measures (again referring to the comprehensive screening) was assessed as good. High rates of child cooperation and rare child refusal (3%) demonstrated high acceptance of the screening tool by the pre-schoolers. In short, the time economy of the screening and its feasibility in pre-school was assessed as "very good." According to the operators, the material is designed in an appealing way and was well-accepted by the children. The time required was also rated as satisfactory, as were the personal effort and the personal burden. Around 92% of the participants would recommend the screening to others.

DISCUSSION

This study investigated the accuracy and feasibility of the newly developed screening measure LOGiK-S in identifying an increased risk of LDs in three sequentially recruited cohorts of bilingual pre-schoolers (n = 270, mean age 58.6 months) with German as their second language. A study to develop the screening measures, including initial validation, preceded the comprehensive validation study. The screening was intended for use within the established universal language screening procedure by speech-language-therapists in the penultimate year of pre-school (age 4–5 years) within the regular pre-school settings, and within a constrained time-frame.

The whole study sample was screened and subsequently assessed using standardized language tests. For the validation sample the results of ROC analyses demonstrated high accuracy of the EG screening, with an excellent AUC (0.953). Using a cutoff of 1, the rate of screening fails was 17%, and sensitivity (0.940) and specificity (0.936) were found to be high. In 51.5% (positive predictive value) of these children, a LD was confirmed by standardized language assessments and the application of bilingual norms.

The RG component of the screening did not increase the screening accuracy achieved by the expressive subtest and was therefore regarded as a non-essential component of the screening procedure. However, since limited receptive skills have been found to predict the persistence of LDs (52), the use of receptive screening as a second-step measure for those who screened positive in the EG component might be considered as a tool that helps to better estimate the probability of a persisting LD requiring speech-language therapy. However, for an evidence-based recommendation of a two-step screening, the prospective predictive quality of the receptive measure requires confirmation.

Despite some diversity in the characteristics of the three cohorts (length of L2 exposure, age, and sex) and in pre-school settings (urbanization level), LOGiK-S demonstrated high predictive accuracy in all samples. This can certainly be considered a strength in an instrument to be used with a variety of children in diverse pre-schools. The non-significant effect of length of time of L2 exposure on the screening results may initially be surprising. However, because in many pre-schools attended by children with German as an L2, a high number, and often the majority of their peers, have family languages other than German, it is very likely that—despite pre-school attendance—the daily quantity of high-quality German language input and particularly the amount of active participation in language interactions in German is limited and highly variable. The quality and quantity of everyday L2 input in the pre-school from peers and caregivers can most likely be considered more relevant to L2 development than the length of L2 exposure (53–55).

Although ASHA (56) proposes that bilingual children be assessed in both languages, a number of practical constraints render the implementation of the guidelines difficult or even impossible. Even obtaining reliable information on first language acquisition and L2 language exposure of all pre-school children with German as a second language is hardly feasible. The present results show that testing in the majority language with norms for learners as L2 can be regarded as a practical and accurate alternative.

LIMITATIONS

The high number of children attending a pre-school with an accumulation of learners of German as an L2 might be considered a limitation of this study because our findings might not be generalizable to the total population of children with a L1 other than German. On the other hand, the majority of children with German L2 acquisition in Austria representing the target group for the screening live in urbanized areas and attend pre-schools with a high percentage of children with migrant backgrounds. The exclusion of children attending the first year of pre-school is a limitation. However, our results show that despite their exclusion, simple EG items were challenging for many bilingual children, as demonstrated by the low cutoff. The lack of a well-defined gold standard for LDs in general and—more specifically—in bilingual children must still be regarded as a significant challenge for developing screening measures.

CONCLUSION

The LOGiK-S EG screening is feasible and identifies LD in children with a variety of first languages other than German. Using a screening measure focusing on the acquisition of German expressive grammar applying specific bilingual norms allows for reliable differentiation between children with and without LDs, even though standardized first language testing is not practical.

AUTHOR'S NOTE

DH is a clinical linguist and director of the center for communication and language at the Institute of Neurology of Senses and Language at the Hospital of St. John of God. His research interests concern the early identification of developmental disorders, the efficacy of interventions in disorders of speech/language/communication, and the association between communication skills and mental health. CW is a social scientist and is working at the University of Education Upper Austria and the Research Institute for Developmental Medicine, Johannes Kepler University Linz. His research interests are quantitative methods and educational inequalities. MJ is a clinical linguist working at the Institute for Neurology of Senses and Language at the Hospital of St. John of God in Linz (Austria). She specializes in the field of diagnosing speech, language, and communication needs in monolingual and bilingual children.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because parents have not given their consent to data sharing. Requests to access the datasets should be directed to daniel.holzinger@bblinz.at.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethikkommission Barmherzige Schwestern und Barmherzige Brüder, Linz. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

DH: conceptualization and formal analysis. CW and DH: methodology. MJ: data curation. MJ, DH, and CW: writing original draft and review and editing. DH and MJ: project administration. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Department of Social Affairs of the Upper Austrian Government. Article processing charge is funded by the Johannes Kepler University Open Access Publishing Fund.

ACKNOWLEDGMENTS

We express our gratitude to all the Upper Austrian speech-language pathologists cooperating with us in the current project.

REFERENCES

1. Leonard LB. *Children With Specific Language Impairment*. Cambridge: MIT Press (2017). p. 490.
2. Neumann K, Holler-Zittlau I, van Minnen S, Sick U, Zaretsky Y, Euler HA. Katzengoldstandards in der Sprachstandserfassung: sensitivität und Spezifität des Kindersprachscreenings (KiSS). *e-HNO*. (2011) 59:97–109. doi: 10.1007/s00106-010-2231-6
3. Norbury CF, Gooch D, Wray C, Baird G, Charman T, Simonoff E, et al. The impact of nonverbal ability on prevalence and clinical presentation of language disorder: evidence from a population study. *J Child Psychol Psychiatry*. (2016) 57:1247–57. doi: 10.1111/jcpp.12573
4. Tomblin JB, Records NL, Buckwalter P, Zhang X, Smith E, O'Brien M. Prevalence of specific language impairment in kindergarten children. *J Speech Lang Hear Res*. (1997) 40:1245–60. doi: 10.1044/jslhr.4006.1245
5. Bishop DVM, Snowling MJ, Thompson PA, Greenhalgh T, The CATALISE Consortium. CATALISE: a multinational and multidisciplinary Delphi consensus study. Identifying language impairments in children. *PLoS ONE*. (2016) 11:0158753. doi: 10.1371/journal.pone.0158753
6. Norbury CF, Vamvakas G, Gooch D, Baird G, Charman T, Simonoff,E., et al. Language growth in children with heterogeneous language disorders: a population study. *J Child Psychol Psychiatry*. (2017) 58:1092–105. doi: 10.1111/jcpp.12793
7. Thordardottir E. Proposed diagnostic procedures for use in bilingual and cross-linguistic contexts. In: Armon-Lotem S, de Jong J, Meir N. editors. *Assessing Multilingual Children: Disentangling Bilingualism From Language Impairment*. Bristol: Multilingual Matters (2015). p. 331–58.
8. Lindsay G, Dockrell JE, Strand S. Longitudinal patterns of behaviour problems in children with specific speech and language difficulties: Child and contextual factors. *Br J Educ Psychol*. (2007) 77:811–28. doi: 10.1348/000709906X171127
9. Sim F, Thompson L, Marryat L, Ramparsad N, Wilson P. Predictive validity of pre-school screening tools for language and behavioural difficulties: A PRISMA systematic review. *PLoS ONE*. (2019) 14:e0211409. doi: 10.1371/journal.pone.0211409
10. Beitchman JH, Wilson B, Brownlie EB, Walters H, Lancee W. Long-term consistency in speech/language profiles: Developmental I. academic outcomes. *J Am Acad Child Adolesc Psychiatry*. (1996) 35:804–14. doi: 10.1097/00004583-199606000-00021
11. Law J, Rush R, Schoon I, Parsons S. Modeling developmental language difficulties from school entry into adulthood: Literacy, mental health, employment outcomes. *J Speech Lang Hear Res*. (2009) 52:1401–16. doi: 10.1044/1092-4388(2009/08-0142)
12. Paradis J, Crago M, Genesee F. French-English bilingual children with SLI: how do they compare with their monolingual peers? *J Speech Lang Hear Res*. (2003) 46:113–27. doi: 10.1044/1092-4388(2003/009)
13. Buschmann A, Jooss B, Rupp A, Feldhusen F, Pietz J, Philippi H. Parent-based language intervention for two-year-old children with specific expressive language delay: a randomized controlled trial. *Arch Dis Childhood*. (2009) 94:110–6. doi: 10.1136/adc.2008.141572
14. Larson AL, Cycyk LM, Carta JJ, Hammer CS, Baralt M, Uchikoshi Y, et al. A systematic review of language-focused interventions for young children from culturally and linguistically diverse backgrounds. *Early Childh Res Q*. (2020) 50:157–78. doi: 10.1016/j.ecresq.2019.06.001
15. Law J, Garrett Z, Nye C. The efficacy of treatment for children with developmental speech and language delay/disorder. *J Speech Langu Hear Res*. (2004) 47:924–43. doi: 10.1044/1092-4388(2004/069)
16. Roberts MY, Kaiser AP. The effectiveness of parent-implemented language interventions: a meta-analysis. *Am J Speech-Lang Pathol*. (2011) 20:180–99. doi: 10.1044/1058-0360(2011/10-0055)
17. Biedermann H, Weber C, Herzog-Punzenberger, Nagel B. A Auf die Mitschüler/innen kommt es an? Schulische Segregation – Effekte der Schul- und Klassenzusammensetzung in der Primarstufe und der Sekundarstufe I. In: *Bruneforth M*, et al. editors. *Nationaler Bildungsbericht Österreich. Band 2: Fokussierte Analysen bildungspolitischer Schwerpunktthemen*. Graz: Leykam (2016). p. 133–74.
18. Marinis T, Armon-Lotem S, Pontikas G. Language impairment in bilingual children: state of the art 2017. *Ling Appr Bilingual*. (2017) 7:265–76. doi: 10.1075/lab.00001.mar
19. Rothweiler M. Multilingualism and specific language impairment. In: Auer P, Wei L, editors. *Handbook of Multilingualism and Multilingual Communication*. Berlin: De Gruyter (2009). p. 229–46.
20. Gogolin I. Erziehungsziel Mehrsprachigkeit. In: Röhner, C, editor. *Erziehungsziel Mehrsprachigkeit Diagnose von Sprachentwicklung und Förderung von Deutsch als Zweitsprache*. Weinheim/München: Juventa (2005). p. 13–24.
21. Schroeder C, Dollnick M. Mehrsprachige Gymnasiasten mit türkischem Hintergrund schreiben auf Türkisch. In: *[Interdisziplinäres Symposium]. Mehrsprachig in Wissenschaft und Gesellschaft. Interdisziplinäres Symposium zu Mehrsprachigkeit, Bildungsbeteiligung und Potenzialen von Studierenden mit Migrationshintergrund*. Bielefeld (2013). doi: 10.2390/biecoll-mehrspr2013_11
22. Montrul S. *Incomplete Acquisition in Bilingualism. Re-Examining the Age Factor*. Amsterdam: Benjamins (2008). p. 312.
23. Montrul S. Current issues in heritage language acquisition. *Annu Rev Appl Linguist*. (2010) 30:3–23. doi: 10.1017/S0267190510000103
24. Novogrodsky R. Specific Language Impairment (SLI) is not specific enough: Sub-types of SLI and their implications for the theory of the disorder. In: Stavrakaki St, editor. *Language Acquisition and Language Disorders*. Amsterdam: John Benjamins. (2015). p. 113–24.
25. Gillam RB, Peña ED, Bedore LM, Bohman TM, Mendez-Perez A. Identification of specific language impairment in bilingual children: I. Assessment in English. *J Speech Lang Hear Res*. (2013) 56:1813–23. doi: 10.1044/1092-4388(2013/12-0056)
26. Schulz P, Tracy R. *Linguistische Sprachstandserhebung – Deutsch als Zweitsprache (LiSe-DaZ)*. Göttingen: Hogrefe (2011).
27. Hamann C, Abed Ibrahim. L. Methods for identifying specific language impairment in bilingual populations in Germany. *Front Commun*. (2017) 2:16. doi: 10.3389/fcomm.2017.00016
28. Woon CP, Yap NT, Lim HW, Wong BE. Measuring grammatical development in bilingual Mandarin-English speaking children with a sentence repetition task. *J Educ Learn*. (2014) 3:144–57. doi: 10.5539/jel.v3n3p144
29. de Almeida L, Ferré S, Morin E, Prévost P, dos Santos C, Tuller L, et al. Identification of bilingual children with specific language impairment in France. *Linguistic Appr Bilingual*. (2017) 7:331–58. doi: 10.1075/lab.15019.alm
30. Rothweiler M, Chilla S, Clahsen H. Subject-verb agreement in specific language impairment: a study of monolingual and bilingual German-speaking children. *Bilingualism*. (2012) 15:39–57. doi: 10.1017/S136672891100037X
31. Chilla S, Rothweiler M, Babur E. *Kindliche Mehrsprachigkeit. Grundlagen – Störungen – Diagnostik*. München: Ernst Reinhardt (2013). p. 139.
32. Kauschke C, Siegmüller J. *Patholinguistische Diagnostik bei Sprachentwicklungsstörungen (PDSS)*. München: Urban and Fischer (2009). p. 808.
33. Angermaier M. *Entwicklungstest Sprache für Kinder von 4 bis 8 Jahren: ETS 4-8*. Harcourt Test Services (2007).
34. Fox-Boyer AV. *Test zur Überprüfung des Grammatikverständnisses (TROG-D)*. Auflage Idstein: Schulz-Kirchner. (2011). p. 44.
35. Kiese-Himmel C. *Aktiver Wortschatztest für 3- bis 5-jährige Kinder – Revision (AWST-R). Göttingen: Hogrefe* (2005). p. 112.
36. Lenhard W, Lenhard A, Gary S. *cNORM: Continuous Norming*. Vienna: The Comprehensive R Network (2021). Available online at: https://cran.r-project.org/web/packages/cNORM/ (accessed September 14, 2018).
37. The Jamovi Project. *jamovi (Version 1.6) [Computer Software]*. Retrieved from: https://www.jamovi.org (accessed September 1, 2021).
38. Swets JA. Measuring the accuracy of diagnostic systems. *Science*. (1988) 240:1285–93.
39. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*. (2011) 12:77. doi: 10.1186/1471-2105-12-77
40. López-Ratón M, Rodríguez-Álvarez M, Cadarso-Suárez C, Gude-Sampedro F. OptimalCutpoints: a R package for selecting optimal cutpoints in diagnostic tests. *J Stat Softw*. (2014) 61:1–36. doi: 10.18637/jss.v061.i08
41. Plante E, Vance R. Selection of pre-school language tests: a data-based approach. *Lang Speech Hear Serv Sch*. (1994) 25:15–24. doi: 10.1044/0161-1461.2501.15
42. Pepe MS. *The Statistical Evaluation of Medical Tests for Classification and Prediction*. New York, NY: Oxford University Press (2003).

43. Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results will they help me in caring for my patients? The Evidence-Based Medicine Working Group. *JAMA*. (1994)271:703–7. doi: 10.1001/jama.271.9.703
44. Place S, Hoff E. Properties of dual language exposure that influence 2-year-olds' bilingual proficiency. *Child Dev*. (2011) 82:1834–49. doi: 10.1111/j.1467-8624.2011.01660.x
45. Bosch L, Sebastián-Gallés N. Early language differentiation in bilingual infants. In: Cenoz, J, and Genesee, F. editors. *Trends in Bilingual Acquisition.* Amsterdam: John Benjamins (2001). p. 71–93.
46. Neugebauer U, Becker-Mrotzek M. *Die Qualität von Sprachstandsverfahren im Elementarbereich. Eine Analyse und Bewertung* (2021). Available online at: https://www.mercator-institut-sprachfoerderung.de/fileadmin/user_upload/Mercator-Institut_Qualitaet_Sprachstandsverfahren_Web_03.pdf (accessed September 14, 2013).
47. Muthén LK, Muthén B. *Mplus User's Guide*. Los Angeles, CA: Muthén and Muthén (1998–2017).
48. Schermelleh-Engel K, Moosbrugger H, Müller H. Evaluating the fit of structural equation models: tests of significance and descriptive goodness-of-fit measures. *Methods Psychol Res Online.* (2003) 8:23–74.
49. Yu CY. *Evaluating Cutoff Criteria of Model Fit Indices for Latent Variable Models With Binary and Continuous Outcomes.* Los Angeles, CA: University of California (2002).
50. Youngstrom EA. A primer on receiver operating characteristic analysis and diagnostic efficiency statistics for pediatric psychology: we are ready to ROC. *J Pediatric Psychol*. (2014) 39:204–21. doi: 10.1093/jpepsy/jst062
51. Venkatraman ES. A permutation test to compare receiver operating characteristic curves. *Biometrics*. (2000) 56:1134–8. doi: 10.1111/j.0006-341X.2000.01134.x
52. Bishop DV. Why is it so hard to reach agreement on terminology? The case of developmental language disorder (DLD). *Int J Lang Commun Disord.* (2017) 52:671–80. doi: 10.1111/1460-6984.12335
53. Mashburn AJ, Justice LM, Downer JT, Pianta RC. Peer effects on children's language achievement during pre-kindergarten. *Child Dev*. (2009) 80:686–702. doi: 10.1111/j.1467-8624.2009.01291.x
54. Unsworth S. Quantity and quality of language input in bilingual language development. In: Nicoladis E, Montanari S, editors. *Bilingualism Across the Lifespan*. Berlin: De Gruyter Mouton. p. 103–22.
55. Yang N, Shi J, Lu J, Huang Y. Language development in early childhood: Quality of teacher-child interaction and children's receptive vocabulary competency. *Front Psychol.* (2021) 12:649680. doi: 10.3389/fpsyg.2021.649680
56. American Speech-Language-Hearing Association (2004). Available online at: https://www.asha.org/ (accessed September 14, 2021).

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.

frontiers in Pediatrics

ORIGINAL RESEARCH
published: 11 March 2022
doi: 10.3389/fped.2022.750126

Identifying Stuttering in Arabic Speakers Who Stutter: Development of a Non-word Repetition Task and Preliminary Results

Roaa Alsulaiman[1], *John Harris*[1], *Sarah Bamaas*[2] *and Peter Howell*[1]*

[1] *Division of Psychology and Language Sciences, University College London, London, United Kingdom,* [2] *Speech and Language Pathology Division, Jeddah Institute for Speech and Hearing, Jeddah, Saudi Arabia*

Stuttering and other conditions that affect speech fluency need to be identified at an early age in order that effective interventions can be given before the problems becomes chronic. This applies in countries where several languages are spoken including those in which English and Arabic are both widely used which calls for assessment procedures that work across these languages. The 'universal' non-word repetition task (UNWR) has been established as an effective screening tool for discriminating between children who stutter (CWS) and children with word-finding difficulty for a number of languages. However, the UNWR does not apply to languages such as Arabic and Spanish. The present study aimed to: (1) introduce an Arabic English NWR (AEN_NWR); which was developed based on the same phonologically informed approach used with UNWR; (2) present preliminary non-word repetition data from Arabic-speaking CWS and adults who stutter (AWS). The AEN_NWR items comprises twenty-seven non-words that meet lexical phonology constraints across Arabic and English. The set of items includes non-words of two, three and four syllables in length. Preliminary non-word repetition data were collected from ten CWS between the ages of 6;5 and 16;7 ($M_{age} = 12{:}1$) and fourteen AWS between the ages of 19;2 and 31;0 ($M_{age} = 24$). Participants performed the non-word repetition task and provided a sample of spontaneous speech. The spontaneous speech samples were used to estimate %stuttered syllables (%SS). To validate that AEN_NWR performance provides an alternative way of assessing stuttering, a significant correlation was predicted between %SS and AEN_NWR performance. Also, word length should affect repetition accuracy of AEN_NWR. As predicted, there was a significant negative correlation between the AEN_NWR and %SS scores (r (25) $= -0.5$), $p < 0.000$). Overall, CWS were less accurate in their repetition than AWS at all syllable lengths. The AEN_NWR provides a new assessment tool for detecting stuttering in speaker of Arabic and English. Future studies would benefit from a larger sample of participants, and by testing a population-based sample. These studies would allow further investigation of the AEN_NWR as a screening measure for stuttering in preschool children.

Keywords: fluency, stuttering, screening, Arabic, speech disfluency, word-finding, non-word, diversity

OPEN ACCESS

Edited by:
Daniel Holzinger, Hospitaller Brothers of Saint John of God Linz, Austria

Reviewed by:
Riffat Mehboob, King Edward Medical University, Pakistan
Ivana I. Kavecan, University of Novi Sad, Serbia

***Correspondence:**
Peter Howell
p.howell@ucl.ac.uk

Specialty section:
This article was submitted to Children and Health, a section of the journal Frontiers in Pediatrics

Received: *30 July 2021*
Accepted: *04 February 2022*
Published: *11 March 2022*

Citation:
Alsulaiman R, Harris J, Bamaas S and Howell P (2022) Identifying Stuttering in Arabic Speakers Who Stutter: Development of a Non-word Repetition Task and Preliminary Results. Front. Pediatr. 10:750126. doi: 10.3389/fped.2022.750126

Andrews and Harris (1) reported that the lifetime incidence of stuttering is 1% in their study on stuttering in 1,142 families in the United Kingdom that used children born between May and June 1947 in Newcastle, UK. The study ended when the children were 15 years old, and established that the point prevalence of stuttering up to the age 15 was approximately 4.9%. Yairi and Ambrose (2) confirmed that approximately 5% of pre-school age children exhibit episodes of stuttering. Stuttering and other conditions that affect speech fluency need to be identified at an early age so that effective interventions can be given before the problem becomes chronic (3, 4). This applies in countries where several languages are spoken including those in which English and Arabic are both widely used. In countries of the latter type, some children use both languages, and they may be less fluent in the official language used in schools (e.g. English in the UK) than the one they use in their home (Arabic in this example). Whilst it would be possible to wait for fluency in, for instance, English to develop in school before attempting to identify cases of stuttering, this would delay identification and intervention of children affected by fluency issues. Delaying intervention for some time after a disorder has begun may lead to other effects that result in lower educational achievement, and behavioral and social problems in a child's later life (4). Consequently, a school-based screening procedure has been developed for identifying children with Speech Language Communication Needs (SLCN), including stuttering, for use in reception classes (5). Howell's procedure separates fluent children, those children with word-finding difficulty (WFD), which could arise *inter alia* when children use English as an Additional Language (EAL) and those with SLCN. As validation of the procedure, a spontaneous speech sample was obtained from each child, and analyzed for symptoms of stuttering and WFD. Three speech symptoms were used (part-word repetitions, prolongations and word breaks) to identify stuttering which was quantified as percentage of stuttered syllables out of all syllables spoken (%SS) as in Riley's (6) Stuttering Severity Instrument (SSI). Stuttering was identified when children had rates of stuttering symptoms above a threshold %SS which did not preclude them also exhibiting high rates of WWR (7). Any remaining children with %WWR above threshold were designated as having WFD whilst children below %SS and %WWR thresholds were designated fluent.

School staff and teachers do not have much time to dedicate to SLCN because they are under pressure to deliver national curriculums (4). Consequently, Mirawdeli (8) argued that assessment procedures are needed that are quick and practical to use in schools. A further complication is that large numbers of pre-schoolers who use a wide number of native languages need to be screened. This requires forms of assessment that are independent of language spoken. To this end, Howell et al. (9) developed a "universal" non-word repetition task (UNWR) that is convenient to administer and score and applies to at least 20 languages[1] spoken in UK schools. Howell et al. (9) showed that the UNWR is an effective screening instrument for discriminating between fluent children, children who stutter (CWS) and children with WFD. As a *non-word* test, UNWR controls for extraneous influences of lexical knowledge, making it a sensitive marker for children's phonological ability. UNWR uses consonants that occur in all languages that the test applies to. To generate UNWR test items, the overlapping phonotactic properties of onsets and codas for legitimate syllables in the selected languages were identified, and rules for concatenating syllables for constructing multisyllabic non-words were applied. Next, exemplars consisting of all strings meeting these constraints were automatically generated, and bespoke dictionaries created and used to exclude candidates when the strings were words that occurred in any of the languages included in UNWR. The test was administered to 96 children from reception classes in five mainstream primary schools in the United Kingdom of which 20.83% used EAL (9). The spontaneous speech samples from the children were assessed for symptoms of stuttering and WFD, and their performance on the UNWR was measured. Stuttering symptoms (measured by %SS) predicted UNWR scores, whereas WFD scores (measured by %WWR) did not. The findings were interpreted as confirming that UNWR scores differentiate stuttering from WFD.

The issue addressed in this paper arose because UNWR does not apply to languages with vastly different phonological structures from languages like English, such as Arabic, Spanish and Mandarin. This paper attempted to fill one of these gaps by providing a new NWR task for screening Arabic and English children, the Arabic English NWR (AEN_NWR). AEN_NWR is based on the same phonologically-informed approach used with UNWR. The test items were constructed so that they accord with various constraints on lexical phonology common to Arabic and English. This quick and easy-to-administer test includes stimuli that vary in syllable structure, consonant age of acquisition, lexical effects and stress.

A brief review of the literature on NWR as a potential behavioral clinical marker for identifying language disorders is presented next. This literature has focused on non-word repetition and aspects of phonological performance in CWS as well as children with other language disorders such as specific language impairment (now referred to as developmental language disorder, DLD) and dyslexia. NWR can potentially inform work on how phonology pertains to speech fluency. Following this review, a description of Arabic phonology specifying areas that were involved in the generation of Arabic and English non-words is presented.

Non-word Repetition as a Language Assessment Tool

The ability to repeat a novel phonological sequence is a basic language skill that humans possess. Infants under the age of 12 months attend to speech sounds, especially when the sounds are spoken to them by adults (10). Infants can spontaneously imitate the words of others and by the time they are 2 years old they repeat non-word when requested (11). Children also repeat "non-words" spontaneously when they

[1] The 20 languages the UNWR applies to are English, Polish, Romanian, European Portuguese, Bulgarian, Serbo-Croat-Bosnian, Czech, Dutch, French, German, Hungarian, Slovene, Swedish, Danish, Norwegian, Russian, Latvian, Ukrainian, Urdu-Hindi and Bengali.

mimic real words spoken to them by adults (12). Although non-word repetition (NWR) tasks appear simple, they rely on the following cognitive processes. First, the person must process the acoustic signal, extract phonemes and match the signal with phonological representations in memory. Then, the person must plan the articulatory movements for achieving production of the non-word and execute this plan as their response (13). Correlations between phonological working memory and non-word repetition were first examined by Gathercole and Baddeley (14). Performance on non-word tasks has been examined in children with several language disorders such as dyslexia (15) and specific language impairment (16). Performance on NWR tasks presents a challenge for children with these language disorders. For instance, Snowling's (15) study that examined NWR non-word abilities in children with dyslexia tested two groups of children: 22 typical readers and 20 dyslexic children whose age ranged between 7 and 17 years old. The children repeated 30 non-words that were two, three or four syllables long. non-words posed more difficulty for dyslexic children than for the control group with dyslexic children making more repetition errors. Significant differences between the groups were found when four syllable non-words were repeated. The finding was interpreted as indicating a phonological deficit in dyslexic children. Subsequent studies [e.g., (17)] agree that dyslexia should be considered a phonological deficit indicating language weakness rather than impaired low-level auditory difficulties (18).

On the Relationship Between NWR and the Phonological Loop

Before turning to studies on NWR and stuttering, details are given about phonological memory to illustrate its role in repetition of non-words or unfamiliar words. The phonological loop is part of working memory (WM), which is a cognitive system that temporarily holds and manipulates information whilst people perform tasks such as comprehension and learning (19). The WM-model of Baddeley and Hitch (20) proposed three major components: (1) the central executive system, which is the supervisory controlling system that is aided by the other two components; (2) The visuospatial sketchpad, which is concerned with the visuo-spatial memory; that is, it stores and processes information in a visual or spatial form; and (3) the articulatory loop (now referred to as the phonological loop) that is responsible for rehearsing and storing speech-based verbal information. It transforms the verbal stimuli into phonological codes which have the associated acoustic and temporal properties of the stimuli. Matches between the phonological codes and codes that exist in the long-term memory system (i.e., phonemes and words) are sought. The phonological loop can be further divided into two sub-components: the phonological short-term store and the subvocal rehearsal component (21). The phonological short-term store is like an inner ear which holds speech-based information for up to about two seconds. Speech-based information can be maintained by the subvocal articulatory rehearsal component; a process that can be used to enter information into the phonological store. The subvocal rehearsal component is like an inner voice which allows information to be rehearsed, for example, object names that are articulated either overtly or sub-vocally (21). Rehearsal allows a person to remember a telephone number by circulating the phone digits to oneself. **Figure 1** provides a graphical representation of the phonological loop model based on Baddeley (21) adapted from Gathercole (22).

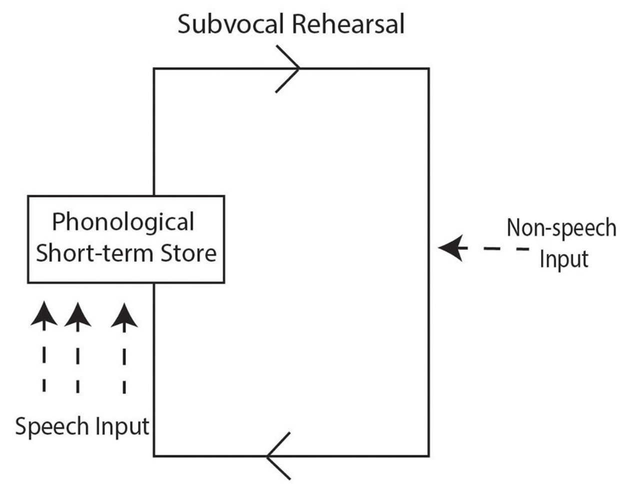


FIGURE 1 | Schematic representation of the phonological loop model based on Baddeley (21) adapted from Gathercole (22), See text for further detail.

We now provide a brief overview of the way the phonological loop operates when repeating non-words. As implied earlier, it has been established that there is a strong relationship between NWR performance and the phonological loop component of WM (23). Repeating non-words requires temporary storage of unfamiliar phonological sequences in the phonological loop; and it is assumed that success when holding the sequences depends on the short-term memory capacity of the phonological loop. The rehearsal component of the phonological loop serially reactivates the unfamiliar phonological sequence stored in the phonological store, where this process does not necessarily involve movements of speech articulators. As long as rehearsal is maintained, the phonological store can hold on to the speech information. Indeed, the process of rehearsal is time-limited; the longer a phonological sequence is, the longer it takes to reactivate the sequence leading to fewer rehearsals in a given time (22).

Non-word Repetition and Stuttering

Several studies have reported how NWR performance is affected in CWS and adults who stutter (AWS). Hakim and Ratner (24) investigated the performance of eight CWS and eight children who do not stutter (CWNS) on non-word repetition where the children's ages ranged between 4 and 8 years. Children attempted to repeat 40 non-words from Gathercole et al.'s (23) Children's Test of Non-word Repetition (CNRep). The task consisted of 40 non-words of 2-, 3-, 4-, and 5-syllables in length. Results showed that CWS were less accurate at repeating non-words at all syllable-lengths, although statistical differences between participant groups only occurred for the 3-syllable non-words but not the 2-syllable and longer 4- and 5-syllable words. Anderson et al. (25) replicated and extended the findings from Hakim and Ratners's (24)'s work in a sample of younger children (aged 3 to 5 years). The authors argued that examining non-word repetition performance at this young age could provide an opportunity to assess phonological memory during a time of critical language development, relative to school aged children. Anderson et al. (25) administered the CNRep to 12 CWS and a matched control group of 12 CWNS. CWS were significantly less accurate in repeating non-words of two and three syllables. However, no significant differences were found between the groups in their accuracy of repeating the longer non-words (four and five syllable non-words). The findings of this study is partially consistent with results from Hakim and Ratner (24). The authors suggested that the significant differences between the groups even with the two syllable non-words were probably because this is a younger group of children; thus, their performance was not impacted by ceiling effects. On the other hand, the lack of significance effects for the longer non-words was due to the impact of floor effects in both groups.

Subsequently, Anderson and Wagovich (26) examined the relationships between measures of linguistic processing speed and between two aspects of cognition: phonological working memory and attention. Nine CWS and a matched control group of 14 CWNS, aged 3 to 5 years, participated. Gathercole et al.'s (23) CNRep test was again used in this study. Children were asked to repeat 40 non-words, where there were 10 each of 2-, 3-, 4-, and 5-syllable non-words. There were significant differences between the two groups in their accuracy of repeating non-words of two and three syllables. However, the differences between the two groups were not significant when repeating non-words of four and five syllables, although CWS performed worse (i.e., were less accurate in their repetition) with them than were controls. These findings are consistent with Anderson et al.'s (25) study and the lack of significant differences on the longer nonwords could be attributed to floor effects in both groups.

Sasisekaran and Byrd (27) investigated NWR accuracy in 14 CWS and a matched control group of CWNS aged between 8 and 15 years of age. Participants repeated a set of 36 non-words consisting of 12 non-words at each syllable lengths (2-, 3-, 4-, and 7-syllables). CWS were less accurate when producing two-syllable non-words compared with the CWNS. However, differences between CWS and CWNS on accuracy at each syllable length was not reported. The non-words at four-syllables posed most difficulty for children in both groups (i.e. had the lowest percent of correct repetitions). Based on this result, the authors suggested that their findings are consistent with previous studies; confirming that CWS show a trend to perform poorly on NWR tasks. Whilst the studies above focused on English-speaking children, Sugathan and Maruthy (28) explored NWR performance in Kannada-speaking school-aged children. Seventeen CWS and a matched group of CWNS were tested. The non-words consisted of 2-, 3- and 4-syllables, and for each syllable length there were 12 non-words. These were language specific in that they conformed with the phonotactic constraints of the Kannada language. CWS were less accurate in producing the non-words compared to the CWNS at all syllable lengths. Significant differences between the two groups were reported for the mean number of correct non-words; however, whether the differences at each syllable length were significant was not reported.

Howell et al. (9) investigated NWR ability in a group of 96, 4–5-year old monolingual English children and children with EAL, who came from diverse language backgrounds, using the UNWR. The goal was to evaluate the effectiveness of the UNWR in distinguishing between CWS and children with WFD, irrespective of which language they speak. Children with EAL often have to produce phonological sequences in English; a language they are not familiar with. This is similar to what happens in NWR tasks where children are required to repeat a novel sequence of phonemes that does not exist as a word in their first language. Howell et al.'s (9) study faced the methodological problem, common to many NWR tasks, that materials tend to be biased toward the language for which the test was developed. Howell et al.'s (9) "universal" NWR task was developed to apply to various languages spoken in UK schools hence the name "universal" and avoids confounds between language ability in the test language and presence of stuttering symptoms. If a person has WFD but no stuttering symptoms, this should not be evident when repeating non-words, whereas CWS are expected to struggle performing the task (9). UNWR scores were predicted by %SS, but not by WFD [as measured by the percentage of whole word repetition (%WWR)]. This relationship between UNWR and %SS provided empirical evidence that UNWR provides a sensitive measure of stuttering. The authors also attributed this relationship to the fact that both measures (UNWR and %SS) reflect phonological planning, whilst WFD is more of a vocabulary problem rather than an articulation one. The results also showed that monolingual English children and children with EAL did not differ in their performance on UNWR. Thus, accuracy in repeating non-words on the UNWR was not affected across language groups who showed different levels of %WWR. This again highlights that the test eliminated the problem in other NWR tasks that favor the language for which the test was designed. In summary, the UNWR has a strong potential as a screening instrument for language-diverse samples that can separate CWS from CWNS based on their accuracy in repeating non-words.

Whilst the reason for developing the AEN_NWR is to establish a stuttering screening instrument for preschool *children*, we argue that assessing adults is also necessary for the following reasons. First, when planning assessment and intervention for speech disorders, it is recommended that users are aware of weakness in phonological processing in older participants [e.g. adolescents in (29)]. Second, testing participants at older ages on their performance on NWR aids in capturing developmental differences in repetition accuracy (30). In relation to that, the previous studies presented generally showed that CWS were performing at ceiling levels for the shorter nonwords and at floor effect for the longer ones. Consequently, it would be of interest to evaluate whether ceiling and floor effects operate for AWS; and if they do, at which syllable length would AWS and CWS be differentiated. In fact, it is desirable when designing an NWR to have performance around the ceiling and floor to assess performance limits of participants on items that vary in difficulty

(25). Thus, in this section, we present a brief overview of studies that examined NWR performance in AWS. These studies have informed our hypothesis on differences between AWS and CWS on accurate repetition of NWR items.

Several researchers have examined NWR abilities in AWS and AWNS. Namasivayam and Van Lieshout (31) had five AWS and five AWNS repeat a set of non-words at two different rates (normal and fast) across three test session (two sessions on the same day, and a third session approximately one week later). AWS and AWNS differed in variables that concern the organizational aspects of speech motor control in terms of: (1) movement stability and (2) the strength of coordination patterns. Thus, the authors interpreted the results as AWS may have limited unique difficulty in the motor leaning of new sound sequences. Smith et al. (32) explored performance on a non-word repetition task in 17 AWS and a matched control group of 17 AWNS. The non-words ranged in length from one to four syllables; these were adapted from the Non-word Repetition Test [NRT; (33)]. Overall, there were no differences between the two groups. In fact, all participants performed at or near ceiling while repeating non-words of 1-,2-, and 3-syllables length. Only for non-words of 4-syllables and 5-syllables, the AWS scored lower than AWNS, although the results were still comparable. These findings could indicate that, unlike children, longer nonwords may be better at revealing differences between AWS and AWNS. Byrd et al. (34) also examined non-word repetition abilities in 14 AWS and a matched control group of 14 AWNS. Participants repeated non-words that consisted of 2-, 3-, 4- and 7-syllables. The two groups repeated the 2-, 3- and 4-syllable length non-words with comparable accuracy. Only for the 7-syllable length non-words, significant differences in accuracy were found between the two groups with AWS being less accurate in their production. Hence, it was suggested that it might be possible that non-words of at least 7 syllables may be needed to distinguish the two groups, whilst non-words at 2–3 syllable lengths are sufficient when to differentiate between CWS and CWNS Byrd et al. (34).

From the review above, it is clear that the available findings, both from children and adults are mixed. In terms of studies on CWS, it appears that there is a general agreement that CWS score lower than CWNS, but this does not hold true for all syllable lengths. Some studies provided information on how CWS and CWNS performed on NWR, overall and at each syllable length. However, other studies provided only general information on the performance of each group. This poses a limitation in comparing results from those studies. With respect to studies on AWS, there are several issues that are worth noting: (1) performance on non-word repetition tasks for AWS has received less attention; (2) Moreover, the methodology of those studies is not consistent. In some of the studies reviewed, NWR performance was part of a large study and there was not sufficient details on performance at each syllable length. Also, a general conclusion is that AWS might struggle repeating long non-words. This could be an indication that stuttering is associated with limited phonological working memory capacity. Thus, because the AEN_NWR is a newly designed task, we employed a sample of CWS and AWS to investigate group differences in performance at non-words that vary in length and phonological complexity.

To wrap up this section, it is worth noting that the NWR tasks employed in most studies were language specific, making them appropriate for English speakers only. Language-specific NWR tests are needed that test cohorts of speakers who use two or more languages (e.g., Arabic and English here) in an unbiased way.

Arabic Language

Arabic language has different phonological and morphological structures to English and the other Indo-European languages that UNWR applies to. Both English and Arabic use a pulmonic egressive airstream mechanism. This means that all the speech sounds of English and Arabic are produced using air from the lungs that exits the vocal tract (35). The two languages, however, still differ in many respects such as the phonemes that occur in each (this applies to consonants and vowels), stress and syllable constraints. Another aspect where the two languages differ is that the Arabic plural forms include singular, dual (i.e. referral to two objects or two persons) and plural while the English include two forms only; singular and plural. Additionally, in Arabic, stress depends on the syllable weight and it is more predictable than it is in English. The final syllable is stressed in cases where there is a long vowel (CVV) or where there is a word-final consonant cluster (CVCC), including geminate consonants. In other words, syllables with consonant clusters carry the main stress (36), which made it impossible to manipulate stress independent from consonant clustering when creating non-word stimuli. These differences, along with the phonotactic constraints that apply to both languages were examined below to generate the AEN_NWR stimuli. The steps involved in generating AEN_NWR stimuli that incorporate phonotactic constraints shared by English and Arabic to equitably test speakers of English and Arabic are described. The features of Arabic that governed design of AEN_NWR stimuli are described next.

Arabic is the most widely spoken Semitic language in the world and is used by more than 250 million people as their first language in the Middle East. Arabic is considered a Diglossic language (37). There is a Standard form, Modern standard Arabic (MSA) and a large number of regional Colloquial Arabic (CA) forms. MSA is used in most Arab countries in different situations and places including on street signs, in newspapers, Television news, university, schools and books. When spontaneous speech samples are collected from participants, speakers use their own CA dialect. However, when a passage is read, participants use MSA because CA forms are not usually used in written texts (although CA is used in, for example, text messaging). **Figure 2** summarizes the situations where MSA and CA are typically used.

Arabic has a complex morphology (38). It exhibits a discontinuous morphology that is based on the combination of the root and the word pattern (39). The root is exclusively made up of a sequence of consonants (usually three) that carry the core semantic information. The word pattern specifies the phonological structure and the morphosyntactic properties of the vowels, prefixes and suffixes that are then attached to the root to derive lexical meanings. In Arabic and in other Semitic languages, roots and word patterns are intertwined to form words across

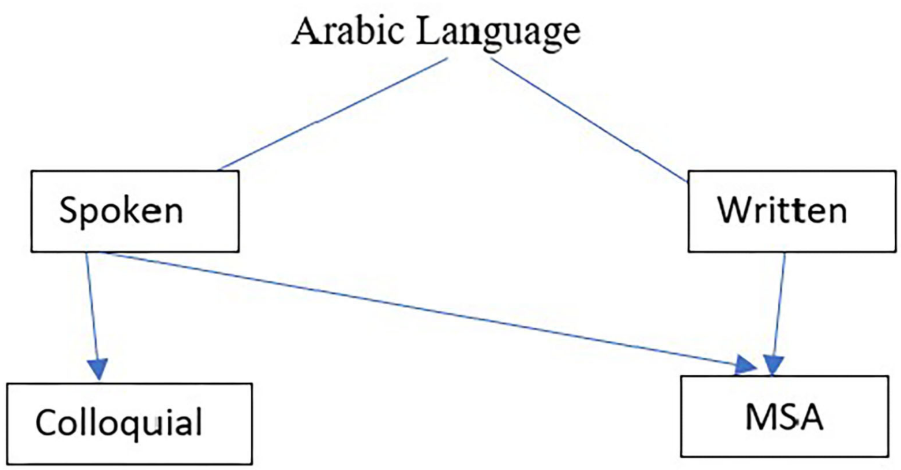


FIGURE 2 | Usage of MSA and CA.

TABLE 1 | Words derived from the /ktb/ root.

Word	Meaning
Kataba	He wrote
Katabat	She wrote
Kitaab	Book
Maktaba	Library; bookstore
Kaatib	Writer
Kutayyib	Booklet
Maktuub	Written

different lexical categories (e.g., nouns, verbs and adjectives). KTB is an example of a root that can be used to show how it implicates the meanings of the concept "write" (40). **Table 1** gives seven examples of distinct word forms derived from this root. The examples are from different lexical categories and involve different vowels and prefixes.

The Present Study

This study has two main goals. First, outlining a set of AEN_NWR stimuli that adheres to Arabic and English phonotactic constraints. Second, obtaining preliminary data on the effectiveness of the AEN_NWR in identifying stuttering in samples of AWS and CWS who speak Arabic as their first language. Specifically, we examined differences between CWS and AWS on accuracy of repeating non-words as the length of the non-words measured in syllables increased. It was expected that accuracy would decrease as non-words increased in syllable length for both groups; provided that non-words are of medium difficulty and thus ceiling and floor effects do not operate. Moreover, to provide support that the AEN_NWR is a reliable measure of stuttering, the empirical study reported here tested for a relation between AEN_NWR scores and the percentage of stuttered syllables (%SS). As advised in Riley's (41) manual, each stuttering instance was considered a single syllable and the %SS was obtained according to the manual. It was hypothesized that AEN_NWR scores would correlate with %SS in PWS, such that a lower AEN_NWR score would be associated with a higher %SS indicating higher levels of stuttering. Thus, results of stuttering assessment using a symptom-based procedure would be validated against AEN_NWR scores. Ideally, this preliminary testing should lead to further investigations of the AEN_NWR test items; and allow further refinement of the test until it is established as a sensitive measure that can equally assess Arabic- and English-speaking children.

MATERIALS AND METHODS

Participants

Ten CWS and fourteen AWS participated. Information on participants' age, gender and stuttering severity are given in **Table 2**. All participants had been previously evaluated by Speech and language pathologist (SLPs) as exhibiting stuttering behaviors. Participants spoke Arabic as their first language. None of the participants had unusual phonological processes that affected syllable structure. Also, none of the participants had neurological deficits. Demographics on the children's age, gender and the dialect of Arabic spoken were collected from children's parents and obtained directly from the older groups before the experiment started. Participants received reimbursement for participation. Ethics approval was granted by UCL's Institutional Review Board number 0078/006.

Stimuli: Considerations for Designing Arabic English NWR Stimuli

This section explains the steps taken to generate the AEN_NWR stimuli.

TABLE 2 | Participants in the study.

	Group	Gender	Age	Dialect	Treatment History	%SS	AEN NWR
1	Group 1	Male	6.5	Hijazi	Current	11	0
2		Male	9	Hijazi	Current	2	8
3		Male	9.0	Hijazi	Previous	2	13
4		Male	14	Kuwait	Previous	7	9
5		Male	11	Hijazi	Current	1.5	12
6		Male	11.5	Najdi	Current	6	9
7		Male	11.5	Gulf	Previous	3.5	6
8		Male	15.10	Omani	No treatment	11.5	0
9		Male	16.0	Egypt	Current	3	11
10		Male	16.7	Najdi	Current	0.8	21
1	Group 2	Female	19.2	Najdi	Previous	2.5	19
2		Female	20.1	Hijazi	Previous	7.5	3
3		Female	20.4	Najdi	No treatment	3.5	19
4		Male	22.0	Hijazi	Current	6.5	4
5		Female	23.1	Najdi	Previous	1	18
6		Female	24.12	Najdi	Previous	3	6
7		Female	24.8	Hijazi	Previous	7.5	2
8		Male	25	Najdi	No treatment	4	14
9		Female	25.1	Hijazi	Previous	6	1
10		Female	26.1	Najdi	Current	0.5	18
11		Male	29.4	Najdi	Previous	6.5	18
12		Female	31.0	Bahraini	No treatment	5	17
13		Male	21.9	Hijazi	Current	8	19
14		Male	21.9	Hijazi	Current	9	19

Step 1

Overlapping phonotactic constraints across English and Arabic were identified to create the non-word candidates. For each syllable template, all possible phone sequences were created according to the following constraints.

Syllable Patterns

Two syllable templates were selected that are permitted in both languages.

1.CV–A short open syllable (Consonant–vowel)

2.CVC–A medium closed open syllable (Consonant–vowel–consonant)

A final dull syllable was also employed word-finally alone to generate word-final C.C clusters. Since the second segment of the dull syllable occupies an onset and a syllable onset must be supported by a nucleus, that onset then must be followed by an empty nucleus, hence the name dull syllable (42).

In Arabic, the syllable is always initiated with a single consonant, which requires an obligatory onset. To emphasize, no word can start with a vowel in Arabic. The maximum number of consonants allowed in the onset position is one in MSA and in most CA dialects. In English, the "C" in the onset is optional (an example where the "C" is absent in the word "eye"). However, there is a strong preference in English for a syllable to begin with a consonant as zero onset syllables (Ø) are rare (43). Thus, all syllables generated are well-formed with one C in the onset. The selected templates were strung together to form polysyllabic non-words that systematically increased in phonological difficulty.

Word-initial clusters were not allowed (they are permitted in English but not Arabic dialects). For example, Hijazi Arabic, which is spoken in the west region of Saudi Arabia precludes word-onset clusters. The non-words were generated to be appropriate for a range of dialects of Arabic, so they are suitable for use with participants from diverse geographical and socioeconomic backgrounds. Therefore, care was taken that the phonotactic constraints were not specific to one Arabic dialect because speakers of any dialect are potential users of AEN_NWR. Sequences of more than two consonants do not occur in any syllable position in Arabic and a vowel is obligatory within a syllable except for a final dull syllable. Because of the conflicting points of view about the existence of a coda in Arabic, and about the phonotactic restrictions on the consonant clusters that appear word-finally (whether they constitute a coda, or they are just adjacent consonants that appear word-finally), only the phonotactic constraints that the cluster should adhere to are discussed.

All possible phone sequences were created for each syllable template, with the following additional constraints:

Consonants and Vowels Selection

To be included in the AEN-NWR, a consonant or vowel had to occur as a phoneme or an allophone in both languages. The specific consonants and vowels that were selected were as follows:

Consonants

Consonant phonemes that exist in English and Arabic are the following: /f/ , /b/, /d/, /m/, /n/, /s/, /z/, /k/, /g/, / ʃ /, /θ/ , /t/ /r/ and /l/ . The remaining consonant phonemes that exist in English or Arabic were not selected when the consonant phoneme does not have a corresponding phoneme in the other language. There are no corresponding phoneme for /tˤ/, /dˤ/ /ðˤ/, /sˤ/, /x/, /ɣ/ and /h/ in English. Also, there are no corresponding consonant phoneme for /p/, /v/ and /tʃ/ in Arabic. The consonant phoneme /t/ was selected although it is realized differently in the two languages. In Arabic, /t/ is dental involving simultaneous contact with the upper front teeth and the tongue tip and stopping the oral passage of air. In English, however, it is alveolar alone except when it occurs before dental fricatives (e.g., in words like "eighth" or between words like "at that") when it is dental. These are sub-phonemic differences, and the phoneme category is used in the NWR test, as it is one of the phonemes that has an early age of acquisition.

Consonant Acquisition. The age of acquisition of Arabic phonemes is generally similar to the corresponding ones in English, for common consonants although some allophonic variations between the two have been noted (44). In the latter study, a consonant was considered to have been acquired when at least 75% of children tested in each age group produced a consonant phoneme within a single word correctly in all positions of the word; initially, medially and finally. An Arabic-speaking child acquires /b/, /d/, /k/, /f/, /m/, /n/, /l/, /w/ in early childhood (2:0 to 3:10) and /s/ , /h/ and / ʃ / in later childhood (4:0 to 6:4) (44). Other consonants like /θ/ and /z/ are acquired later in Arabic (after 6:4) than in English. The acquisition criteria for English in the present study were adapted from Sander (45). According to Sander, children acquire /m/, /n/, /f/ and /w/ before the age of 3. The consonant phonemes /s/ and / ʃ / are acquired in later childhood at the age of 4.5. **Table 3** shows the age of acquisition for each of the selected consonant phonemes in both Arabic and English as reported in the studies that were mentioned above. It is important to note that these are average age estimates and the upper age limit in Sander (45) stops at an age level at which 90% of children are customarily producing the consonant phoneme. Information on this table were taken into account when designing non-words where consonants that are acquired earlier constituted most of the stimuli created making it possible for young children to produce them.

TABLE 3 | Age of acquisition in Arabic and English.

Consonant phoneme	Age of consonant acquisition in Arabic (44) and Bahakeem (46)	Age of consonant acquisition in English Sander (45).
/f/	(2;0 to 3;10)	(2;5 to 4;0)
/b/	(2;0 to 3;10)	(2;0 to 3;0)
/d/	(2;0 to 3;10)	(2;0 to 4;0)
/m/	(2;0 to 3;10)	(2;0 to 3;0)
/n/	(2;0 to 3;10)	(2;0 to 3;0)
/s/	(3;0 to 3;5)	(3;0 to 8;0)
/z/	(6;0 to 6;4)	(3;5 to 8;0)
/k/	(3;0 to 3;5)	(2;0 to 4;0)
/g/	(3;0 to 3;5)	(2;0 to 4;0)
/θ/	(3;0 to 3;11)	(4;5 to 7;0)
/ð*/	(4;0 to 4;5)	(5;0 to 8;0)
/j/	(2;6 to 2;11)	(4;0 to 7;0)
/t/	(2;6 to 2;10)	(2;0 to 6;0)

TABLE 4 | Short vowels mapping across Arabic and English.

Arabic short vowels	English short vowels
/ɪ/ Front high unrounded short	/ɪ/ Front high unrounded lax
/ʊ/ Back high unrounded short	/ʊ/ Back high unrounded lax
/a/ Central low unrounded short	/æ/ Front low unrounded

Vowels

Three short vowels were selected [i, a, u] because each has an equivalent, or a near equivalent, that can be mapped across languages as shown in **Table 4**. Thus, a speaker of either of the two languages is not expected to have difficulty perceiving and then producing those vowels. Long vowels, however, were excluded because of differences between the two languages that may result in the speaker having difficulty perceiving contrasting forms because of the absence of some vowels from the participant's first language. Vowels in Arabic vary little among speakers of different Arabic dialects which employ the three short vowels that were listed above and another three long vowels. There are cases where long vowels might be analyzed as a sequence of two nuclei rather than a single branching nucleus (47). English has a larger number of vowels that do not have an equivalent in Arabic and thus may be problematic for an Arabic speaker to identify and produce. The quality of English long vowels also varies considerably between English accents. AlShanqiti (48) investigated how Saudi Arabic learners of English perceive and produce English vowels. She examined the problematic phonemic contrasts for learners of British English. The results showed that vowels that do not have counterparts in Arabic were more challenging for Arabic listeners to recognize. Also, Shafiro et al. (49) investigated the perception of American English vowels and consonants by native Arab speakers and Arab-English bilinguals. Vowel perception was less accurate than consonant perception in both groups. The authors attributed low accuracy in perceiving vowels to the bilingual participants' mapping of the larger Arabic English inventories to the smaller inventories of Arabic vowels. Moreover, a speaker might substitute one phoneme with another according to the speaker's first language. For example, in an Arabic word such as /ka:n/ (where the vowel is front low unrounded and long), English speakers would be expected to assimilate to the nearest vowel, which is the back low unrounded vowel [α] such as in the word /calm/ as reported in a study by Huthaily (35) where the difficulties in producing vowels for English learners of Arabic were examined. Huthaliy presented participants with the Arabic word [ra:tib] where the long vowel is front unrounded and long. All participants substituted the long vowel with [α] and used [α] to substitute for [a:] whenever it occurred. Since long vowels are commonly substituted, long vowels were excluded. For AEN_NWR, short vowels were included because they only differ in narrow phonetic details that should not be problematic for children or adults to recognize when listening to stimuli.

Phonotactic Constraints

The following phonotactic constraints in both languages were implemented to ensure that the non-words that were created abide by the constraints of both languages.

Coda-Onset Clusters

"Coda" refers to a word-internal consonant, not a word-final consonant. This applies when using a syllable template with a coda like CVC for creating a non-word with two syllables or more, for example: CVC.CV, CVC.CVC. Word-internal codas are restricted by the following six constraints: (1) A coda must be a sonorant /m/,/n/,/w/, /r/,/l/ or an obstruent /k/ or fricative /s/, /f/ ; (2) A post-coda onset must be a plosive /t/, /k/, /d/, /g/, /b/; (3) No geminates are allowed as they are not allowed in English; (4) A nasal must be homorganic with a following onset: [mb, nt, nd]; (5) [s] can only appear before a plosive voiceless onset: [st, sk, sf]; and (6) [r] can appear word-finally in General American English but not in non-rhotic accents such as British English hence its usage is restricted to onset positions alone. All of the clusters allowed by these restrictions are shared by Arabic and English.

Word-Final Clusters

As mentioned above, the phonotactics of the word-final consonants in English are similar to those of the internal coda. In other words, a word-final cluster behaves like an internal C.C because it is also a coda onset cluster. Consequently, because of the parallelism between the two domains, the phonotactic constraints need to be stated once only (42). Generally, the two phonemes in the word-final consonant cluster in English must conform to the sonority sequencing phonotactic principle (SSP). SSP was first introduced by (50) and it aimed to characterize the syllable structure in terms of sonority. Thus, in a C.C coda cluster, the first coda consonant should be higher in sonority relative to the second coda consonant. There are exceptions when word-final pairs of consonants violate the SSP. These occur in a sequence of two stops that are not homorganic (e.g., act), and a sequence of a stop + /s/ (e.g., lapse and tax) (51). Moreover, there are final sequences that show no evidence of phonotactic constraints which are usually generated by suffixation. In Arabic, however, the SSP is not a reliable predictor of the sequence of two consonant phonemes that occur word-finally. There are many Arab words, in MSA and in several other Arabic dialects that violate the SSP. AlTamini and AlShboul (52) conducted a study of MSA coda cluster phonotactics to assess applicability of SSP. They assumed that any consonant that follows the last vowel of a word belongs to a coda. The authors used a sample of around 500 CVCC lexical items that were collected from The Hans Wehr Dictionary of Modern Written Arabic. It is important to note here that looking at the words that were selected from the dictionary, many of them are nouns of high frequency and they are usually pronounced the same way in MSA and Arabic dialects as in /ˁaks/ "reverse", /s'ubh/ "morning", /wadJh/ "face". The results showed that contrary to what is widely reported in the literature on the compliance of the CC coda with SSP, this was only true for 42% of the cases. The remaining 58% of the cases violated the SSP sonority hierarchy as follows: (1) reversal (49% of the C.C coda clusters showed a rise in sonority in which the coda first consonant had lower sonority relative to the coda second consonant); or (2) plateau in which the C.C cluster consonants are of almost equal sonority. The results challenge the fact that the phonotactics of the Arabic C.C coda are sonority based because SSP was violated in more than 50% of cases. The authors raised the idea of re-considering a more theoretical model outside the scope of SSP that has been long thought of as governing complex coda syllables.

Step 2

Non-word candidates were selected for each syllable length as follows: First, all permitted syllable combinations which could serve as templates were created for each syllable length; 100 of these templates per syllable length were then selected at random. Next, consonant and vowel phones were selected at random, apart from those barred by the constraints listed above (i.e. consonants and vowel selection). The selected phones were entered into the template for each syllable length. Finally, individual syllables were combined. Given all the constraints above, the AEN_NWR test should meet the goals of including segments that a speaker of Arabic or English can pronounce and do so in comparable ways across these languages provided that the speaker is equally fluent in both languages. In the study, even when phonologically complex materials for the two languages are tested, testing starts with "easy" materials with minimal use of clusters. Then, complexity is increased systematically in terms of the number of segments and syllable structures. **Appendix 1** shows the orthographic transcription of the final two, three and four syllable AEN_NWR stimuli.

Step 3

Stress, as a linguistic phenomenon, occurs in both languages. The two languages are similar in terms of the association between stress and heavy syllables; Arabic and English are both quantity-sensitive languages where heavy syllables attract stress (53, 54). In order to meet the phonological requirement of testing across the two languages, language-specific stress contrasts were avoided by producing all syllables with equal stress, except for the CVCC heavy syllables which mark the end of some stimuli. Additionally, another set of identical non-word unstressed syllables was developed where short vowels were reduced to schwa; hence the vowels were homogenized (Note: this set was not used in the current experiment). This was done because there are differences in vowel quality and stress between Arabic and English. When a participant's repetition of an AEN_NWR stimulus was evaluated, any differences in produced stress patterns or vowel quality were ignored and only consonants were considered in scoring.

Step 4

Checks were made to ensure none of the phone sequences are words in either of the languages using Aralex (55). Aralex is a lexical database for MSA that provides token frequencies of roots and word patterns that integrates information from two sources: (1) a 40 million word corpus derived from different newspapers covering various topics such as politics, sport and culture; and (2) a dictionary compromised of 37,494 entries that provides information and token and type frequencies of Arabic words and morphemes. Although Aralex was created using MSA rather than a spoken dialect of Arabic, it still meets the requirements for a lexical database that can be checked for lexicality effects for the following reasons. (1) Despite the existence of many dialects across different Arabic countries, speakers of the language have a single inventory of phonemes. (2) MSA and spoken Arabic forms present with different phonological, syntactic and lexical systems where each fulfills distinct sociolinguistic functions. The database was built from MSA using text from newspapers but when dialectal words were found they were retained in the corpus but flagged as such (55). It is worth noting here that Arab dialects are rarely written and they are mostly used for speech communication alone. The Aralex database (2010) has a user-friendly interface that consists of 12 boxes where filled boxes can be used to input a search string. The user has the option of displaying the results in English or Arabic Unicode. For example, the orthographic form window takes as input an Arabic or English script and displays the selected results either in Arabic or English.

Experimental Procedure

Participants were tested in a quiet setting in one session of approximately 15 min. The experimenter conducted two tasks: (1) elicitation of spontaneous speech samples and (2) administration of the AEN_NWR. The study used a within-participants design as participants completed both tasks; and the two tasks were conducted in randomized order. The complete session was recorded on a Sony DAT audio-recorder using a Sennheiser K6 microphone and audacity software. For the spontaneous speech, samples of 200 syllables from all participants were obtained. These were elicited during conversational speech with the researcher using topics of interest, such as school, travel, books and hobbies. When necessary, picture material from Riley (6) was also used to elicit speech from children. For the AEN_NWR, the 28 non-words listed in **Appendix 1** were used as follows. All stimuli from the three sets (i.e., two syllables, three syllables and four syllables) were presented in standard order, in which stimulus length systematically increased. Participants were exposed to all stimuli regardless of their performance. Participants were informed that they would hear made-up words; i.e., ones that do not exist either in Arabic or English. The examiner then gave the following instructions to the participant: "I am going to play some made-up words to you through the headphones and I want you to repeat them as accurately as you can. You will have to listen carefully because you will only hear them once". Participants were allowed as much time as was necessary to respond. There were two practice items to make sure the output volume was appropriate, and the participant understood the nature of the task. The non-words were pre-recorded to ensure that factors such as differences in word stress patterns and accent would not affect the results. Recording of materials the participants heard took place in an anechoic chamber and were obtained from a male professional phonetician who was phonetically trained in Arabic and English[2].

Data Processing, Scoring and Reliability

Spontaneous Speech

The audio recordings of the participants' speech were orthographically transcribed after replaying them as many times as necessary by the researcher. The percentage of stuttered syllables (%SS) from the first 200 syllables were obtained following guidelines from Riley (6). Todd et al. (56) confirmed that a 200-syllable long speech sample was sufficient to obtain a reliable SSI score. Guidelines from the original SSI-4 (6) were followed to score the first 200 syllables of all samples collected. %SS was calculated by determining the number of stuttered syllables and the total number of syllables. The number of stuttered syllables was then divided by the total number of syllables and multiplied by 100, resulting in the percentage of stuttered syllables %SS. A syllable was stuttered if the speaker exhibited one of the three disfluency characteristics: (1) repetition indicated by multiple repetition of a sound or a syllable that was not a word; (2) prolongation indicated by abnormal lengthening during the production of a phoneme; (3) break within a syllable.

AEN_NWR Scoring

The recorded non-words were scored offline at the whole-word level. The non-word phonemic response transcriptions and target transcriptions for each non-word were aligned to identify how the two transcriptions differed. A correct response occurred when all consonants in each non-word were pronounced correctly. Errors on consonant production included deletion, substitution or insertion.

Responses were scored as "incorrect" if they contained one or more consonant errors. Vowels were ignored during scoring because of the subtle differences between dialects. To ensure reliability, responses from 20% of the participants, selected randomly, were scored independently by a second trained phonetician. Agreement on the number of correct stimuli was 88%. Data from the first author was used in statistical analysis.

RESULTS

Performance of CWS and AWS on the AEN_NWR

To review, we were interested in examining the effect of age group and non-word length on accuracy of repeating non-words. A repeated Measures ANOVA with the between-participants factor of Group (CWS vs. AWS) and a within-participants factor of Syllable Length (2-, 3, and 4-syllables), was conducted. The

[2]The authors would like to thank Christopher Lucas, SOAS University of London, for recording the non-word stimuli. The recorded materials can be.

dependent variable was the mean % of correctly repeated non-words. Results revealed a main effect for Syllable Length F (2,44) =11.84, P<0.000), partial $\eta_p^2 = 0.4$. Overall, there was no significant between-participant difference F (1,22) = 3.1, p = 0.09, partial $\eta_p^2 = 0.1$.

The differences between the CWS and AWS were inspected at each syllable length using independent samples *t*-tests. Significant difference between the two groups were found for non-words at the 4-syllable length only F (1,22) = 7.60. $p = 0.01$). No significant differences were noted at the 2-syllable, or 3-syllable lengths. **Figure 3** summarizes the results.

Validating AEN_NWR Against Symptom-Based Stuttering Measure

Simple linear regression was used to assess whether AEN_NWR scores predicted %SS scores. A correlation analysis using Pearson's r can be used since the variables were measured on a continuous scale. A significant negative correlation between the AEN_NWR and %SS scores was found using Pearson's product-moment correlation coefficient (r (25) = −0.5), $p < 0.000$), indicating that higher AEN_NWR scores were associated with lower %SS, making AEN_NWR scores a significant predictor of %SS. Performance of each group was inspected again in a separate analysis. A significant negative correlation between the AEN_NWR and %SS for CWS was found (r (10) = −0.9. $p < 0.000$). However, the correlation between AEN_NWR and %SS was not significant for AWS ($p > 0.05$). **Figure 4** summarizes the relationship with separate panels for AWS and CWS.

DISCUSSION

The purpose of this paper was to describe how a set of stimuli for an Arabic-English non-word repetition (AEN_NWR) task were developed, and to examine preliminary data from a small clinical sample of Arabic-speaking CWS and AWS. The study was motivated by Howell et al.'s (9) finding that showed that the "Universal" NWR (UNWR) can identify stuttering in preschool children, irrespective of their first language. However, the UNWR does not apply to Semitic languages such as Arabic. This issue was addressed in this paper by constructing AEN_NWR items that obeyed the phonotactic constraints of Arabic and English. The test had three groups of items that varied in length between two and four syllables. Based on the literature on phonological working memory and stuttering, we expected that, performance on AEN_NWR should decrease as the number of syllables in the stimuli increased. Because performance on the NWR develops with age (23), we recruited clinical samples that comprised CWS and AWS to examine developmental changes. Finally, we validated overall AEN_NWR accuracy by comparing participants' scores against a standard symptom-based procedure, the percentage of stuttered syllables %SS.

Accuracy in Repeating AEN_NWR Items

Regardless of age, non-word length affected accuracy of performance. This was demonstrated by the main effect of syllable length. The 2-syllable non-words were repeated more accurately than 3-syllable non-words, and 3-syllable non-words were repeated more accurately than 4-syllable non-words. These findings add evidence about the effect of phonological memory on individuals who stutter [e.g., (24, 34)]. As non-words increase in length, they take additional time to be perceived and repeated, and consequently their phonological representations may be prone to decay before they can be rehearsed and articulated (57).

As the AEN_NWR is a new measure, we performed separate analysis for CWS and AWS to examine the effect of age on accuracy of performance. Although AWS consistently performed better at all syllable lengths, significant differences were reached only at the 4-syllable length non-words. These findings could be interpreted as preliminary evidence that non-words of at least 4-syllable length are required for the AEN_NWR to be sensitive to age differences. However, it is too early to make such a generalization without increasing participant numbers in both groups. Having a large number of participants from a wide age range could also inform decisions such as: (1) how many items are required at each syllable length; and (2) at what syllable length do CWS and AWS show differences in performance. Ebert et al. (58) noted that non-word stimuli should include enough variants to capture the wide range of non-word repetition ability levels. Hence, the variation in average performance on the AEN_NWR across ages could inform design considerations of the task to make it suitable to test preschool children.

Preliminary Validation of AEN_NWR

Evidence was provided that the AEN_NWR is a reliable measure of phonological skills and hence of fluency of speech. It was hypothesized that, overall, AEN_NWR scores should correlate negatively with %SS (i.e., percentages of stuttered syllables, %SS). This hypothesis was partially supported; it appears that participants who showed a higher percentage of stuttered syllables/dysfluent events scored lower on the AEN_NWR. These findings are consistent with the results by Howell et al. (9) who showed a relationship between %SS and UNWR scores. Thus, this provides support that NWR as an established measure of phonological skills for participants with stuttering symptoms. The correlation with the %SS could be interpreted as an indication that the test has good potential for identifying preschool children with speech disfluency. However, such a conclusion need qualifying until the current results are compared with results of control groups of adults who do not stutter and children who do not stutter. The importance of examining developmental differences between AWS and CWS is linked to the ceiling and floor effects, which are common to NWR tasks (25). Ceiling effects operate for non-words of short syllable lengths and floor effects operate for the longer ones, and this can limit the ability to detect individual differences Munson (59).

When the two age groups used in this study were examined separately, the correlation coefficient for CWS as a group was

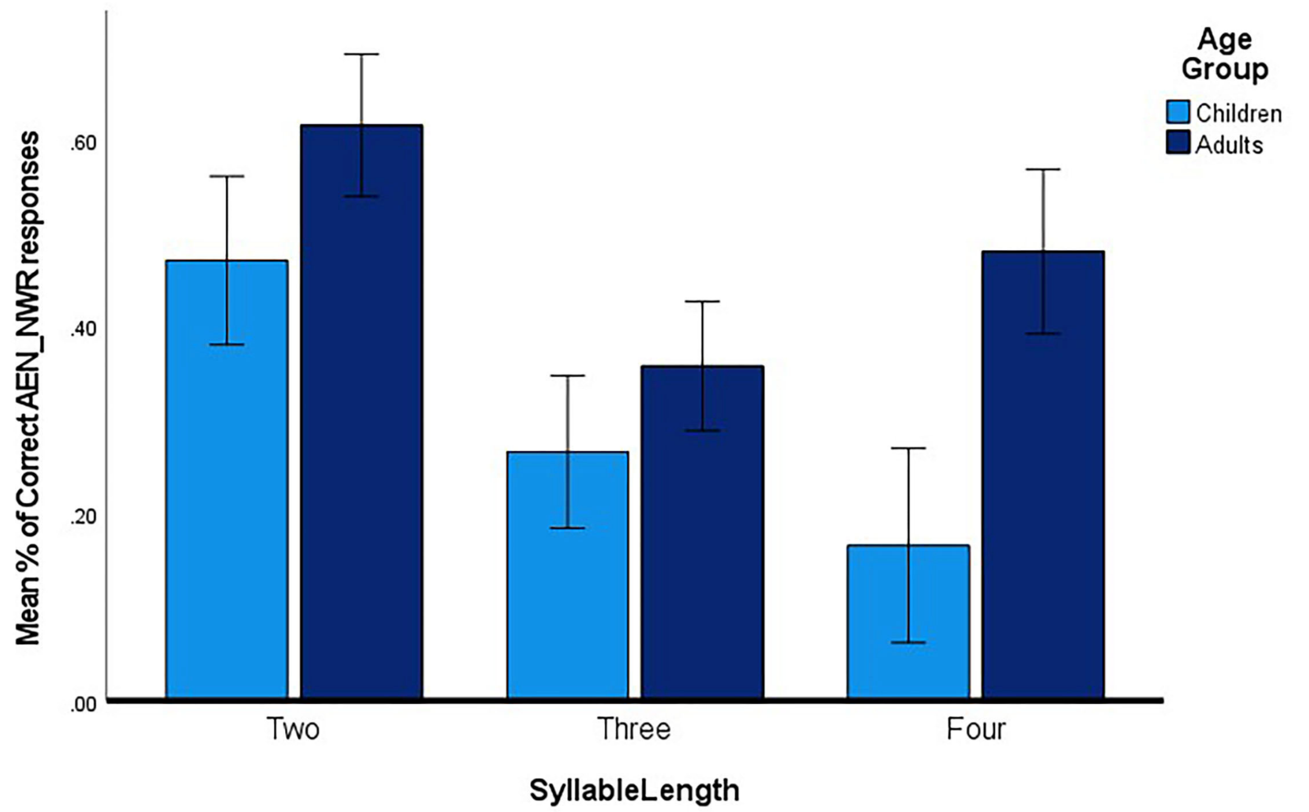


FIGURE 3 | The mean % of accurate productions on the AEN_NWR for producing the non-word at each syllable length for CWS and AWS. Age group is indicated in the inset.

higher than that for the AWS. This finding could be attributed to the fact that some AWS had severe stuttering symptoms. That is, their mean %SS was 4.94; SD 3.1 and seven AWS scored above the mean. This potentially created a skewed sample, which in turn had an effect on the relationship with AEN_NWR. It is worth mentioning that many of the AWS had therapy after their teenage years but despite that their stuttering persisted, although they continued receiving speech therapy services to improve their fluency. Moreover, the fact that AWS showed better performance than CWS could be interpreted as an indication of floor effects operating, particularly for the longer syllables. In fact, ceiling and floor effects may be complicating factors for many studies on NWR accuracy (54). Also in the stuttering population, longer nonwords might pose an additional challenge for CWS relative to AWS. This is particularly true because scoring is done at the word level rather than phonemic level, another factor that could obscure individual differences in performance.

CONCLUSION

The results indicated that the proposed non-word stimuli of the AEN_NWR were broadly appropriate for the stuttering population tested. Floor effects appeared to be influencing performance of CWS, but as mentioned, these are common in NWR tasks. The design consideration behind the AEN_NWR involved eliminating any influences that language history could have on an individual's performance. At this stage, only Arabic speakers were tested, but future work needs to evaluate the performance of English speakers and to compare the two groups (including people who are fluent or who stutter). The proposed set of stimuli conforms to accepted standards for NWR tasks including the following: language-specific phonotactic constraints (Arabic and English), avoiding later-developing consonants, and minimizing potential resemblance between real words and non-words. Furthermore, the test does not require knowledge of lexical semantics for either of the two languages. This aids in reducing any biases that speakers are relying on existing semantic knowledge from their stored language memory.

Limitation and Future Direction

This paper makes a novel contribution to the subject area of stuttering as an aspect of language disorder, and to screening procedures of stuttering; particularly in the Arabic language, in which the literature is sparse. To the best of

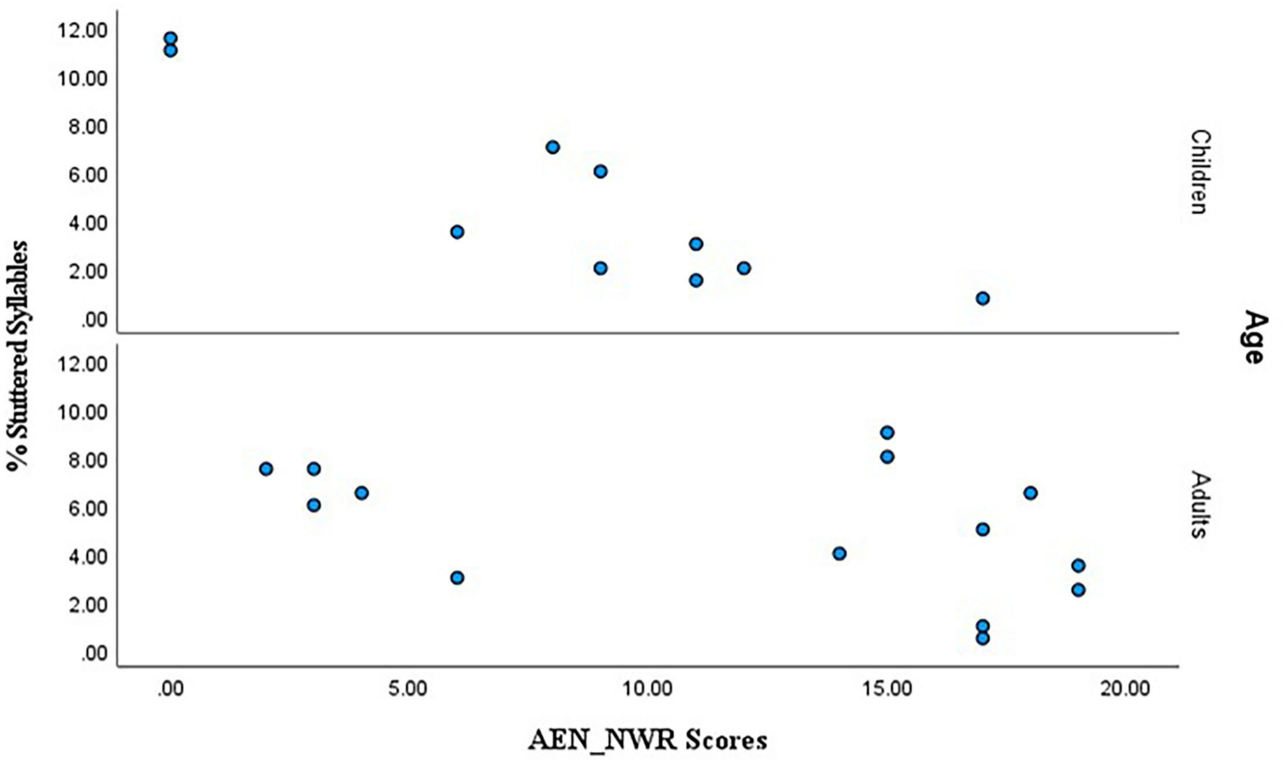


FIGURE 4 | Scatter plot of raw %SS (frequency scores) (Y-axis) against total AEN_NWR scores (X-axis) with panels for CWS and AWS at top and bottom.

our knowledge, this is the first time that non-word repetition abilities have been investigated in Arabic speakers in stuttering research. While the current preliminary data provide promising results, there are some limitations that should be raised. First, all participants in this study come from a clinical sample and had been diagnosed as displaying stuttering symptoms; However, different SLPs may use different guidelines to diagnose stuttering due to the lack of norms for assessing fluency in Arabic. The speech samples obtained in this study were in Arabic as the aim was to develop and standardize an instrument to assess fluency for Arabic preschool children; however for analysis of speech symptoms, the guidelines of SSI were followed with respect to what is counted as a stuttering symptom and SSI was developed for English PWS. Although a strong correlation was found between the %SS and the AEN_NWR, which might suggest that the guidelines can be generalized to Arabic, having guidelines designed specifically to Arabic could result in even a stronger correlations. To explain in more detail, the data used to evaluate SSI statistically were collected from English speakers who stuttered, and it used English passages that had to be read as well as others that were spoken spontaneously. Consequently, the norms do not apply to the Arabic version as the standardization has not been conducted. Moreover, the test is not appropriate for assessment of Arabic in terms of the procedures used in SSI for counting the number of production units (i.e. syllables for English) as well as the specifications of stuttered events. Thus, in parallel to standardizing the AEN_NWR, work is ongoing to establish clear guidelines for assessing spontaneous speech samples in Arabic. This includes instructions on syllables counts and disfluency counts. Second, consistency of the results should be examined by test-retest reliability, which is a necessary measure for an NWR to be considered reliable for diagnostic purposes (60). Third, probably the biggest limitation of this study concerns the number of participants. Although the sample size is reasonable compared to other studies that examined the non-word repetition abilities in PWS [e.g., Hakim and Ratner (24) had eight CWS and Sugathan and Maruthy (28) had 17 CWS], it is important to replicate the results of this preliminary study using larger number of participants. This is particularly the case because participants in this study were from different age groups and there may be developmental differences between them. Having more participants could also permit assessing phonological performance at every non-word length to examine the relationship between repetition accuracy and phonological complexity.

Finally, it would be of interest to examine matched groups of: (1) controls of Arabic speakers who do not stutter; (2) English speakers who stutter and examine their performance on the AEN_NWR. Arabic speakers of dialects other than Saudi could also be examined as additional comparison cohort; this would allow investigating the influence of dialect on AEN_NWR performance.

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 UCL Ethics Research Committee. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

PH and RA designed the study, performed the research, analyzed the data, and wrote the manuscript. PH, JH, and RA developed the non-word stimuli. SB assisted in recruiting participants. All authors contributed to the article and approved the submitted version.

FUNDING

The authors extend their appreciation to the King Salman Centre for Disability Research for funding this work through Research Group No. KSRG-2022-001.

REFERENCES

1. Andrews G, Harris M. *The syndrome of stuttering. Spastics Society Medical Education*. (1964).
2. Yairi E, Ambrose NG. *Early Childhood Stuttering*. Austin, TX: Pro-Ed. (2005).
3. Snow P. Oral language competence and the transition to school: Socio-economic and behavioural factors that influence academic and social success. *International Journal on School Disaffection*. (2014) 11:3–24.? doi: 10.18546/IJSD.11.1.01
4. Bercow J. *The Bercow Report: A Review of Services for Children and Young People (0–19) with Speech*. Language and Communication Needs. Nottingham: DCSF Publications. (2008)
5. Howell P. Screening school-aged children for risk of stuttering. *J Fluency Disord.* (2013) 38:102–23. doi: 10.1016/j.jfludis.2012.09.002
6. Riley G. The stuttering severity instrument for adults and children (SSI-4). Austin, TX. (2009).
7. Howell P, Chua LY, Yoshikawa K, Tang HHS, Welmillage T, Harris J, et al. Does working-memory training given to reception-class children improve the speech of children at risk of fluency difficulty?. *Front Psychol*. (2020) 11. doi: 10.3389/fpsyg.2020.568867
8. Mirawdeli A, Howell P. Is it necessary to assess fluent symptoms. duration of dysfluent events, and physical concomitants when identifying children who have speech difficulties? *Clini Linguist Phonet.* (2016) 30:696–719. doi: 10.1080/02699206.2016.1179345
9. Howell P, Tang K, Tuomainen O, Chan K, Beltran K, Mirawdeli A, et al. Identification of fluency and word-finding difficulty in samples of children with diverse language backgrounds. *Int J Lang Commun.* (2017) 52:595–611. doi: 10.1111/1460-6984.12305
10. Zangl R, Mills DL. Increased brain activity to infant-directed speech in 6-and 13-month-old infants. *Infancy.* (2007) 11:31–62. doi: 10.1207/s15327078in1101_2
11. Gathercole SE. Non-word repetition and word learning: The nature of the relationship. *Appl Psycholinguist.* (2006) 27:513–43. doi: 10.1017/S0142716406060383
12. Coady JA, Evans JL. Uses and interpretations of non-word repetition tasks in children with and without specific language impairments (SLI). *Int J Lang Commun.* (2008) 43:1–40. doi: 10.1080/13682820601116485
13. Munson B, Kurtz BA, Windsor J. The influence of vocabulary size, phonotactic probability, and wordlikeness on nonword repetitions of children with and without specific language impairment. *J Speech Lang Hear Res.* (2005) 48:1033–47. doi: 10.1044/1092-4388(2005/072)
14. Gathercole SE, Baddeley AD. Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. *J Mem Lang.* (1989) 28:200–13. doi: 10.1016/0749-596X(89)90044-2
15. Snowling MJ. Phonemic deficits in developmental dyslexia. *Psychol Res.* (1981) 43:219–34. doi: 10.1007/BF00309831
16. Archibald LM, Gathercole SE. Non-word repetition in specific language impairment: More than a phonological short-term memory deficit. *Psychonomic Bulletin and Review.* (2007) 14:919–24. doi: 10.3758/BF03194122
17. Broom YM, Doctor EA. Developmental phonological dyslexia: A case study of the efficacy of a remediation programme. *Cogn Neuropsychol.* (1995) 12:725–66. doi: 10.1080/02643299508251400
18. Snowling MJ. From language to reading and dyslexia 1. *Dyslexia.* (2001) 7:37–46. doi: 10.1002/dys.185
19. Baddeley A. *Working Memory / Alan Baddeley: Oxford Psychology Series, No. 11*. Oxford: Oxford Press (1986).
20. Baddeley AD, Hitch G. Working memory. In: *Psychology of learning and motivation.* Academic press. (1974).p. 47–89. doi: 10.1016/S0079-7421(08)60452-1
21. Baddeley AD. The concept of working memory. *Models of Short-Term Memory*. (1996) 1–27. doi: 10.1093/acprof:oso/9780198570394.003.0001
22. Gathercole SE. Working Memory. In: Byrne J, editors. Learning and Memory: A Comprehensive Reference, 2, 1st Edition, 33–51. Oxford: Academic Press. (2008). doi: 10.1016/B978-012370509-9.00179-0
23. Gathercole SE, Willis CS, Baddeley AD, Emslie H. The children's test of non-word repetition: A test of phonological working memory. *Memory.* (1994) 2:103–27. doi: 10.1080/09658219408258940
24. Hakim HB, Ratner NB. Non-word repetition abilities of children who stutter: an exploratory study. *J Fluency Disord.* (2004) 29:179–99. doi: 10.1016/j.jfludis.2004.06.001
25. Anderson JD, Wagovich SA, Hall NE. Nonword repetition skills in young children who do and do not stutter. *J Fluency Disord.* (2006) 31:177–99. doi: 10.1016/j.jfludis.2006.05.001
26. Anderson JD, Wagovich SA. Relationships among linguistic processing speed, phonological working memory, and attention in children who stutter. *J Fluency Disord.* (2010) 35:216–34. doi: 10.1016/j.jfludis.2010.04.003
27. Sasisekaran J, Byrd C. Non-word repetition and phoneme elision skills in school-age children who do and do not stutter. *Int J Lang Commun Dis.* (2013) 48:625–39. doi: 10.1111/1460-6984.12035
28. Sugathan N, Maruthy S. Non-word repetition and identification skills in Kannada speaking school-aged children who do and do not stutter. *J Fluency Disord.* (2020) 63:105745. doi: 10.1016/j.jfludis.2019.105745
29. Preston J, Edwards M. Phonological processing skills of adolescents with residual speech sound errors. *Lang Speech Hear Serv Sch.* (2007) 38:297–308. doi: 10.1044/0161-1461(2007/032)
30. Windsor J, Kohnert K, Lobitz KF, Pham GT. Cross-language nonword repetition by bilingual and monolingual children. *Am J Speech-language Pathology.* (2010) 19:298–310. doi: 10.1044/1058-0360(2010/09-0064)
31. Namasivayam AK, Van Lieshout P. Investigating speech motor practice and learning in people who stutter. *J Fluency Disord.* (2008) 33:32–51. doi: 10.1016/j.jfludis.2007.11.005

32. Smith A, Sadagopan N, Walsh B, Weber-Fox C. Increasing phonological complexity reveals heightened instability in inter-articulatory coordination in adults who stutter. *J Fluency Disord.* (2010) 35:1–18. doi: 10.1016/j.jfludis.2009.12.001
33. Dollaghan C, Campbell TF. Nonword repetition and child language impairment. *Journal of Speech, Language, and Hearing Research.* (1998) 41:1136–46. doi: 10.1044/jslhr.4105.1136
34. Byrd CT, Vallely M, Anderson J, Sussman H. Non-word repetition and phoneme elision in adults who do and do not stutter. *J Fluency Disord.* (2012) 37:188–201. doi: 10.1016/j.jfludis.2012.03.003
35. Huthaily K. *Contrastive Phonological Analysis of Arabic and English [Master's Thesis]*. University of Montana. (2003).
36. Shaalan S. *Investigating Grammatical Complexity in Gulf Arabic Speaking Children with Specific Language Impairment [Doctoral thesis]*. University College London. (2010).
37. Ferguson CA. Diglossia. *Word.* (1959) 15:325–40. doi: 10.1080/00437956.1959.11659702
38. Abu-Rabia S. The role of morphology and short vowelization in reading morphological complex words in Arabic: Evidence for the domination of the morpheme/root-based theory in reading Arabic. *Creative Education.* (2012) 3:486. doi: 10.4236/ce.2012.34074
39. Boudelaa S, Marslen-Wilson WD. Discontinuous morphology in time: Incremental masked priming in Arabic. *Lang Cogn Process.* (2005) 20:207–60. doi: 10.1080/01690960444000106
40. Tucker MA. *The morphosyntax of the Arabic* verb: Toward a unified syntax-prosody. (2011).
41. Riley G. Stuttering severity instrument for young children (SSI-3). *Stuttering Severity Instrument for Young Children (SSI-3)*. (1994).
42. Harris J, Gussmann E. Word-final onsets. *UCL Working Papers in Linguistics.* (2002) 14:1–42.
43. Harris J. *English Sound Structure*. Oxford, Blackwell. (1994).
44. Amayreh MM, Dyson AT. The acquisition of Arabic consonants. *J Speech, Language, and Hearing Research.* (1998) 41:642–53. doi: 10.1044/jslhr.4103.642
45. Sander EK. When are speech sounds learned? *Journal of speech and hearing disorders. J Speech Lang Hear Res.* (1972) 37:55–63. doi: 10.1044/jshd.3701.55
46. Bahakeem M. *Phonological Development In: Saudi-Arabic-Speaking Children*. Unpublished Doctoral thesis. UCL (University College London). (2016).
47. Kaye J. Phonology: a cognitive view. Routledge. Tucker MA. The morphosyntax of the Arabic verb: Toward a unified syntax-prosody. (2013). doi: 10.4324/9780203056806
48. Alshangiti WMM. Speech production and perception in adult Arabic learners of English: A comparative study of the role of production and perception training in the acquisition of British English vowels. [Doctoral thesis]. University College London. (2015).
49. Shafiro V, Levy ES, Khamis-Dakwar R, Kharkhurin A. Perceptual confusions of American-English vowels and consonants by native Arabic bilinguals. *Language and speech.* (2013) *56*:145–61. doi: 10.1177/0023830912442925
50. Sievers E. *Grundzüge der phonetik (Fundamentals of phonetics)*. Leipzig: Breitkopf und Hartel (1881).
51. Yavas M, Babatsouli E. Acquisition of /s/-clusters in a Greek-English bilingual child: Sonority or OCP? In: Ball MJ, Müller N, editors. *Sonority Across Languages*. Sheffield, UK: Equinox Publishing. (2016) pp. 337–354.
52. Al Tamimi YA, Al Shboul Y. Is the phonotactics of the Arabic complex coda sonority-based? *J King Saud Univ Sci.* (2013) 25:21–33. doi: 10.1016/j.jksult.2012.12.003
53. Watson JCE. Word stress in Arabic. In: *The Blackwell companion to phonology*. Wiley-Blackwell, Oxford. (2011) 2990–3019. doi: 10.1002/9781444335262.wbctp0124
54. Archibald J. *The acquisition of English stress by speakers of nonaccentual* languages: Lexical storage versus computation of stress. *Linguistics.* (1997) 35:167–82. doi: 10.1515/ling.1997.35.1.167
55. Boudelaa S, Marslen-Wilson WD. Aralex: A lexical database for modern standard Arabic. *Behav Res Methods.* (2010) 42:481–7. doi: 10.3758/BRM.42.2.481
56. Todd H, Mirawdeli A, Costelloe S, Cavenagh P, Davis S, Howell P. Scores on Riley's Stuttering Severity Instrument versions three and four for samples of different length and for different types of speech material. *Clinical Linguistics and Phonetics.* (2014) 28:912–26. doi: 10.3109/02699206.2014.926991
57. Archibald LM. The promise of non-word repetition as a clinical tool. *Canadian Journal of Speech-Language Pathology and Audiology.* (2008) 32:21–8.
58. Ebert KD, Kalanek J, Cordero KN, Kohnert K. Spanish non-word repetition: Stimuli development and preliminary results. *Commun Disord Q.* (2008) 29:67–74. doi: 10.1177/1525740108314861
59. Munson B. Nonword repetition and levels of abstraction in phonological knowledge. *Appl Psycholinguist.* (2006) 27:577. doi: 10.1017/S0142716406290398
60. Gray S. Diagnostic accuracy and test–retest reliability of nonword repetition and digit span tasks administered to preschool children with specific language impairment. *J Commun Disord.* (2003) 36:129–51. doi: 10.1016/S0021-9924(03)00003-0

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.

APPENDIX

TABLE A1 | Orthographic transcription of the AEN_NWR stimuli.

Two syllable	Three syllable	Four syllable
Sibad	Dabibum	Lisakubam
Damif	Sifakuf	Zimtakazum
Fibil	Natadulb	Rifatanult
Manib	Sigadilk	Dakanufast
Tundan	Lazafusk	Kabalikift
Nastim	Ristudab	
Bundaf	Mundatis	
Nambik	Randitak	
Saftif	Luntambilf	
Takisk	Rimbadusk	
Bamift		

frontiers | Frontiers in Pediatrics

ORIGINAL RESEARCH
published: 25 April 2022
doi: 10.3389/fped.2022.833603

Screening for Language Difficulties in Disadvantaged Populations on Entry to Early Years Education: Challenges and Opportunities

Julie E. Dockrell[1]*, *Claire L. Forrest*[1], *James Law*[2†], *Sandra Mathers*[3] and *Jenna Charlton*[2]

[1] Department of Psychology and Human Development, Institute of Education, University College London, London, United Kingdom, [2] School of Education, Communication and Language Sciences, Newcastle University, Newcastle upon Tyne, United Kingdom, [3] Department of Education, Social Sciences Division, University of Oxford, Oxford, United Kingdom

Children aged 3–4 years (n = 876) were recruited from deprived areas in England, and a significant minority of the sample were second language learners. Oral language ability was assessed using child administered standardized measures, and parents reported on children's language. We adapted the Language Use Inventory [LUI; (1)] to capture carer's reports of the children's structural language in the language of instruction and their home language (where appropriate). The final measure included six subscales from the original: use of simple words, requests for help, gaining attention, talking about activities/actions, interactions with others, and building sentences. Children's language abilities and non-verbal abilities were below norms on all standardized tests administered except non-word repetition. Factor analysis indicated that all the six scales of the adapted parent completed measure loaded on one language factor. The revised total scale score correlated significantly ($p < 0.0005$) with child assessed language measures, specifically expressive vocabulary and grammar. Different patterns across gender, language status and parental education were examined. Sensitivity and specificity of the scale to identify children with the greatest delays were evaluated. These preliminary data indicated that parent-reported information on children's language skills at 3 years of age has the potential to provide a reliable indicator to inform pedagogy and practice at the start of nursery school. Study limitations are examined and avenues for future development explored.

Keywords: language, preschool, social disadvantage, education, English as an additional (second) language, LUI

Edited by:
Johannes Fellinger, Hospitaller Brothers of Saint John of God Linz, Austria

Reviewed by:
Meghana Wadnerkar Kamble, University of East Anglia, United Kingdom
Giuliana Pinto, University of Florence, Italy

***Correspondence:**
Julie E. Dockrell
J.dockrell@ucl.ac.uk

†Deceased

Specialty section:
This article was submitted to Children and Health, a section of the journal Frontiers in Pediatrics

Received: *11 December 2021*
Accepted: *18 March 2022*
Published: *25 April 2022*

Citation:
Dockrell JE, Forrest CL, 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

INTRODUCTION

Children from areas of social disadvantage experience delays in their language development, with significant numbers of children entering nursery classes with limited oral language skills. These difficulties are exacerbated when a child's home language is not the language of instruction. This causes challenges for early years practitioners in developing the classroom language learning environment and in targeting resources for the children. There is a need for teachers to access tools to profile children's language skills to inform their practice. Parent completed measures provide a viable means of assessing oracy skills on entry to nursery school. We examined the feasibility and validity of using a modified parental completed language checklist in areas of social disadvantage as an indicator of children's language skills as they enter nursery classes.

Language is a foundational skill supporting interaction with others, accessing the curriculum, and developing literacy skills. Proficiency in oral language provides children with a vital tool for thinking and learning (2). Early language delays are a significant risk factor for later literacy and learning difficulties and for longer term unemployment (3). Universal community surveillance estimates prevalence rates of language delays as between 3 and 8% of the population at 30 months of age (4), with children from areas of social disadvantage experiencing disproportionate delays relative to their more advantaged peers (5, 6). Social disadvantage can be measured in different ways, but typically refers to the distal factors (e.g., household income, parental occupation status, parental education, neighborhood poverty) that influence development. Language minority children in areas of social disadvantage experience a double disadvantage (7, 8). There is widespread agreement that additional support is needed to address these inequities, but continued disagreement about the ways in which this should be done. The use of child health surveillance programs to capture language delays (9), targeted parental packages (10), and specific language interventions (11) have all been proposed as ways of addressing the impact of children's language delays [see also (12)]. These targeted approaches have had varying degrees of success but overall point to the importance of ensuring that all children are provided with evidenced informed universal language support when they enter education [see for example (13)].

Children in at risk contexts are likely to need more systematic "targeted" universal interventions in educational contexts [see (9, 14)]. Despite this reported need, limited use of language supporting strategies have been recorded in these settings in areas of social disadvantage [see for example, (15, 16)], and teachers' knowledge of oral language pedagogy is not always optimal (17). An essential component of supporting young children's oracy skills is to provide researchers with reliable and time efficient tools to capture language development at school entry, and early years practitioners with ways of profiling the language performance of children who enter their settings. This is of particular concern in areas of social disadvantage where there is an enhanced need for methodologically robust interventions (18). In this study, we examined whether a parental questionnaire, the Language Use Inventory [LUI, (19)][1], that we abbreviated and modified with respect to its standard administration, would provide reliable and valid language performance data for children from areas of significant social disadvantage on entry to nursery classes in schools in England, at the age of three. Nursery settings were included in the study if they were within the lowest quintile of deprivation as measured by the United Kingdom (UK) government Indices of Deprivation 2015 (20). We argue that knowledge of children's language levels as they enter early learning environments provides educational professionals with the evidence to develop enhanced language learning activities, thereby reducing the likelihood of poorer language learning opportunities (21). However, to achieve this objective, education professionals need access to reliable and valid tools which are cost efficient and can be completed in a time efficient manner.[2]

Staff in early years settings in England are often tasked with profiling and monitoring children's language levels (22). Access to reliable and valid measures of language performance for children between the ages of three and four is, however, limited (23–26). Many of the assessments that do exist are neither designed for teacher use, nor are they appropriate for monitoring progress. These formal assessments typically take a long time to administer, can be complicated to score, and have specifications about the length of time required between assessments due the psychometric properties of the measures. For example, measures considered gold standard by clinicians for children of this age, such as the Clinical Evaluation of Language Basics Preschool [CELF; (27)], take over 40 min to administer and require clinical training for administration and interpretation. It is perhaps not surprising that teachers report a lack of tools to assess children's language levels and regard this as a challenge for practice (28).

An additional challenge in the evaluation of children's oral language for teachers and researchers is the diversity of home languages spoken by children. Schools, where a significant minority of pupils do not speak the language of instruction at home, is now a frequent occurrence (29). In England, over 1 in 5 children has English as an additional language (EAL) (30), and in areas of social disadvantage, these figures are significantly higher. As such, many children are faced with learning and being assessed in the official language of education which is often not their home language. While there are ideological and practical debates about the ways of dealing with multilingual classrooms, the classroom reality is that teachers are working in multilingual contexts in the United Kingdom and Europe (31). An understanding of the language learning needs of language minority children should enhance teaching and learning.

Standardized parental questionnaires provide an alternative means of capturing children's language levels either in English or a child's home language, but these have been typically designed for infants and toddlers or for older children. For example, the MacArthur communicative development inventories (CDI) scales can be used for children 36 months and younger (32). For older children, checklists have emphasized wider aspects of communication. The Children's Communication Checklist—2 [CCC-2, (33)] is a brief clinically relevant caregiver questionnaire which assesses pragmatic language impairments in children and adolescents. The CCC-2 has been the focus of much research, predominantly with children with disabilities and those of school age (34, 35). The measure is not appropriate for use with children younger than 4 years. To address this gap, the LUI (19) was developed. It is a parent-completed questionnaire designed to assess children's pragmatic language between the ages of 18 and 47 months (1). The original measure, designed for use with children whose primary language is English, is an important contribution to tools available and correlates with later outcomes at school entry on standardized language measures (e.g., CELF-P2; CCC-2), demonstrating high levels of specificity (93%) and Negative Predictive Value (92%) (36). The LUI has been used in a

[1]We would like to thank Dr. O'Neill for sharing her measure with us, providing a critical appraisal of our use of the measure and reading the final manuscript.

[2]The author of the original LUI is currently working on an abbreviated version of the full measure. Please contact Dr. O'Neill for details.

range of contexts and translated into several languages, allowing for use in different country contexts (37–39).

The current study focused on children entering preschool settings. As such, we abbreviated the original scale to focus on structural language (vocabulary and grammar). Both vocabulary and grammar are key dimensions of children's oral language as they enter school in England (40), and profiling performance on both dimensions has the potential to inform classroom pedagogy for at risk groups from areas of significant social disadvantage. In addition, given the multilingual context of English schools where, for a significant minority of children, their home language is not the language of instruction, we modified the original administration and scoring (19) to allow parents to complete the form with reference to both English and their home language (see "Materials and Methods" section for details).

To capture language profiles at this point in development for young children at risk of language delay, it is important to establish which language skills should be measured to provide staff with indicative levels of performance and need. It is well established that children who enter school with small vocabularies and limited grammatical skills (41, 42) experience difficulties in accessing the curriculum and have lower levels of attainments in school (43). Significant numbers of children from lower-socioeconomic status (SES) backgrounds have large gaps in their vocabulary (5), negatively impacting learning (44). Growth in oral language during early childhood reflects a continuous development of lexical representations (vocabulary) and the development of an implicit understanding of the rules of grammar. Both these skills have been described as core language components (40, 45). These skills underpin later development of narrative language which can be reliably captured in typically developing children above 4 years of age (40, 45).

There is also increasing evidence that capturing expressive language is crucial. Not only is the development of the ability to use high quality talk in the classroom increasingly recognized as a key component of children's education (46), children's expressive language predicts the quality of language provision they receive (47). Teachers also use spoken language to assess learning (48), and their perceptions of expressive language ability have been significantly correlated with their perceptions of a child's overall development (49). As such, reliable measures of expressive language can provide teachers with data to make evidence based decisions. Finally, agreement between parent report and direct assessment is stronger for language production than for comprehension (50). These studies suggest that, to profile oral language skills for pedagogical purposes, both expressive vocabulary and grammar should be sampled in young children whose language is likely to be delayed.

There are several advantages in using parental report as children first enter nursery school. Parents have had the opportunity to observe and interact with their children across various situations (50), opportunities that nursery staff will not yet have experienced. Moreover, children take time to adapt to nursery settings and their respective routines and, as such, may not use language as competently as they can in familiar settings with familiar people. It is also the case that many of the children in these at-risk settings come from families where their home language is not the language of instruction, English, in the context of this study. As such, parents will have a broader understanding of the diverse ways in which children can use language. Parental reporting, at this point in development, has already been systematically embedded within routine pediatric developmental screening, often to identify at risk children (51, 52). Involving parents in providing information about children's language abilities also brings benefits in terms of enhancing parent-school relationships (since parental knowledge is being valued and requested by the school) and parent involvement in children's schooling, both of which have been shown to predict children's academic and social outcomes (53). The potential for using similar approaches in early years educational settings to develop language learning pedagogy requires further exploration.

THE CURRENT STUDY

We reasoned that a modified version of the LUI (1), focusing on children aged 3–4 years of age as they take their first steps in universal education, had the potential to provide teachers with a profile of children's oral language strengths and needs. Given our focus on children from areas of significant social disadvantage, the LUI items were modified in three ways. Firstly, to shorten completion time for parents, the four subscales which are not included in the original LUI Total Score and its norms were omitted (Subscales A and B on gestures, and Subscales E and L with open-ended questions about the child's interests in play and talk). Secondly, we prioritized subscales with items related to language use with developmentally later-emerging expressive vocabulary and grammar, given their importance for later academic achievement [see for example (54, 55)] and that evidence that expressive narrative skills in 3-year-olds from disadvantaged settings may be too limited to provide a useful focus of assessment (56). Thirdly, we sought to prioritize subscales with potential to provide feedback to teachers about their children's language levels and ultimately inform targeted interventions (57). Six of the 10 scored LUI subscales met our criteria, with three focusing primarily on vocabulary (*C: Types of words your child uses, F: How your child uses words to get you to notice something, I: Your child's use of words in activities with others)* and three focusing on more extended language use and grammar (*D: Your child's requests for help, H: Your child's questions and comments about themselves or other people, N: How your child is building longer sentences and stories*). These six subscales all had high factor loadings on the original scale of 0.83–0.95 [see (19)]. We will refer to our abbreviated version of the LUI as the LUI-6 henceforth. Scales are reported as described in the original LUI.

Given the focus on educational settings, we also modified the original administration and scoring (19). We adapted the LUI from one response column to two response columns. This allowed parents to separately record children's abilities in English and their home language. The LUI-6 English is used to refer to parents' reported use of English. When we are referencing reported use of English on the LUI-6, we refer to it as LUI-6 English. Language status was operationalized in the current context as either monolingual English or different home language. We first examined responses for English to establish whether the LUI-6 was reliable and valid and reflected the same factor structure as the original. Secondly, to examine validity, we compared responses for the LUI-6 with scores on a standardized child administered measures of oral language. Thirdly, we examined whether performance on the LUI-6 provided reliable data to identify the children who performed below 1.5 SDs from the mean on standardized measures of oral language. Finally, we explored the data provided by a sub-sample of parents who recorded their child's language in both English and a home language to establish whether there were differences between reports for English and home language. Given the current understanding of oral language development at age three, we predicted that responses for the six subscales from the LUI-6 would load on a single factor as in the original research (1). We also anticipated that, as with the original LUI, the LUI-6 would correlate significantly with child administered standardized measures of the children's expressive and receptive oral language. We anticipated that the focus on expressive language in the LUI-6 would result in high levels of sensitivity and specificity in identifying children who were struggling most with English at this point in development. Assessment of home language skills raises many challenges and, while lags in the language of instruction are often evident, there have been fewer attempts to capture differences in performance between the language of instruction and home language in disadvantaged populations.

TABLE 1 | Demographics of sample.

	London (n = 438)	Teesside (n = 384)	Total sample (N = 822)
Child characteristics			
Mean Age in months (SD)	43.32 (3.90)	43.82 (3.95)	43.56 (3.93)
Female (%)	49.5	54.4	51.8
First born (%)	33.3	31.9	32.6
Reported language concern (%)	20.1	12.6	16.5
Home language (%)			
Monolingual	30.0	87.7	57.6
English as an additional Language	70.0	12.3	42.4
Parent/carer characteristics			
Relationship to child (%)			
Parent	98.5	95.9	97.4
Other	1.4	4.1	2.8
Highest level of education (%)			
Primary school	8.2	8.5	8.3
GCSE level	19.9	24.8	22.2
Above	71.9	66.7	69.5
Frequency of book reading (%)			
Every day	38.6	41.4	39.9
At least once a week	60.1	57.3	58.8
Never	1.2	1.4	1.3

MATERIALS AND METHODS

Ethics

Ethical approval was granted by Institute of Education (IOE) Research Ethics Committee (IOE REC 1118: Empowering Staff to Enhance Oral Language in the Early Years).

Participants

The current sample formed part of a larger intervention study investigating the language learning needs of children from disadvantaged areas in England (58). A power calculation using G*Power (Cohen's $d > 0.20$, $p < 0.05$) determined that a target sample of 600 participants was needed to observe an intervention effect. Schools that were in the lowest quintile for deprivation in the United Kingdom and contained a nursery class were recruited in Greater London and Teesside (in the Northeast of England). Settings were recruited via flyers advertising the study, presentations given to school partnerships, and a meeting of Early Years leaders from local primary schools. The project was also publicized through local charities, language leads, and social media. Deprivation information was generated from nursery postcodes using the Income Deprivation Affecting Children Index (IDACI), one of the subscales of the UK government's English Indices of Deprivation 2015 (20). The IDACI measures the proportion of all children aged 0–15 years living in income-deprived families and ranks small geographical areas in England from 1 (most deprived) to 32,844 (least deprived). IDACI Ranks were generated from each nursery's postcode. These scores were converted to a z-score using the normsinv function on excel and dividing the IDACI Rank by 32,844, as recommended by Bishop (59). Thirty-seven nurseries in Greater London expressed interest, of which 28 met inclusion criteria for deprivation. Five nurseries did not respond to invitations to participate, and two nurseries declined to participate. Thirty nurseries in Teesside which met inclusion were approached. Seven nurseries did not respond to invitations and three nurseries declined to participate. In total, 41 nurseries were recruited into the study, but two nurseries, one from each area, withdrew, resulting in a final sample of 39 nurseries (London n = 20; Teesside n = 19).

A total of 876 children were recruited into the sample, 29 of whom were not eligible because either: their parent or teacher reported special educational needs (SEN) (n = 14); children no longer attended the nursery (n = 9); children were wrongly recruited (too young or in a different class) (n = 4); or teacher reported no English (n = 2). A further 25 children did not complete the assessments due to absences (n = 9) or refusals (parent withdrew consent/incomplete consent n = 3; child declined to participate n = 13).

Demographic details were collected from the parents, and parents were also asked whether they had any concerns about their child's language development. Demographic details are presented in **Table 1**. As the table shows, 42.4% of the children spoke English as an additional language (EAL, n = 331). There were 47 different languages spoken in the sample (n = 19 European, n = 14 African, n = 6 South Asian, n = 5 East Asian, n = 2 Middle Eastern and n = 1 Caribbean). More children were from bilingual homes in the London sample [$\chi^2(1, 823) = 276.97$, $p < 0.001$], and more children attended nursery part-time in

Teesside [$\chi^2(1, 801) = 57.44$, $p < 0.001$]. No other differences were statistically significant.

Measures

Children's language abilities were measured in two ways: first, by direct assessment of the child's language skills in the nursery setting and, secondly, by asking parents to complete the LUI-6. Details of the measures are discussed below.

Child Assessment

Expressive Language

The Naming Vocabulary subtest of the British Ability Scales 3rd Edition [BAS-3; (60)] assesses children's knowledge of nouns. Children are required to name colored pictures of objects shown one at a time. Successful performance depends on expressive language ability, picture recognition skills, and long-term memory skills. Naming vocabulary has a test-retest reliability of 0.92, and internal reliability coefficients range from 0.70 to 0.92 between the ages of 3 and 4 years.

The sentence repetition component of the Grammar and Phonology Screening test [GAPS; (61)] is another measure of expressive language, designed to highlight significant markers of language impairment and reading difficulties. Children repeat 11 sentences that target subject-verb agreement, tense marking (past, future), phrasal embedding, dative construction, object question formation, reversible passive construction, and anaphoric and pronomial reference. All words have an early age of acquisition, simple phonological structure, and knowledge of the words is reported to be independent of socioeconomic and cultural bias. The sentence repetition component has a reliability of $\alpha = 0.86$ and is significantly correlated with the CELF Sentence Structure ($r = 0.52$). Sentence repetition tasks are very sensitive markers for identifying children with developmental language disorder (62).

Receptive Language

The Verbal Comprehension subtest from the BAS-3 (60) provides a measure of receptive language by assessing children's understanding of oral instructions involving basic language concepts. Children respond by either pointing to a picture, handing objects to the researcher, or placing objects in different positions according to the instructions. Questions assess children's understanding of object names, commands (e.g., "Put the horse in the box"), the functions of objects (e.g., "Give me the one we draw with"), prepositions (e.g., "Put the car under the bridge"), and complex instructions (e.g., "Give me all the red shapes except the square."). Verbal Comprehension has a test-retest reliability of $r = 0.78$ and internal reliability coefficients range from 0.85 to 0.91 between the ages of 3 and 4 years.

Phonological Skills

The non-word repetition subtest from the GAPS (61) requires children to repeat eight non-words which vary in complexity (e.g., *dremp, bademper, difimp,* etc.). It has reliability of $\alpha = 0.73$ and construct validity of $r = 0.58$ with the Children's Test of Non-word Repetition [CNRep; (63)].

Narrative Skills

The Bus Story test (64) is a measure of story retell abilities while also drawing on verbal comprehension. The researcher reads aloud a story about a "naughty bus," and children follow along looking at the pictures. Children are then asked to tell the researcher what happened in the story using the pictures as guides. The Bus Story was administered on a laptop, and children's responses were recorded with the in-built microphone and by hand. Research assistants transcribed the recordings and then scored the responses for information (the amount of information the child conveys when telling the story) using the published coding system. Participants receive a score of 2, 1, or 0 for each item, depending on the amount of detail that has been reported. There are a total of 54 points that can be obtained. Information has a test-retest reliability of 0.79 and construct validity of 0.98. Ten randomly selected transcripts were coded by the lead researcher (CF) and two research assistants using the scoring guidelines. Any discrepancies in scoring were discussed, along with the rationale for scores, until agreement was reached. Any decisions were documented, and a further ten random transcripts were coded using these guidelines. Again, discrepancies were discussed until agreement was reached. The researchers then coded 20% of the complete Bus Story assessments ($n = 149$ transcripts) which were randomly selected. Inter-rater reliability on these 149 transcripts was calculated using intra-class correlations (ICC). A mean-rating ($k = 3$), absolute-agreement, two-way mixed effects model was used to calculate ICC in SPSS v25.0 (65). This model treats the rater as a fixed effect and the Information Score as a random effect because we are only interested in the reliability between the three raters who coded the same 149 transcripts (66). Inter-rater reliability coefficients were excellent [ICC = 0.9 (95% CI = 0.97–0.99)] (67).

Non-verbal Abilities

The non-verbal reasoning ability cluster from the BAS-3 (60) consists of the Matrices scale and the Picture Similarities subtest. Both subtests were administered on a laptop. The Matrices scale assesses children's perception and application of relationships among pictures or abstract figures. Children are shown a 4×4 matrix, three of which are colored objects and one is blank. Below the matrix are four options from which to choose the missing object that fits the pattern. There are minimal oral instructions, and children respond by pointing. Internal reliability for the Matrices subtest ranges from 0.68 to 0.83 between the ages of 3 and 4 years and has a test-retest reliability of $r = 0.64$. The Picture Similarities subtest requires children to match a picture to one of the four pictures displayed. This task requires minimal oral instruction, can de demonstrated by gesture, and children do not need to respond verbally. The Picture Similarities subtest has an internal reliability ranging from 0.68 to 0.83 between the ages of 3 and 4 years and a test-retest reliability of $r = 0.64$. Overall, the non-verbal reasoning ability cluster has a test-retest reliability of $r = 0.77$ and an internal validity of $r = 0.46$.

Parental Report of Language Ability

The LUI (19) is a standardized parent-report measure of language use in daily life for children aged 18–47 months. Parents

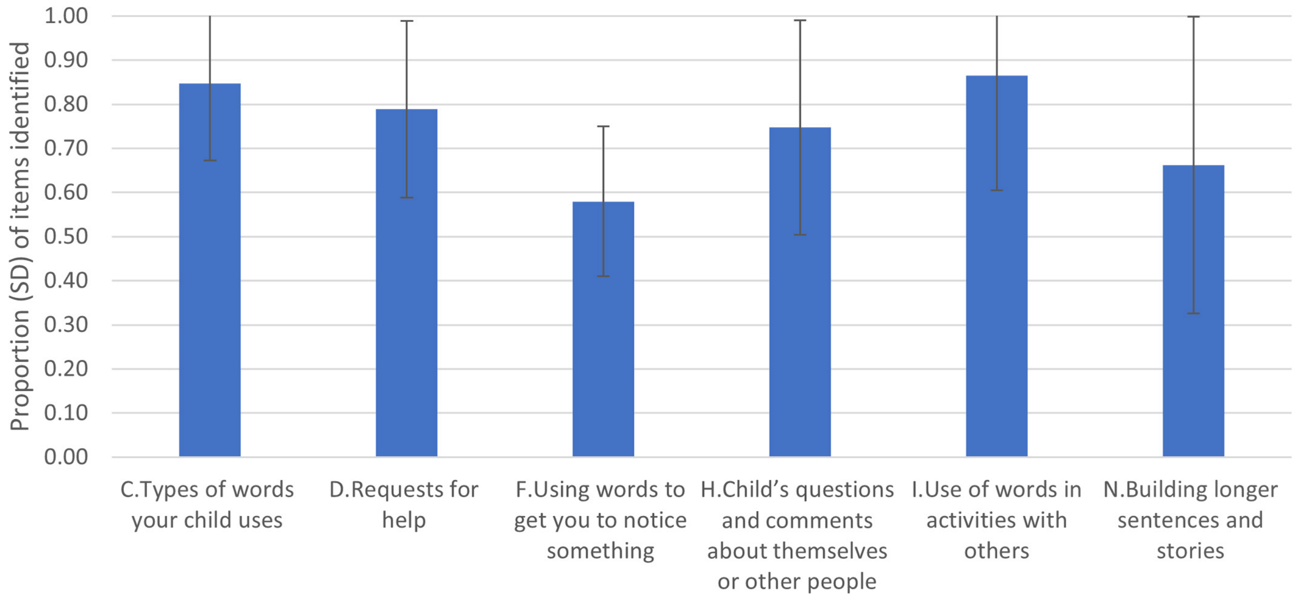


FIGURE 1 | Mean proportion of items (SD) recorded correct for use of English for each of the LUI subscales employed.

report on children's current language and communication skills using a nominal scale (Yes/No) or ordinal scale (Never/Rarely/Sometimes/Often). Parents were instructed to complete the questionnaire in 1 day and were encouraged to ask others for help if needed (e.g., partner, child's grandparent, child's nursery teacher). The LUI was designed to assess young children's spoken pragmatic language use with a focus on the functions of language influenced by children's developing cognitive abilities and increasing social interactions during this age (1). As described above, six of the original 10 scored LUI were used, namely, *C: Types of words your child uses; D: Your child's requests for help; F: How your child uses words to get you to notice something; H. Your child's questions and comments about themselves or other people*; *I: Your child's use of words in activities with others;* and *N: How your child is building longer sentences and stories*.

The items chosen assesses later appearing and more sophisticated uses of language that involve greater vocabulary, grammatical, and conversational skills (68). The original version of the LUI has good internal reliability, ranging from 0.83 to 0.98 (1, 19). Five minor word changes were made to example phrases provided to reflect a British sample (e.g., mommy to mummy; airplane to airplane).

Finally, the format of administration was modified. The original instructions allow parents to reply regardless of language used by the child. This was modified so that, in this study, parents could indicate, for each item, in their reply whether the child was able to produce the target item in English or their home language (if appropriate). The current data report on the nominal items (omitting 9 ordinal items) for which parents replied with respect to their child's use in English (the LUI-6 English) and, where appropriate, the home language, with a total of 111 English items scored 1 (Yes) or 0 (No).

In the current study the standardization from the LUI does not apply given the use of selected subscales from LUI such that the full range of abilities were not sampled, and parents were not given the standard instructions as per the LUI Manual.

Procedure

Nursery settings were recruited into the intervention study spring/summer 2019. Informed consent and background questionnaires were delivered to parents by nursery staff. A team of five research assistants and one research associate conducted the 1:1 child assessment sessions during the autumn term. All assessors completed comprehensive training, and their performance was assessed at the end of the training. Each researcher spent approximately 1 week at each setting administering the assessments in two 20-min sessions. Children received a sticker after each session as a thank you for their participation. Parents received an envelope containing the LUI-6 after their child had been assessed with instructions to complete and return to the nursery within a week. Envelopes and questionnaires were labeled with a unique ID for each child to maintain anonymity. Nursery staff were encouraged to support parents in the completion of the questionnaires when necessary.

Statistical Analysis

Data were analyzed using SPSS v25. Chi-square analysis was used to compare group differences between those who did and did not return the LUI-6, examining background demographics and child performance on direct assessments of language. Factor analysis of the LUI-6 was conducted using principal component analysis with varimax rotation to examine the factor loading for the LUI-6. Subsequently, group differences on the LUI-6 factor score were examined using an ANOVA with gender, language status (monolingual or bi/multilingual), and parent report of language concern entered as fixed factors. Concurrent validity of the LUI-6 factor score was analyzed using zero-order correlations to explore the relationship between the LUI-6 factor score and scores from the directly assessed language measures. The effect of directly assessed language ability on the LUI-6 factor was assessed using stepwise regression by entering language status first. A receiver operating characteristic (ROC) analysis was conducted to measure the sensitivity and specificity of the LUI-6 in identifying children with the lowest language ability (more than 1.5 SDs below the mean). Finally, differences between English language and home language use as reported by the LUI-6 were compared using t-tests, and principal component analysis was used to explore the factor loadings of the LUI-6 when home language was reported.

TABLE 2 | Children's raw scores on the six LUI subscales based on parental reported use of English (sole or in addition to home language) (N = 335).

LUI Subscale	Number of items	Mean	*SD*
C: Types of words your child uses	21	17.78	3.67
D: Your child's requests for help	7	5.52	1.4
F: How your child uses words to get you to notice something	6	3.48	1.02
H: Your child's questions and comments about themselves or other people	36	26.9	8.74
I: Your child's use of words in activities with others	14	12.1	3.64
N: How your child is building longer sentences and stories	36	23.82	12.12

RESULTS

The results are presented in five sections. In the first section, we examine differences between the parents who returned a completed LUI-6 for their children (for English, home language or both) and those who did not. Significant numbers of parents did not complete or return the questionnaire, and we reasoned that these data inform interpretation of the subsequent results. Section 2 provides data on the pattern of responses across the six subscales used and examines the factor structure of the LUI-6 using principal component analysis to establish whether the six subscales formed unitary construct (as with the original LUI) or several constructs. In Section 3, we examine the concurrent validity of the LUI-6 by exploring parental responses with direct child assessment on the language measures collected from the children. Section 4 explores the sensitivity and specificity of the LUI-6 in identifying children with the most delayed language, and, finally, Section 5 explores responses related to the children's home language, where reported.

Section 1: Parental Completion of the Language Use Inventory-6

Of the final 876 participants in the sample, a total of 338 LUI-6 forms were returned (38.6% of the total sample). Of these 338 completed forms, 225 were completed with respect to the English only (66.6%), and a further 110 (32.5%) were completed with reference to both spoken English and the child's *spoken* home language. Only three forms (1%) were completed with respect to solely the child's use of their home language, and these are not considered further in the analyses.

The LUI-6 was more likely to be completed by parents of monolingual English children [$\chi^2(1, N = 794) = 9.80, p = 0.002$] who were from less deprived catchment areas [$t(837.255) = 3.34$, $p = 0.001$; LUI-6 completed $M = -1.60$, $SD = 0.469$; not completed $M = -1.71$ $SD = 0.61$], and who had children who scored higher on the BAS-3 oral language measure [$t(794) = 2.68$, $p = 0.008$: LUI-6 completed $M = 87.68$, $SD = 14.97$; not completed $M = 84.81$, $SD = 14.65$]. Where significant differences were evident, either the association was weak (monolingual English Cramer's $V = 0.105$) or the effect sizes were small (Indices of deprivation Cohen's $d = 20$; oral language measure Cohen's $d = 0.19$). There were no differences between the educational levels of parents who returned the form and those who did not [$t(642.313) = -4.77$, $p = ns$: LUI-6 completed $M = 3.15$, $SD = 1.11$; not completed $M = 3.26$, $SD = 1.03$), child age $t(865) = 1.162$, $p = ns$: LUI-6 completed $M = 43.66$, $SD = 3.83$; not completed $M = 43.34$, $SD = 3.99$)], or parental concern about the child's language [$\chi^2(1, N = 808) = 0.005$, $p = ns$]. In sum, there was evidence of some bias in response rates where forms were less likely to be completed by the parents of bilingual/multilingual children, children with lower language levels, and those from the most deprived settings.

Section 2: Performance on Language Use Inventory-6 Subscales

We first consider responses related to children's use of English (either when only completed in English or when completed in addition to home language), reflecting the language of instruction and the original development of the LUI scale. In the final section where data are reported (see **Table 7**), we examine responses for home language (among those who also reported use of English) and compare these with responses for English.

Table 1 reports the responses to the different six subscales used from the original LUI. **Figure 1** shows the proportion correct for each scale to allow comparison across the six subscales. To test the overall internal consistency of all the subscales together, we first computed Cronbach's alpha. As with the full LUI, Cronbach's alpha was high for the selected subscales, 0.979. Children's mean scores (and SD) for each subscale are provided in **Table 2**. Apart from subscale N, *How your child is building longer sentences and stories,* the subscales were negatively skewed, indicating that children's reported performance was high. This was most marked for subscale C, *Types of words your child uses* (skewness -3.224, $SE = 0.133$).

To explore relative performance across the subscales, proportion scores for each subscale were calculated. These are shown in **Figure 1**. Both subscale F *'How your child uses words to get you to notice something'* and subscale N "*How your child is building longer sentences and stories*' were subscales where the lowest proportion of items were reported by the parents for the children in the study. A non-parametric Friedman test of

TABLE 3 | LUI-6 subscales principal component analysis, rotated factor loadings based on parental reported use of English (sole or in addition to home language) (N = 335).

LUI subscales	Factor loadings
C: Types of words your child uses	0.867
D: Your child's requests for help	0.887
F: How your child uses words to get you to notice something	0.901
H: Your child's questions and comments about themselves or other people	0.943
I: Your child's use of words in activities with others	0.942
N: How your child is building longer sentences and stories	0.806

differences on proportion scores was conducted, and there was a significant difference between the subscales [$\chi^2(5) = 822.27$, $p < 0.0005$]. Dunn-Bonferroni *post hoc* tests were carried out, and there were significant differences between the subscale F '*How your child uses words to get you to notice something*' and all other subscales (all $ps < 0.0005$). Performance on subscale N "*How your child is building longer sentences and stories*" was significantly lower than subscale C "*Types of words your child uses*" ($p < 0.0005$) and subscale I "*Your child's use of words in activities with others*" ($p < 0.0005$). The data reflect developmental trends in performance found in the original studies (1, 19).

Following O'Neill (19), we examined the factor structure of the revised scale. Principal Component Analysis was conducted with varimax rotation. As **Table 3** shows, all subscales loaded on a single factor accounting for 79.6% of the variance, with loadings for each scale on the factor all above 0.80. As one factor was identified, the regression-based factor score is used in subsequent analyses to identify ranking on the latent variable and allow follow-up analyses. Factor scores are standard scores with a Mean = 0 and Variance = squared multiple correlation between items and factor. This procedure maximizes validity of estimates. This latent variable is referred to as LUI-6 factor.

We examined differences on the LUI-6 factor score by gender, language status (monolingual or bi/multilingual), and whether parents had reported concerns about child's language. An analysis of variance yielded a main effect for monolingual English [$F(1,301) = 12.16$, $p = 0.001$, $\eta_p^2 = 0.04$ (monolingual English speakers LUI-6 English factor score $M = 0.202$, $SD = 0.856$; bilingual/multilingual LUI-6 English factor score $M = -0.308$, $SD = 0.1.109$)] and reported language concern [$F(1,301) = 8.02$, $p = 0.003$, $\eta_p^2 = 0.03$ (no language concern LUI-6 English factor score ($M = 0.101$, $SD = 0.938$); language concern LUI-6 English factor score ($M = -0.383$, $SD = 0.1.101$)], but there was no effect for gender [F (1,301) = 0.031, $p = ns$ (female LUI-6 English factor score $M = 0.114$, $SD = 0.91$); Male LUI-6 English factor score ($M = -0.072$, $SD = 0.1.040$)]. In sum, differences on the LUI-6 factor score were evident for language status and concern about children's language skills, but effect sizes were small, accounting for only 4% of the variance for language status and 3% of the variance for parental concern. There were no significant interaction effects.

Section 3: Concurrent Validity of Language Use Inventory-6 Factor Score Based on Parental Reported Use of English (*N* = 335)

Given the six subscales from the original LUI had been selected to reflect vocabulary and grammatical aspects of oral language, we examined the LUI-6 factor score based on parental reported use of English (sole or in addition to home language) in relation to the scores children achieved on the direct assessments of oral language captured in the study.

As **Table 4** shows, there were significant correlations between the LUI-6 English factor score and all language measures except the Bus Story sentence length. As predicted, correlations with standardized measures of expressive vocabulary and grammar were large, demonstrating high levels of concurrent validity between parent reported English use on the LUI-6, and these standardized directly administered child measures. Relationships between the LUI-6 and non-word repetition and narrative skills, while significant, were lower. Correlations between naming vocabulary and sentence repetition and the LUI-6 factor score were statistically significantly higher than those between the LUI-6 factor score and Bus Story sentence length ($Z = 10.732$, $p < 0.001$), non-word repetition ($Z = 8.331$, $p < 0.001$), Bus Story information ($Z = 6.311$, $p < 0.001$), and Verbal comprehension ($Z = 2.337$, $p = 0.01$), providing further evidence that the LUI-6 is capturing lexical and grammatical expressive oral language in this population.

Using regression analyses, we examined whether direct assessments of children's language contributed significantly to the LUI-6 English factor score once language status (monolingual or bi/multilingual) was accounted for. Language status was entered first, followed by the standardized measures of oral language and parental concern about children's language levels. This allows a test of whether performance on the standardized language measures accounted for performance on the LUI-6 score once language status was considered. We hypothesized that both language status and expressive language would account for the most variance in the LUI-6 factor score. A significant model was found when only language status was used in the model [$F(1, 334) = 25.16$, $p < 0.0001$, $R^2 = 0.07$, $R^2_{Adjusted} = 0.067$], but the model only accounted for a small proportion of the variance. Inclusion of the directly assessed language measures in the second step significantly improved the model, resulting in a significant R^2 *change* = 0.266 and a model that accounted for over 30% of the variance in the LUI-6 score [$F(6, 334) = 27.66$, $p < 0.0001$, $R^2 = 0.346$, $R^2_{Adjusted} = 0.324$]. As **Table 5** shows the final model language status is no longer significant and responses to the LUI-6 are explained by children's performance on direct assessments of naming vocabulary and verbal comprehension.

Section 4: Capturing the Children With the Greatest Levels of Language Learning Need

To identify whether scores on the LUI-6 English factor score accurately identified children who had the poorest levels of

TABLE 4 | Zero order correlations between LUI-6 factor score and directly assessed language skills (n = 312).

	LUI-6 factor score	GAPS: sentence repetition	GAPS: non-word repetition	BAS: verbal comprehension	BAS: Naming vocabulary	Bus Story information
GAPS: sentence repetition	0.411**					
GAPS: Non-word repetition	0.206**	0.585**				
BAS: Verbal comprehension	0.533**	0.647**	0.431**			
BAS: Naming vocabulary	0.640**	0.543**	0.273**	0.723**		
Bus Story: Information	0.322**	0.555**	0.315**	0.527**	0.502**	
Bus Story: Sentence length[a]	0.062	0.368**	0.292**	0.215*	0.300**	0.531**

*[a]Only 127 children produced Bus Story narratives that were sufficiently long to complete a sentence length score. *Correlation is significant at the 0.01 level. **Correlation is significant at the 001 level.*

TABLE 5 | Final regression model examining predictors of performance on the LUI-6 factor based on parental reported use of English (sole or in addition to home language) (N = 335).

	B	Std error	Beta	t	Sig
Language status	−4.645	2.834	−0.801	−1.639	*ns*
GAPS sentence repetition	0.618	0.541	0.78	1.143	*ns*
GAPS non-word repetition	0.036	0.579	0.003	0.061	*ns*.
Verbal comprehension	0.281	0.085	0.219	3.319	0.001
Naming vocabulary	0.371	0.066	0.340	5.600	<0.0005
Bus story information	−0.174	0.400	−0.024	−0.436	*ns*

ns, non significant.

language, we computed a ROC analysis. ROC analysis provides data on which variables offer the best discriminatory power in this study to identify children with the lowest levels of language. If the area under the curve is 1, this would illustrate perfect discrimination with 0.5 being chance.

We used standard scores on the BAS-3 verbal ability measure as our benchmark level of language performance. Following data which suggest that significant children struggling with oral language can be captured when scores are more than 1.5 SDs below expectation, children who performed more than 1.5 SDs from the mean were classified as the poorest performers (33.4%, n = 108). To capture the utility of the LUI-6 English factor score, we compared performance on the parent-completed LUI-6 questionnaire with two direct assessments of children's expressive language not used to identify poorest performance in oracy (GAPS sentence repetition and the Bus Story information score). As **Table 6** shows, all measures demonstrated good sensitivity and specificity in identifying the children with the poorest levels of language and, importantly, the LUI-6 English factor score performed at a similar level to direct assessments of the children's language levels.

Section 5: Language Use Inventory-6 Subscale Scores Reported for Home Language (in Addition to English)

There is a dearth of knowledge about children's performance on home languages in multilingual settings. As such, we explored parental responses for children's home language use (which, among this group, was always in addition to reporting English use). A significant minority of our sample completed the LUI-6 with a reference point of the child's home language (n = 110) in addition to English language performance. As such, we examined the LUI-6 home language scores. The internal consistency for the LUI-6 Home Language was good with Cronbach's alpha = 0.865. Our first step was to compare reported performance in English and the child's home language on the LUI-6 subscales. As **Table 7** shows, across all six subscales, parents reported greater levels of oracy skills in English than in the home language. For all comparisons, effect sizes were large.

As with the LUI-6 English, we examined the factor structure of the LUI-6 Home Language. Principal Component Analysis was conducted with varimax rotation. As **Table 8** shows, as for English, all six subscales loaded on one LUI-6 Home Language factor which accounted for 81.1% of the variance with all subscales loading above 0.80 on the factor.

As with performance on the LUI-6 English, we examined whether there were differences by the child's gender and parents report of concern about the children's language development. As with the score LUI-6 score in English, there was no statistically significant difference by gender [$t(103) = 0.165$, *ns*], but scores for children where a concern about oral language was reported (n = 19) were significantly lower [$t(32.07) = 2.36$, $p = 0.024$].

DISCUSSION

Oral language skills are foundational for learning and attainment in school, and it is well established that, on average, children coming from disadvantaged backgrounds experience challenges with both vocabulary and grammar (3, 5, 69, 70). Despite the evidence, there has been little attention paid to enhancing teachers' knowledge about how to capture children's language learning skills in the early years. Such evidence could inform classroom pedagogy and practice in the early years of schooling (71). To address this gap in the literature, we used the LUI (19), abbreviated to six subscales, for use with children from areas of social disadvantage as they entered nursery classes in England. As predicted, the LUI-6 subscales all loaded on a single language factor. In addition, the LUI-6 English provided reliable and valid data, correlating with child administered measures of oral language. As anticipated, the LUI-6 resulted in high levels of sensitivity and specificity in identifying children who were struggling most with English at this point in development. By contrast, our exploration of the parents' responses in children's home language did not provide evidence of differential performance in the child's home language. These results are discussed in detail below.

TABLE 6 | ROC analysis for LUI-6 English factor score and direct language assessments in identifying children with the greatest levels of language learning need: area under curve and 95% CI.

	Area under the curve	Std error	Significance	Lower bound	Upper bound
LUI-6 English factor score	0.780	0.033	<0.0005	0.714	0.845
GAPS sentence repetition	0.806	0.027	<0.0005	0.752	0.859
Bus story information	0.795	0.028	<0.0005	0.740	0.849

TABLE 7 | Comparison between scores for LUI-6 subscales when reported for English language use and home language use for the same child.

LUI subscale	LUI-6 reported Home Language *M (SD)*	LUI-6 reported English M (*SD*)	*t* all sig < 0.0005	Cohens *d*	95% CI
C: Types of words your child uses	10.10 (7.48)	15.58 (5.63)	5.529	0.54	0.33–0.74
D: Your child's requests for help	2.95 (2.13)	4.79 (1.68)	4.923	0.48	0.28–0.68
F; Your child's use of words to get you to notice something	1.57 (1.28)	2.98 (1.37)	5.918	0.59	0.37–,78
H: Your child's questions and comments about themselves and other people	12.85 (12.91)	22.26 (11.52)	4.818	0.47	0.27–0.64
I: Your child's use of words in activities with others	5.72 (5.89)	10.21 (5.00)	5.033	0.49	0.29–69
N: How your child is building longer sentences and stories	8.84 (12.12)	17.20 (12.62)	4.378	0.43	0.23–0.63

TABLE 8 | LUI-6 home language principal component analysis, rotated factor loadings.

LUI subscale	Factor loadings
C: Types of words your child uses	0.868
D: Your child's requests for help	0.901
F; Your child's use of words to get you to notice something	0.920
H: Your child's questions and comments about themselves and other people	0.947
I: Your child's use of words in activities with others	0.925
N: How your child is building longer sentences and stories	0.840

The subscales of the original LUI (19) were ordered to capture early to later emerging social pragmatic functions of language use, as validated in studies with the LUI and its standardization (1, 19, 39). Our data demonstrated similar developmental trends across the six subscales when the focus was primarily on expressive vocabulary and grammar as these are foundational for later language use. In a sample of children from areas of significant social disadvantage, the data showed that language performance can be captured by parental data. There were statistically significant differences between the scales selected from Part 2 of the original LUI (Your child's communication with words) and those selected from Part 3 (Your child's longer sentences). The majority of the children in our sample were using words for earlier appearing functions (e.g., Subscale C: *Types of words your child uses*), whereas performance was lower for later-appearing functions more likely to involve longer sentence constructions (e.g., Subscale N: *How your child is building longer sentences and stories*) and *use of words to get you to notice something* (Subscale F) were significantly lower.

Data are preliminary but, if validated in wider settings, there are potential pedagogical implications. Children who rely on short or single word utterances at the age of three are at risk. The data from the Bus Story indicated that the children had difficulties in producing narratives and most of the sample (both monolingual and those with English as an additional language) could not produce five sentences to compute an average sentence length score on the standardized measure. Research has shown that difficulties in producing extended discourse and using language to engage others limits children's ability to communicate with others in social settings and their ability to actively engage with classroom activities (72, 73). Discourse skills, critical for later achievement, are built on vocabulary and grammar but also from children's ability to engage in conversational turns. There is a need to move beyond a focus on vocabulary and reducing the word gap in interventions (41) to provide opportunities for extended discourse, to enhance the classroom language learning environment (74, 75), and to develop strategies and resources for their children at entry to nursery (76).

The LUI-6 English was statistically significantly associated with robust child administered measures of oral language. This speaks to the validity of the LUI-6 English, but also has important practical implication. Standardized language tests that need to be administered are time consuming and, typically, teachers do not have access to these standardized measures. The regression analysis indicated performance on the LUI-6 English was predicted by children's performance on the standardized tests and parental concern about language development, and not on whether they were monolingual English. As such, it was performance as captured by the LUI-6 English that identified children's language performance, not whether they spoke English as an additional language. These results could provide teachers with data to target their universal support (12). In addition, ROC analyses indicated that the LUI-6 English identified the children who were struggling most with oral language in English. Combined, these data point to the utility of the scale in capturing performance in English on entry to nursery school, providing teachers with evidence to embed more targeted support for oral language with these children and monitor progress systematically (57). The use of a parent-completed measure has the potential to supplement teachers' own assessments of children's language and, arguably, to provide a more nuanced picture since parents can provide information about the diverse ways in which their child uses language across multiple contexts and in multiple languages (where relevant). Fostering acceleration requires high-quality, intentional language modeling, and instruction within preschool classrooms (77). Knowing which oral language dimensions to target and which children to provide additional support is the basis for evidence-based practice (78).

We found no differences in reported performance on the LUI-6 for gender either reported for English or home language. This appears to contrast with other studies documenting the early (age 2) linguistic superiority of girls (79) in children who are disadvantaged (80) and the heightened risk of boys for developmental language difficulties (81). Data vary in how long this difference between genders is maintained, with the difference being less evident from 28 months when boys seem to have an increased language learning trajectory and catch up to girls (82). The mean age of our participants was 3;6, and the data suggest that in this population, at this age, gender differences were not evident.

The data also indicated that parental concern, while accounting for a small proportion of variance, was also a predictor of children's language status in both English and home language. These data further support the importance of schools engaging with parents to explore their views about their children's language development. While the checklist format of the LUI-6 reduced parents' need to provide written text and the readability of the questionnaire was of primary school level, some parents may have found understanding the questions and making judgments challenging, thereby resulting in a reduced completion rate. While questionnaires to parents often result in low completion rates, the significant number of unreturned forms (61.4%) raises challenges for use in disadvantaged populations. In particular, this group was more likely to consist of parents who reported speaking a language other than English, who came from more deprived areas, and who had concerns about their child's language. These findings raise questions about the most effective ways to engage parents in research given the important role they can play in profiling their child's language skills.

There is limited research on the language learning profiles of dual language learners, especially dual language learners in areas of significant social disadvantage (83). Our attempt to consider children's home language use on the same scale was both exploratory and novel. The same unitary structure was evident for the scale in both English and home language. Overall, as expected, children who were monolingual performed better on the language measures in English. However, as the regression analyses indicated, it was not whether children were monolingual that predicted their factor score on the LUI-6 English, but rather their performance in English as assessed by direct measures of oral language. Previous studies have shown better performance on the oral language assessments was associated with more English language exposure and more exposure from native English speakers (8, 84). These are factors which can be addressed in early years settings. This is not to minimize the importance of home language, but rather to empower children in the current language of education in England. In this context, it is noteworthy that overall, where parents reported for English and home language, skills were always reported to be more advanced in English. Further work is required to establish whether these differences reflected the heterogeneity of the population sampled, or the way in which the scale was administered. Performance of language minority pupils is more strongly associated with the concentration of social disadvantage than with the concentration of pupils who speak a different language than the one taught at school [(85) p. 19]. Capturing children's proficiency across their languages provides opportunities for settings to build on language diversity in an evidence informed way.

Limitations

Despite our large unique sample and the robust evidence for the validity of the modified version of the LUI, there are several significant limitations which need to be addressed in future studies. Firstly, although many parents completed the form, there was evidence of selective completion of the scale. In particular, even in using the abbreviated LUI, fewer responses were received from parents of bilingual/multilingual children, children with lower language levels, and those from the most deprived settings. The reasons for non-completion are not known, but may reflect engagement with the study, time available to complete the scale, and/or low literacy levels in the parents. This raises concerns both about the representativeness of the data and the importance of supporting completion in these disadvantaged and often multilingual settings. Secondly, our knowledge of children's home language use was limited. Information about the frequency with which the languages spoken at home was neither available nor did we have data on parents' English language competence, which limits are understanding of the impact of language status. Finally, although no measures exist to capture all the languages used in the settings we sampled, the LUI-6 was presented in the English language, and this may impact on parental understanding as well as completion rates.

CONCLUSIONS AND FUTURE DIRECTIONS

To close the gaps in the oracy skills of children from lower-SES homes, children need to develop language skills at an accelerated rate (86, 87). Early support in educational settings is critical to address this gap. We reasoned that a tool which supported teachers' ability to profile children's language skills would be an important lever to for these early years settings. Parents completed a measure which was simple to score and that provided a profile of children's language in an efficient way for practitioners. The shortened version has the potential to be used in a range of settings, but may be particularly useful in areas of social disadvantage with more EAL speakers and lower levels of literacy. In sum, the current study demonstrates that the LUI-6 English was an effective measure of language abilities in young children. The data also raise important new avenues for research to capture the language learning needs of children as they enter nursery school. The need to capture economically and linguistically diverse populations in interventions (88) requires the development and use of valid tools. The LUI-6 aimed to capture expressive vocabulary and grammar, and therefore missed other aspects of social pragmatic information. It may be that completion of the 10 main scales of the LUI could capture greater diversity in the population, especially if more reliable and valid data were collected about children's home language and their use and proficiency in this language. Given the diverse population of children that enter nurseries in areas of social disadvantage, providing teachers with ways of reliably capturing children's language performance to map progress and evaluate interventions is of paramount importance.

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 Institute of Education, UCL Ethics Committee. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

JD and JL conceptualized, designed, and managed the study. SM contributed to the work in the early years settings and staff needs. CF identified key subscales for the study and completed all pilot work for the abbreviated measure. CF and JC collected and entered data. All authors except JL were responsible for the final submission. All authors contributed to the article and approved the submitted version.

REFERENCES

1. O'Neill DK. 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.* (2007) 50:214–28. doi: 10.1044/1092-4388(2007/017)
2. Bruner J. Play, thought, and language. *Peabody J Educ.* (1983) 60:60–9.
3. Law J, Rush R, Schoon I, Parsons S. Modeling developmental language difficulties from school entry into adulthood: literacy, mental health, and employment outcomes. *J Speech Lang Hear Res.* (2009) 52:1401–16. doi: 10.1044/1092-4388(2009/08-0142)
4. Sim F, O'Dowd J, Thompson L, Law J, Macmillan S, Affleck M, et al. Language and social/emotional problems identified at a universal developmental assessment at 30 months. *BMC Pediatrics.* (2013) 13:206. doi: 10.1186/1471-2431-13-206
5. Law J, Rush R, King T, Westrupp E, Reilly S. Early home activities and oral language skills in middle childhood: a quantile analysis. *Child Dev.* (2018) 89:295–309. doi: 10.1111/cdev.12727
6. Nelson KE, Welsh JA, Trup EMV, Greenberg MT. Language delays of impoverished preschool children in relation to early academic and emotion recognition skills. *First Lang.* (2011) 31:164–94. doi: 10.1177/0142723710391887
7. Alsford E, Ralephata A, Bolderson S, Curtin M, Parish E, Klaber V, et al. The wrong side of the tracks: starting school in a socially disadvantaged London borough. *Child Lang Teach Ther.* (2017) 33:145–56. doi: 10.1177/0265659016654954
8. Hoff E. Interpreting the early language trajectories of children from low-SES and language minority homes: implications for closing achievement gaps. *Dev Psychol.* (2013) 49:4–14. doi: 10.1037/a0027238
9. Law J, Charlton J, Gascoigne M, Dockrell J, McKean C, Theakston A. *Early Language Development: Needs, Provision, and Intervention for Preschool Children from Socio-Economically Disadvantaged Backgrounds*. (2017). Available online at: https://educationendowmentfoundation.org.uk/public/files/Law_et_al_Early_Language_Development_final.pdf (accessed November 30, 2021).
10. Davies C, McGillion M, Rowland C, Matthews D. Can inferencing be trained in preschoolers using shared book-reading? A randomised controlled trial of parents' inference-eliciting questions on oral inferencing ability. *J Child Lang.* (2020) 47:655–79. doi: 10.1017/s0305000919000801
11. Fricke S, Bowyer-Crane C, Haley AJ, Hulme C, Snowling MJ. Efficacy of language intervention in the early years. *J Child Psychol Psychiatry.* (2013) 54:280–90. doi: 10.1111/jcpp.12010
12. Ebbels SH, McCartney E, Slonims V, Dockrell JE, Norbury CF. Evidence-based pathways to intervention for children with language disorders. *Int J Lang Commun Disord.* (2019) 54:3–19. doi: 10.1111/1460-6984.12387
13. Dobinson KL, Dockrell JE. Universal strategies for the improvement of expressive language skills in the primary classroom: a systematic review. *First Lang.* (2021) 41:527–54. doi: 10.1177/0142723721989471
14. Law J. The changing perception of communication needs-a litmus test for the Warnock legacy. *Front Educ.* (2019) 4:42. doi: 10.3389/feduc.2019.00042
15. Mathers S, Smees R. *Quality and Inequality - Do Three and Four Year Olds in Deprived Areas Experience Lower Quality Early Years Provision*. London: Nuffield Foundation (2014).
16. Phillips BM, Zhao YT, Weekley MJ. Teacher language in the preschool classroom: initial validation of a classroom environment observation tool. *Early Educ Dev.* (2018) 29:379–97. doi: 10.1080/10409289.2017.1408371
17. Mathers SJ. Using video to assess preschool teachers' pedagogical knowledge: explicit and higher-order knowledge predicts quality. *Early Childh Res Q.* (2021) 55:64–78. doi: 10.1016/j.ecresq.2020.10.010
18. Greenwood CR, Schnitz AG, Carta JJ, Wallisch A, Irvin DW. A systematic review of language intervention research with low-income families: a word gap prevention perspective. *Early Childh Res Q.* (2020) 50:230–45. doi: 10.1016/j.ecresq.2019.04.001
19. O'Neill DK. *The Language Use Inventory (LUI) and Manual*. Waterloo, ON: Knowledge in Development (2009).
20. Smith T, Noble M, Noble S, Wright G, McLennan D, Plunkett E. *The English Indices of Deprivation 2015*. London: Department for Communities and Local Government (2015).
21. Neuman SB, Kaefer T, Pinkham AM. A double dose of disadvantage: language experiences for low-income children in home and school. *J Educ Psychol.* (2018) 110:102–18. doi: 10.1037/edu0000201
22. Department for Education [DE]. *Statutory Framework for the Early Years Foundation Stage*. (2021). Available online at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/974907/EYFS_framework_-_March_2021.pdf (accessed December 01, 2021).
23. Dockrell JE, Marshall CR. Measurement issues: assessing language skills in young children. *Child Adolesc Ment Health.* (2015) 20:116–25.
24. Friberg JC. Considerations for test selection: how do validity and reliability impact diagnostic decisions? *Child Lang Teach Ther.* (2010) 26:77–92.
25. Hoffman LM, Loeb DF, Brandel J, Gillam RB. Concurrent and construct validity of oral language measures with school-age children with specific language impairment. *J Speech Lang Hear Res.* (2011) 54:1597–608. doi: 10.1044/1092-4388(2011/10-0213)
26. Law J, Roy P. Parental report of infant language skills: a review of the development and application of the Communicative Development Inventories. *Child Adolesc Ment Health.* (2008) 13:198–206.
27. Semel E, Wiig E, Secord W. *Clinical Evaluation of Language Fundamentals-UK (CELF-4 UK)*. New York, NY: Pearson Education (2006).
28. Dockrell JE, Howell P, Leung D, Furgad A. Children with speech language and communication needs in England: challenges for practice. *Front Educ.* (2017) 2:35. doi: 10.3389/feduc.2017.00035
29. OECD. *The Resilience of Students with an Immigrant Background: Factors that Shape Well-being, OECD Reviews of Migrant Education*. Paris: OECD Publishing (2018). doi: 10.1787/9789264292093-en

30. Strand S, Malmberg L, Hall J. English as an Additional Language (EAL) and Educational Achievement in England: An Analysis of the National Pupil Database. (2015). Available online at: https://www.bell-foundation.org.uk/app/uploads/2017/05/EALachievementStrand-1.pdf
31. Dockrell JE, Papadopoulos TC, Mifsud CL, Bourke L, Vilageliu O, Besic E, et al. Teaching and learning in a multilingual Europe: findings from a cross-European study. *Eur J Psychol Educ.* (2021) 1–28. doi: 10.1007/s10212-021-00539-z
32. Fenson L, Dale PS, Reznick JS, Bates E, Thal DJ, Pethick SJ, et al. Variability in early communicative development. *Monogr Soc Res Child Dev.* (1994) 59:i–185.
33. Bishop DVM. Development of the children's communication checklist (CCC): a method for assessing qualitative aspects of communicative impairment in children. *J Child Psychol Psychiatry Allied Discip.* (1998) 39:879–91. doi: 10.1017/s0021963098002832
34. Song S, Yang Y, Kim YT, Yim D. A Meta-analysis of children's communication checklist (CCC) for children & adolescents with pragmatic language impairment. *Commun Sci Disord.* (2016) 21:436–50. doi: 10.12963/csd.16302
35. Wellnitz SAC, Kastel I, Vllasaliu L, Cholemkery H, Freitag CM, Bast N. The revised children's communication checklist-2 (CCC-R): factor structure and psychometric evaluation. *Autism Res.* (2021) 14:759–72. doi: 10.1002/aur.2467
36. Pesco D, O'Neill DK. Predicting later language outcomes from the language use inventory. *J Speech Lang Hear Res.* (2012) 55:421–34. doi: 10.1044/1092-4388(2011/10-0273)
37. Bialecka-Pikul M, Filip A, Stepien-Nycz M, Kus K, O'Neill DK. Ratunku! or just tunku! Evidence for the reliability and concurrent validity of the language use inventory-polish. *J Speech Lang Hear Res.* (2019) 62:2317–31. doi: 10.1044/2019_jslhr-l-18-0230
38. Longobardi E, Lonigro A, Laghi F, O'Neill DK. Pragmatic language development in 18-to 47-month-old Italian children: a study with the language use inventory. *First Lang.* (2017) 37:252–66. doi: 10.1177/0142723716689273
39. Pesco D, O'Neill D. Assessing early language use by French-speaking Canadian children: introducing the LUI-French. *Can J Speech Lang Pathol Audiol.* (2016) 40:198–217.
40. Massonnié J, Llaurado A, Sumner E, Dockrell JE. Oral language at school entry: dimensionality of speaking and listening skills. *Oxf Rev Educ.* (2022) 1–24. doi: 10.1080/03054985.2021.2013189
41. Hindman AH, Wasik BA, Snell EK. Closing the 30 million word gap: next steps in designing research to inform practice. *Child Dev Perspect.* (2016) 10:134–9. doi: 10.1111/cdep.12177
42. Levine D, Pace A, Luo R, Hirsh-Pasek K, Golinkoff RM, de Villiers J, et al. Evaluating socioeconomic gaps in preschoolers' vocabulary, syntax and language process skills with the Quick Interactive Language Screener (QUILS). *Early Childhood Res Q.* (2020) 50:114–28.
43. Snowling MJ, Adams JW, Bishop DVM, Stothard SE. Educational attainments of school leavers with a preschool history of speech-language impairments. *Int J Lang Commun Disord.* (2001) 36:173–83. doi: 10.1080/13682820010019892
44. Hammer CS, Farkas G, Maczuga S. The language and literacy development of head start children: a study using the family and child experiences survey database. *Lang Speech Hear Serv Sch.* (2010) 41:70–83. doi: 10.1044/0161-1461(2009/08-0050)
45. Karlsen J, Hjetland HN, Hagtvet BE, Braeken J, Melby-Lervag M. The concurrent and longitudinal relationship between narrative skills and other language skills in children. *First Lang.* (2021) 41:555–72. doi: 10.1177/0142723721995688
46. Oracy All-Party Parliamentary Group [Oracy APPG]. *Speak for Change: Initial Findings and Recommendations from the Oracy All-Party Parliamentary Group Inquiry.* All Party Parliamentary Group (2020). Available online at: https://d5119182-bdac-43d5-be55-e817e7736e5b.filesusr.com/ugd/2c80ff_33e3208ce4dd4764b154682488c53ef7.pdf
47. Justice LM, McGinty AS, Zucker T, Cabell SQ, Piasta SB. Bi-directional dynamics underlie the complexity of talk in teacher-child play-based conversations in classrooms serving at-risk pupils. *Early Childh Res Q.* (2013) 28:496–508. doi: 10.1016/j.ecresq.2013.02.005
48. Petersen DB, Gillam SL, Spencer T, Gillam RB. The effects of literate narrative intervention on children with neurologically based language impairments: an early stage study. *J Speech Lang Hear Res.* (2010) 53:961–81. doi: 10.1044/1092-4388(2009/09-0001)
49. Vega V, Osman Y, Katz B, Giordano K. Hidden expectations: do teachers use expressive language ability as a proxy for overall development? *Early Child Dev Care.* (2020) 190:1682–90. doi: 10.1080/03004430.2018.1546176
50. Sachse S, Von Suchodoletz W. Early identification of language delay by direct language assessment or parent report? *J Dev Behav Pediatr.* (2008) 29:34–41. doi: 10.1097/DBP.0b013e318146902a
51. Emerson ND, Morrell HE, Neece C. Predictors of age of diagnosis for children with autism spectrum disorder: the role of a consistent source of medical care, race, and condition severity. *J Autism Dev Disord.* (2016) 46:127–38. doi: 10.1007/s10803-015-2555-x
52. Johnson CP, Myers SM. Identification and evaluation of children with autism spectrum disorders. *Pediatrics.* (2007) 120:1183–215.
53. Powell D, Son S, File N, San Juan R. Parent-school relationships and children's academic and social outcomes in public school pre-kindergarten. *J Sch Psychol.* (2010) 48:269–92. doi: 10.1016/j.jsp.2010.03.002
54. Bleses D, Makransky G, Dale PS, Hojen A, Ari BA. Early productive vocabulary predicts academic achievement 10 years later. *Appl Psycholinguist.* (2016) 37:1461–76. doi: 10.1017/s0142716416000060
55. Savage R, Kozakewich M, Genesee F, Erdos C, Haigh C. Predicting writing development in dual language instructional contexts: exploring cross-linguistic relationships. *Dev Sci.* (2017) 20:12406. doi: 10.1111/desc.12406
56. Dockrell JE, Stuart M, King D. Supporting early oral language skills for English language learners in inner city preschool provision. *Br J Educ Psychol.* (2010) 80:497–515. doi: 10.1348/000709910x493080
57. Hulme C, Snowling MJ, West G, Lervag A, Melby-Lervag M. Children's language skills can be improved: lessons from psychological science for educational policy. *Curr Direct Psychol Sci.* (2020) 29:372–7. doi: 10.1177/0963721420923684
58. Dockrell JE, Law J, Mathers S. *Empowering Staff to Enhance Oral Language in the Early Years. AEA RCT Registry*. (2021). doi: 10.1257/rct.4379-2.0
59. Bishop DVM. *An Index of Neighbourhood Advantage from English Postcode Data. BishopBlog*. (2018). Available online at: http://deevybee.blogspot.com/2018/09/an-index-of-neighbourhood-advantage.html (accessed November 30, 2021).
60. Elliott CD, Smith P. *British Ability Scales (BAS-3)*. 3rd ed. London: GL Assessment (2011).
61. Gardner H, Froud K, McClelland A, van der Lely HK. Development of the grammar and phonology screening (GAPS) test to assess key markers of specific language and literacy difficulties in young children. *Int J Lang Commun Disord.* (2006) 41:513–40. doi: 10.1080/13682820500442644
62. Conti-Ramsden G, Botting N, Faragher B. Psycholinguistic markers for specific language impairment (SLI). *J Child Psychol Psychiatry.* (2001) 42:741–8. doi: 10.1111/1469-7610.00770
63. Gathercole SE, Willis CS, Baddeley AD, Emslie H. The children's test of nonword repetition: a test of phonological working memory. *Memory.* (1994) 2:103–27. doi: 10.1080/09658219408258940
64. Renfrew CE. *Bus Story Test: A Test of Narrative Speech*. Milton Keynes: Speechmark (2010).
65. IBM Corp. *IBM SPSS Statistics for Windows, Version 25*. Armonk, NY: IBM Corp (2017).
66. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropract Med.* (2016) 15:155–63. doi: 10.1016/j.jcm.2016.02.012
67. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol Assess.* (1994) 6:284–90. doi: 10.1037/1040-3590.6.4.284
68. Tomblin JB, Zhang X. The dimensionality of language ability in school-age children. *J Speech Lang Hear Res.* (2006) 49:1193–208. doi: 10.1044/1092-4388(2006/086)
69. Law J, Clegg J, Rush R, Roulstone S, Peters TJ. Association of proximal elements of social disadvantage with children's language development at 2 years: an analysis of data from the Children in Focus (CiF) sample from the ALSPAC birth cohort. *Int J Lang Commun Disord.* (2019) 54:362–76. doi: 10.1111/1460-6984.12442
70. McKean C, Wraith D, Eadie P, Cook F, Mensah F, Reilly S. Subgroups in language trajectories from 4 to 11years: the nature and predictors of stable, improving and decreasing language trajectory groups. *J Child Psychol Psychiatry.* (2017) 58:1081–91. doi: 10.1111/jcpp.12790

71. Walker D, Sepulveda SJ, Hoff E, Rowe ML, Schwartz IS, Dale PS, et al. Language intervention research in early childhood care and education: a systematic survey of the literature. *Early Childh Res Q.* (2020) 50:68–85. doi: 10.1016/j.ecresq.2019.02.010
72. Apel K, Apel L. Identifying intraindividual differences in students' written language abilities. *Top Lang Disord.* (2011) 31:54–72.
73. Snow P. Oral language competence and the transition to school: socio-economic and behavioural factors that influence academic and social success. *Int J Sch Disaffect.* (2014) 11:3–24. doi: 10.18546/IJSD.11.1.01
74. Dockrell JE, Bakopoulou I, Law J, Spencer S, Lindsay G. Capturing communication supporting classrooms: the development of a tool and feasibility study. *Child Lang Teach Ther.* (2015) 31:271–86. doi: 10.1177/0265659015572165
75. Law J, Tulip J, Stringer H, Cockerill M, Dockrell J. Teachers observing classroom communication: an application of the communicating supporting classroom observation tool for children aged 4-7 years. *Child Lang Teach Ther.* (2019) 35:203–20. doi: 10.1177/0265659019869792
76. Zauche LH, Thul TA, Mahoney AED, Stapel-Wax JL. Influence of language nutrition on children's language and cognitive development: an integrated review. *Early Childh Res Q.* (2016) 36:318–33. doi: 10.1016/j.ecresq.2016.01.015
77. Cabell SQ, Justice LM, Piasta SB, Curenton SM, Wiggins A, Turnbull KP, et al. The impact of teacher responsivity education on preschoolers' language and literacy skills. *Am J Speech Lang Pathol.* (2011) 20:315–30. doi: 10.1044/1058-0360(2011/10-0104)
78. Odom SL, Wolery M. A unified theory of practice in early intervention/early childhood special education: evidence-based practices. *J Special Educ.* (2003) 37:164–73. doi: 10.1177/00224669030370030601
79. Galsworthy MJ, Dionne G, Dale PS, Plomin R. Sex differences in early verbal and non-verbal cognitive development. *Dev Sci.* (2000) 3:206–15. doi: 10.1111/1467-7687.00114
80. Craig HK, Washington JA. Oral language expectations for African American preschoolers and kindergartners. *Am J Speech Lang Pathol.* (2002) 10:59–70.
81. Viding E, Spinath FM, Price TS, Bishop DVM, Dale PS, Plomin R. Genetic and environmental influence on language impairment in 4-year-old same-sex and opposite sex twins. *J Child Psychol Psychiatry Allied Discip.* (2004) 45:315–25. doi: 10.1111/j.1469-7610.2004.00223.x
82. Bouchard C, Trudeau N, Sutton A, Boudreault M, Deneault J. Gender differences in language development in French Canadian children between 8 and 30 months of age. *Appl Psycholinguist.* (2009) 30:685–707. doi: 10.1017/S0142716409990075
83. Hammer CS, Hoff E, Uchikoshi Y, Gillanders C, Castro DC, Sandilos LE. The language and literacy development of young dual language learners: a critical review. *Early Childh Res Q.* (2014) 29:715–33. doi: 10.1016/j.ecresq.2014.05.008
84. Hoff E, Ribot KM. Language growth in English monolingual and Spanish-English bilingual children from 2.5 to 5 Years. *J Pediatr.* (2017) 190:241–5.e1. doi: 10.1016/j.jpeds.2017.06.071
85. European Commission. *Language Teaching and Learning in Multilingual Classrooms*. Luxembourg: European Commission (2015). doi: 10.2766/766802
86. Biemiller A, Boote C. An effective method for building meaning vocabulary in primary grades. *J Educ Psychol.* (2006) 98:44–62. doi: 10.1037/0022-0663.98.1.44
87. Marulis LM, Neuman SB. The effects of vocabulary intervention on young children's word learning: a meta-analysis. *Rev Educ Res.* (2010) 80:300–35. doi: 10.3102/0034654310377087
88. Walker D, Sepulveda SJ, Hoff E, Rowe ML, Schwartz IS, Dale PS, et al. Language intervention research in early childhood care and education: a systematic survey of the literature. *Early Child Res Q.* (2020) 50:68–85.

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.

frontiers | Frontiers in Pediatrics

ORIGINAL RESEARCH
published: 11 May 2022
doi: 10.3389/fped.2022.863477

Predicting Word Reading Deficits Using an App-Based Screening Tool at School Entry

Martin Schöfl[1,2]*, *Gabriele Steinmair*[1], *Daniel Holzinger*[2,3] *and Christoph Weber*[1,2]

[1] Department of Educational Sciences, University of Education Upper Austria, Linz, Austria, [2] Research Institute for Developmental Medicine, Johannes Kepler University Linz, Linz, Austria, [3] Institute of Neurology of Senses and Language, Hospital of St. John of God, Linz, Austria

Background: Reading is a crucial competence associated with academic development, mental health, and social adaptation. Reading difficulties are often detected at a late stage, with a possible negative impact on long-term reading development and secondary developmental disadvantages. The first manifestations of reading difficulties can be identified by word reading deficits in first and second grade, paving the way for specific interventions. For widespread implementation, instruments must be easy to use and motivating for children.

Objectives: Development and validation of an economical, well-accepted, and accurate screening tool composed of the domains of phonological information processing, language skills, and non-verbal intelligence in regular school settings.

Design: In 2020, the screening tool was used on a sample of 409 first graders between the second and fifth weeks of school in a one-to-one setting. Additionally, information on parental education and the use of German and/or other languages by the child was collected using a parental questionnaire. A follow-up involving the use of established standardized word reading tests was conducted at the end of the first school year.

Results: A five-variable screening tool consisting of the dimensions of phonological information processing (letter knowledge, rapid naming, and phonological awareness) and linguistic skills (receptive vocabulary and morphosyntax) showed statistical relevance (AUC = 0.78; sensitivity 0.80, specificity 0.74) for predicting word reading problems concerning reading speed (<16th percentile) at the end of first grade, whereas gender, first language, and age of first exposure to the German language did not contribute to the prediction. The instrument was well accepted by the children and screeners and can be administered within an acceptable time frame.

Conclusion: Word reading deficits at the end of first grade can be predicted by the use of an app-based screening tool at school entry that includes phonological information processing and language skills. Further validation and assessment of empirical feasibility data are needed to support the screening instrument for German orthography.

Keywords: word reading, primary school, predictive power, language, app-based screening

OPEN ACCESS

Edited by:
Maria Pia Bucci,
Centre National de la Recherche Scientifique (CNRS), France

Reviewed by:
Adrian Pasquarella,
University of Delaware, United States
Stephanie Ducrot,
Aix-Marseille Université, France

***Correspondence:**
Martin Schöfl
Martin.Schoefl@jku.at

Specialty section:
This article was submitted to Children and Health, a section of the journal Frontiers in Pediatrics

Received: *27 January 2022*
Accepted: *01 April 2022*
Published: *11 May 2022*

Citation:
Schöfl M, Steinmair G, Holzinger D and Weber C (2022) Predicting Word Reading Deficits Using an App-Based Screening Tool at School Entry. Front. Pediatr. 10:863477. doi: 10.3389/fped.2022.863477

INTRODUCTION

Word reading consists of several components: phonological analysis of the written word (1); orthographic processing in the sense of the "ability to form, store, and access orthographic representations" [(2), p. 4049], and lexical access to word meaning. In German orthography, which is characterized by low complexity and high transparency, reading difficulties manifest at an early stage as reduced word reading speed (3–10).

In an established reading model by Perfetti (11) and Perfetti and Hart (12), fluent word decoding and fast and effortless access to the orthographic lexicon predicted sentence and text reading fluency and thus represents an essential component of academic learning (13).

Reading weaknesses are associated with significant disadvantages throughout the school years and beyond, with impacts on school achievement (14, 15) and later employment (16). Nordström et al. (17) stressed the importance of schools investigating children's early word decoding ability. Another large-scale longitudinal study (18) followed up individuals who had weak word decoding at the age of 7, finding that they had lower school achievement and income as adults compared to good and average readers. Stanovich summarized the logic of this finding as early as 1986 as the "Matthew Effect": Good readers are intrinsically motivated to read and therefore read a lot, consequently, their reading skills continuously improve. Children starting school with poor reading skills often lack the motivation to read and consequently read less. Soon, a gap begins to open that is increasingly difficult to close.

Longitudinal studies have shown this very development in different orthographies [for German: (4, 5, 19, 20)].

An Australian research team conducted a comprehensive review of over 100 articles investigating the emotional consequences of slow reading in children over a period of 30 years, finding an increasingly negative impact on self-esteem and anxiety (21). German scientists replicated the findings of increased internalization of problems and resulting social withdrawal in children with reading or spelling deficits compared to children without learning disorders (22). Mammarella et al. (23) found that sustained academic failure and perceived low self-esteem increased the risk of anxiety and depression in children with reading problems. Earlier detection of risk factors connected with specific interventions could counteract this trend of reading deficits with consequences for education, employment, and wellbeing (24).

Screening tools [c.f. (25)] need to meet the following criteria: to be stable over time; to accurately predict reading achievement (criterion: "validity"); and to be objectively applicable, evaluable, and interpretable (criterion: "objectivity"). Additional criteria relate to their application in schools: screening must require little training of the test instructors; their administration should be time-efficient and administrable with limited staff resources (criterion: "test economy"). Furthermore, screening should be motivating and not overburdening for the children (criterion: "reasonableness"), and no child should be disadvantaged by the way it is conducted or the language used (criterion: "fairness"). Finally, the results should be available to teachers quickly and unambiguously and should allow conclusions to be drawn for schools, such as assignments to support groups or the adaptation of teaching methods (criterion: "usefulness").

The use of app-based screening technology by children around school-entry age appears promising in terms of both test economy and feasibility, as demonstrated by the assessment of vocabulary performance in the last year of kindergarten (26) in Austria. Internationally, acceptable clinical screening accuracy is reported only close to or at school entry and not in the prior years (27). This is because data on the highly relevant and directly literacy-related factors (e.g., letter knowledge) can only be collected close to the beginning of school entry, in addition to more general predictors, such as non-verbal intelligence or linguistic skills.

Child-Related Predictors

Linguistically based skills on the one hand and visual skills on the other hand have been found to predict word reading [for an overview, see (28)]. Predictors associated with visual processing such as visual memory span at kindergarten age (29) have been researched experimentally, but to our knowledge, there is still a lack of established test paradigms shown to be feasible within school-based screenings. Related to linguistically based predictors in alphabetic languages, letter knowledge, phonological awareness, and Rapid Automatized Naming-speed (RAN) have been demonstrated as robust predictors of word reading even across different orthographies and a number of reliable and practicable test paradigms have been developed (30). These factors, often summarized as phonological information processing (31, 32), are frequently supplemented by phonological working memory (33). Only recent studies have focused on the prerequisites for these factors, namely, linguistic skills. In a longitudinal study, Snowling et al. (34) demonstrated the influence of lower levels of linguistic competencies on the development of specific learning disorders. Moreover, linguistic deficits in already-diagnosed reading problems are observed retrospectively by parents more often than in average or good readers (35). Thus, asking parents about their children's language performance and assessing it as an additional factor at school entry is anticipated to be a central component of a valid screening tool for written language skills. Non-verbal IQ as a general predictor of school success contributed little to direct variance explanation of word reading or writing performance in previous longitudinal studies. Rather, non-verbal IQ determined the level of profiles in profile analyses, such as the large-scale study by Ozernov-Palchik et al. (36). Non-verbal IQ as a general predictor of school success contributed little to direct variance explanation of word reading or writing performance in previous longitudinal studies. Rather, non-verbal IQ determined the level of profiles in profile analyses, such as the large-scale study by Ozernov-Palchik et al. (36). Latent profile analysis in kindergarteners showed specific effects and interactions of the known predictors RAN, phonological awareness, verbal working memory, and letter knowledge. The level of (non-verbal) IQ helped to identify groups of children with average or overall (non-specific) slightly below-average performance. With regard to screening IQ as an additional criterion does not lead to better identification of children that require specific reading promotion and is therefore not included.

Environmental Factors

Environmental factors influencing children's reading competence have been highlighted in established reading socialization models [e.g., the multilevel model of family reading Hurrelmann et al. (37)]. They describe learners' reading experiences in different social contexts and their influence on the development of motivation, interests, and skills. Rosebrock and Nix's [(38), p. 16] reading literacy model includes three levels of reading competence: subjective, cognitive, and social (39). At the social level, the influence of the family as language and reading role models is emphasized as a moderator variable.

Lack of familiarity with the language spoken at school might be another factor that could affect word reading due to underspecified phonological representations or irregular letter-sound correspondence for L2 (40). Growing up with a primary language (L1) other than that used at school (L2) is usually related to having an immigrant background, including culture-specific home environments related to literacy. Although the majority of studies have found similar word reading skills in these children compared to their native peers (41–43), only a few reported better outcomes for native students (44, 45). For Dutch, there is a body of evidence in support of word decoding from kindergarten being highly comparable in Dutch as the first language (L1) and Dutch as the second language (L2) learners (46). Nevertheless, group differences at different stages of reading development have been documented, such as differences between L1 and L2 learners on rapid naming assessments in Grade 1, which disappeared in Grade 2 (47). For German (48), German L2 learners' reading fluency was mostly predicted by non-verbal intelligence, whereas for L1 learners, phonological awareness tasks in the last year of kindergarten best predicted reading fluency.

Scientific Aim

The aim was to construct and validate a time-efficient screening tool for word reading ability for use in community school settings around school entry. Child-related predictors concerned phonological information processing (phonological awareness, letter knowledge, and phonological working memory), language (vocabulary and grammar), and non-verbal intelligence. Children's gender, additional environmental predictors, first language, and exposure to the German language were analyzed as potential moderating variables.

MATERIALS AND METHODS

Participant Recruitment

The majority of the recruited children were from an Upper Austrian district with four big community-based schools. They were informed and invited to participate in the study project firstly via telephone and then by personal visits. All headmasters agreed to participate. Additionally, four more schools asked for participation and joined the project. Finally, parents of 409 children (100%) gave their written permission for including their children in the study. The final study sample reflected a heterogeneous distribution of children, comparable with Austrian primary schools in terms of gender, the proportion of children with a non-German dominant language, and parental educational levels.

The individualized screening started in autumn 2020 within 2 weeks after school onset. Within 3 weeks, 85% of the sample had been assessed. In the two subsequent weeks, those children who had been ill or unable to attend during the first survey period were surveyed. A total of 27 children were not included in the analysis because they were listed as "Vorschüler" (preschoolers) in their first year of learning. A total of 86 children were sick or out of school when the reading test was conducted at the end of the first school year; these children did not differ significantly from the analyzed sample in terms of age, gender, and most importantly, the screening variables. At the end of first grade, word reading was assessed in the classroom setting. **Figure 1** shows the recruitment pathways and timeline.

Participant Characteristics

In school statistics for the 2019/2020 school year, of the 344,282 Austrian elementary school children, 48.2% were female (49). Nearly the same proportion is found in the present study, where 48.9% (N = 407) are female.

In the 2019/2020 school year, 106,498 out of a total of 344,282 children in Austrian elementary schools had German not as their first language, which corresponds to 26.8% (49). Hence, around seven out of ten children in Austrian elementary schools have German as their first language [(49), p. 12]. In all nine schools studied in the project, 74.2% of children speak German exclusively as a first language. The proportions from the research project thus correspond to the Austrian distribution.

Socioeconomic status is approximated by parents' highest educational attainment. The sample consisted of parents from all educational backgrounds: among the mothers, 4.9% had maximum educational attainment of an elementary school diploma, 14.4% had a high school diploma, and 29.8% had a university diploma. Educational levels for fathers were comparable (6.9% with maximum educational attainment of a primary school diploma, 15.4% with a high school diploma, and 21.5% with a university degree). Overall, the educational level of the Austrian population are comparable: 6% of parents have the highest educational attainment in elementary school, 22% in high school, and 27% have a university diploma. The given sample contains a variety of educational levels.

Measures and Procedure

This 1-year prospective study followed children from the beginning to the end of first grade. The research design included two steps:

(1) Screening of phonological information processing, intelligence, and language in the first weeks of first grade, before the formal teaching of reading and spelling, had

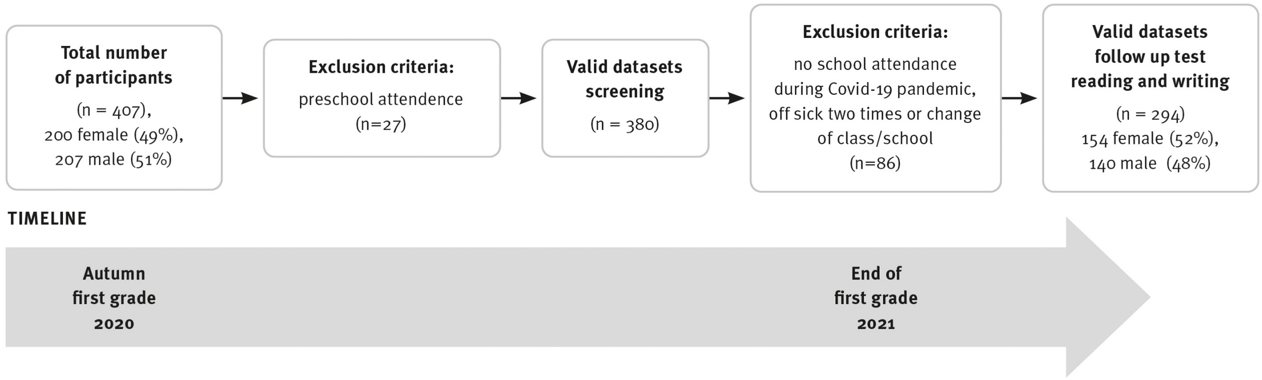


FIGURE 1 | Recruitment pathways and timeline.

begun. The classification of the first spoken language was done by means of a questionnaire by the parents. If it was indicated that the first language was only German or contact with German occurred from birth up to and including the age of 2, then children were classified as L1. Children whose contact with German occurred only after the age of 2 were classified as L2.

(2) Standardized assessment of word reading at the end of first grade.

Screening Measures

The screening tool consisted of 13 subtests, which can be systematized into the three domains of phonological information processing, language, and non-verbal intelligence. A total of seven of the subtests are well-established standardized tests, whereas six of the screening tests have been newly designed; see **Table 1** for an overview of the tasks. All tasks were app-supported, although for some subtests the child had a paper version to look at or an audio presentation of the stimuli was played to them on a tablet, which was used to enter results.

Phonological Information Processing

Phonological awareness was assessed by three tasks, with one task intended to differentiate in the lowest performance range (rhyming), one in the middle (syllable count), and one close to written language acquisition (initial phoneme detection). Phonological awareness tasks were newly constructed despite the presence of existing tests in order to meet the quality criterion of the economy for the use of the instrument in the school setting. Existing test procedures in German-speaking countries are well constructed but are too time-consuming for universal use in schools [e.g., (50)]. Each of the tasks was introduced by three practice items including feedback, followed by ten test items. Tasks were constructed from high-frequency words from the childLex database (51) for the youngest age group (6–8 years). Syllable count was controlled for the target items, and distractor tasks consisted of phonologically similar structures to the target items.

For the rhyming task, the child selected the words that rhyme from a set of three words (picture and word presented). Ten examples (7 one-syllable and 3 two-syllable) were presented. For example: "What rhymes: house, mouse, man?" An explorative factor analysis (EFA) for binary items conducted with Mplus 8 (52) showed a dominant factor with an eigenvalue of 5.351 (53% explained variance). A second factor with an eigenvalue of 1.418 was not interpretable. Moreover, the one-factor solution yielded an acceptable fit [$\chi^2(35) = 0.9405$, root mean square error of approximation (RMSEA) = 0.065; comparative fit index (CFI) = 0.918]. Therefore, we choose the single-factor solution. Internal consistency was (Kuder–Richardson KR-20) low but acceptable at 0.610. Internal consistency was (Kuder–Richardson KR-20) low but acceptable at 0.61.

Syllable count was also assessed by one- to four-syllable words, presented by means of pictures and spoken language; visual cues (one clapping a hand to four clapping hands) were used to indicate the number of syllables. Notably, an EFA yielded an inadequate fit for a one-factor solution [$\chi^2(35) = 211.4$, RMSEA = 0.112, CFI = 0.850], but a good fit for a two-factor solution [$\chi^2(26) = 38.9$, RMSEA = 0.035, CFI = 0.989]. The analyses revealed a factor focusing on two or more syllable words (7 items; eigenvalue = 4.563, 46% explained variance) and a factor focusing on one-syllable words (3 items; eigenvalue = 1.876, 19% explained variance). Thus, we used two different syllable count scores in this paper. Internal consistency was good for the one-syllable factor (KR-20 = 0.764) and acceptable for the two or more-syllable factor (KR-20 = 0.651).

For initial phoneme detection, we presented a letter visually and as a speech sound simultaneously.

From a selection of three pictures, those with the same first phoneme had to be selected ("Which word begins with I like Ines: Hase, Igel, Spiegel?"). Although an EFA yielded three factors with eigenvalues greater than 1 (3.772, 1.418, 1.064), the one-factor solution showed an adequate fit [$\chi^2(35) = 74.3$, RMSEA = 0.053, CFI = 0.908]. Thus, we choose the single-factor solution. Internal consistency was adequate (KR-20 = 0.632).

For the assessment of Rapid Automatized Naming (RAN), two conditions were chosen: objects and digits. The RAN

TABLE 1 | Subtests and domains of the screening measures.

Domain	Subtest	Type of subtest	Number of practice items	Number of test items	Presentation mode	Target selection mode
Phonological information processing	Phonological awareness	Newly designed	3	10	Tablet	Children using tablet
	Rhyme detection					
	Phonological awareness syllable count	Newly designed	3	10	Tablet	Children using tablet
	Phonological awareness	Newly designed	3	10	Tablet	Children using tablet
	Initial phoneme detection					
	Rapid Automatized Naming, RAN (1)	Newly designed	5	30	Paper	Instructor
	Objects					
	Rapid Automatized Naming, RAN (2)	Denckla and Rudel (53) following Landerl et al. (54)	5	30	Paper	Instructor
	Digits					
	Letter knowledge	Newly designed	None	26	Paper	Instructor
	Phonological working memory	Newly designed	None	Adaptive	Tablet	Instructor
	Word list memory					
	Phonological working memory	IDS-II, Grob and Hagmann-von Arx (55)	None	Adaptive	Instructor	Instructor
	Letter–number-span forward					
	Phonological working memory	IDS-II, Grob and Hagmann-von Arx (55)	None	Adaptive	Instructor	Instructor
	Letter–number-span backward					
Linguistic skills	Receptive vocabulary	GraWo; Seifert et al. (58)	2	30	Tablet	Children using tablet
	Sentence repetition	Adapted from Hamann and Abend Ibrahim (57)	None	15	Tablet	Instructor
Intelligence	Complete matrices	PITVA (59)	None	Adaptive	Paper	Instructor
	Picture series	PITVA (59)	None	Adaptive	Paper	Instructor

object condition was designed through five high-frequency monosyllabic words (cow, hand, ice, tree, and mouse). First, the items of the RAN tasks were presented app-based, and the task was given to repeat these items. Once the investigator ensured that the instruction was understood and the items were known, the test session started. The items were presented on paper repeatedly in a different order over six lines. The investigator pressed a button on the tablet to time the test and noted any incorrect responses on the tablet by pressing a button. When the last item was reached, the time measurement was stopped manually again, and the time distance was calculated automatically.

Rapid Automatized Naming in the digit condition was based on the work of Denckla and Rudel (53), following (54), and was presented and rated analogously to the object condition with monosyllabic digits (2, 8, 1, 6, 3).

Letter knowledge: All letters of the alphabet were offered as capital letters in random order on paper. Each page contained three to four letters. Children were asked: "I know you haven't learned these letters yet at school. Maybe you still know one? Please name it!." Positive scores were given for letter names or sounds and ticked off on the tablet.

Phonological working memory was assessed by two subtests of a broad-range intelligence test battery [IDS-2; Intelligence and Development Scales for Children and Adolescents, (55)] testing memory of letter-number sequences forward and backward. The child was asked to repeat a series of digits mixed up with letters (3-A, 5-M-2) in the same (forward condition) way or form back and forth (backward condition). The investigator clicked correct solutions on the tablet. The difficulty level of the tasks was determined by the length of the spans, and the termination occurred after three unsolved or incorrectly solved tasks. The longest possible range of letter and number sequences was of the target value. Reliability is described as fair; Cronbach's Alpha was 0.89 (end of first grade). Retest reliability was rtt = 0.93 (first grade).

Wordlist memory was also used to analyze phonological working memory: a list of 10 words (5 single-syllable words and 5 two-syllable words) was presented *via* an audio file. The child was then asked to freely reproduce as many of them as possible. The investigator ticked off the words in the mentioned order (including repetitions and wrong words). The sum of all memorized items yielded the overall performance.

Language

Morphosyntactic skills were assessed by an adapted sentence-repetition task. The German version was constructed according to the LITMUS (Language Impairment Testing in Multilingual Children) principles (56) by Hamann et al. (57) following the COST Action IS0804. A block of 15 items representing morphosyntactic constructions with varying degrees of complexity was selected and scored according to whether or not the sentence was completed correctly. Correctness was judged and noted in the app by the examiner when the sentence structure was reproduced completely correctly, regardless of articulatory deficits.

Internal consistency (KR-20) was high at 0.877.

In a digital form of the Graz Vocabulary Test [GraWo; (58)], receptive vocabulary was tested by 30 matching tasks. The child was required to choose from four pictures the one that matched the audio-presented word. Reliability data are given for the paper form of the GraWo: Cronbach's Alpha ranged from 0.89 (end of first grade) to 0.82 (end of second grade). Retest reliability was rtt = 0.93 (first grade).

Non-verbal Intelligence

Two subtests of the PITVA (59) were used to assess non-verbal intelligence: Complete Matrices (Cronbach's Alpha 0.83 for 6 year old/Retest reliability rtt = 0.9) and Picture Series (Cronbach's Alpha 0.86). The child was shown matrices and sequences of items and had to click on the correct condition from a selection directly on the tablet.

Word Reading

Word reading and writing tasks were administered in a classroom setting at the end of the first school year, exclusively by research staff.

The ELFE II word reading test (60) was used to assess decoding fluency at the word level in silent reading. For each picture, the appropriate written word from a selection of four had to be selected. The test duration was limited to 3 min. Representative norm scores are available from the end of the first school year to the beginning of the seventh grade; reliability data are presented as excellent (split-half $r = 0.98$, retest $r = 0.83$). A cut-off score of 13 represents M – 1 SD.

Procedure

Before the implementation of the screening tool, the principals of participating schools received information about the testing process. They were also given a letter to send to parents, including consent forms and questions about children's first language and language use as well as parents' educational background.

Teachers entered children's names into an online database, which converted the names to IDs for use on the tablet. The testers, all of them were student teachers, were enrolled in a student seminar in order to learn about the tasks and testing procedure, through which the teachers received student credit for the study (amounting to 4 h). The materials for testing were brought to the schools by a research coordinator. On the test mornings, it was agreed with the school administration that the children would be selected alphabetically by the test team (student teachers and core study project staff) from the classrooms. The assessment took place one at a time, with the child and instructor seated across a table from each other. After a brief welcome, the child was handed the tablet, in which the friendly dragon SCHWUPP was introduced right at the beginning. The app navigation was designed in a way that the child can use it independently, but if necessary, the test leader intervened in the navigation of the dragon from one task to the next. All instructions essential for the child were recorded as audio files, opened automatically, and could be repeated if necessary. A yellow background on the app signaled to the test administrator that the child was making test selections independently (such as in the phonological tasks). A gray background meant that the test administrator had to take the tablet to read the instructions from the tablet and give the corresponding instructions. This was especially true for tasks with material (for example, the letter cards). The assessment including all subtests took an average of 38.4 min (SD = 9.3) per child.

METHODS OF ANALYSIS

First, we used receiver operator characteristic (ROC) analyses to evaluate the diagnostic accuracy of each subscale. Following Swets (61), AUCs $\geq$ 0.9 are regarded as excellent, AUCs $\geq$ 0.8 and <0.9 as good, AUCs $\geq$ 0.7 and <0.8 as fair, and tests with AUCs < 0.7 as poor. ROC analyses were conducted using the pROC package (62) in R.

Second, to construct a time-efficient screening to predict word reading difficulties at the end of Grade 1, we used a logistic regression model with adaptive variable selection to identify important subtests. In detail, we applied the least absolute shrinkage and selection operator [LASSO; e.g., (63, 64); for an application of LASSO for the selection of screening variables, see (65)] as implemented in the glmnet R Package (66), which adequately addresses the problem of overfitting that is pronounced in standard variable selection procedures (e.g., backward or forward selection) and models with many predictors (relative to the sample size). Overfitting occurs when sample regression estimates capture signal and noise and thus are larger than in the population, which in turn limits the generalizability of the regression results. LASSO addresses overfitting and consequently increases generalizability by applying a penalty term (λ) to the likelihood function that protects estimates from inflation. Just as in the backward or forward selection, null predictors are zeroed out (i.e., they are excluded from the prediction model). Notably, LASSO does not provide *p*-values (methods have been developed for linear models, but not for logistic models) and thus it is not possible to refer to the "significance" of predictors. Instead, the selected predictors are meaningful whether their effects are significantly different from zero or not (63, 65).

To evaluate the importance of the screening variables, we z-scored the predictors. Thus, reported estimates are in a standardized metric. Moreover, LASSO requires the selection of an appropriate penalty term. We used 10-fold cross-validation and selected the value for λ that resulted in the highest area under the curve (AUC). Since in some cases this may insufficiently address the problem of overfitting, we also report results for the second value of λ by applying the one standard error rule [i.e., selecting the largest value of λ at which the AUC is within one standard error of the largest AUC; see e.g., (64), p. 216]. Once a set of predictors had been selected, we used the regression coefficients of the LASSO models to estimate the probability of scoring within the 10%-percentile of the word reading test at the end of Grade 1. This probability score (ranging from 0 to 1) was subsequently used as a screening measure.

Third, we compared the screening scores based on the results for λ at the maximum AUC with scores based on the one standard error rule by applying a bootstrapped test that compares the AUCs of paired ROC curves (62).

Fourth, we compared ROC curves between groups defined by the first language (German vs. non-German), German language exposure (≤2 years vs. >2 years), and gender (girls vs. boys). Significantly differing ROC curves between groups indicate variations in diagnostic accuracy, which would consequently limit the generalizability of the screening tool (67). Besides using a bootstrapped test for unpaired ROC curves that compare the AUCs for two groups, we also applied the Venkatraman permutation test (68) that compares actual ROC curves (also implemented in the pROC package). If two ROC curves do not differ significantly between groups, screening scores would yield the same sensitivity and specificity in both groups, and thus, a single cut-off for both groups would be appropriate.

Finally, we determined optimal cutoff scores using the R-OptimalCutpoints package (69). Cut-offs were evaluated based on the following diagnostic accuracy statistics: sensitivity (Se), specificity (Sp), positive predictive values (PPV), negative predictive values (NPV), and diagnostic likelihood ratios for positive and negative screening results (DLR+ and DLR−, respectively). Se and Sp ≥0.9 indicate good diagnostic accuracy, Se and Sp ≥0.80 are regarded as fair, and values below 0.80 indicate an unacceptably high rate of misclassification (70). DLR+ and DLR− are diagnostic accuracy measures that—unlike predictive values—do not depend on the prevalence of the disorder under investigation (71). DLR+ displays the multiplicative change in the pre-screening odds of scoring in the 10%-percentile of the reading test given a positive screening result (i.e., post-screening odds = DLR+ × pre-screening odds). DLR− is the change in the pre-screening odds of scoring in the 10%-percentile given a negative screening result (post-screening odds = DLR− × pre-screening odds). DLR+ values ≥10 and DLR− ≤0.1 indicate large changes in pre-screening odds, DLR+ ≤10 and >5, and DLR− >0.1 and ≤0.2 indicate moderate changes, DLR+ ≤5 and >2, and DLR− >0.2 and ≤0.5 indicate small changes. DLR+ <2 and DLR− >0.5 are rarely important (72).

RESULTS

Table 2 (Section A) shows the AUCs as well as the point-biserial correlations (r_{pb}) for the screening subtests. Notably, rhyme detection, syllable count, letter–number sequences forward, and the IQ subtests are not significantly associated with word reading problems at the end of Grade 1. For all other predictors, correlations are small and only the AUCs for RAN (digits and objects) and letter knowledge could be regarded as fair. Overall, the AUC for RAN objects is largest at 0.726 (DeLong 95%-CI [658, 0.795]), directly followed by letter knowledge (0.723, DeLong 95%-CI [0.645, 0.801]).

The results of the LASSO logistic regression models are reported in **Table 2** (Section B). When selecting the value for the penalty term λ that yields the highest AUC (Model 1),

TABLE 2 | Areas under the curves (AUCs) for subtests and results of the LASSO logistic regression models.

	Section A			Section B	
				Lasso Model 1 – AUC = MAX	Lasso Model 2 – 1 SE rule
	r_{pb}	AUC	95%-CI DeLong	Estimate (OR)	Estimate (OR)
Phonological awareness					
Rhyme detection	−0.042	0.521	(0.433, 0.610)		
Syllable count (one syllable)	−0.075	0.552	(0.467, 0.637)		
Syllable count (two or more syllables)	−0.086	0.555	(0.467, 0.643)		
Initial phoneme detection	−0.211***	0.663	(0.588, 0.738)	−0.061 (0.941)	
Rapid Automatized Naming					
RAN objects	0.287***	0.726	(0.658, 0.795)	0.312 (1.366)	0.174 (1.190)
RAN digits	0.255***	0.717	(0.632, 0.802)	0.098 (1.103)	
Letter knowledge					
Letter knowledge	−0.280***	0.723	(0.645, 0.801)	−0.380 (0.684)	– 0.179 (0.836)
Phonological working memory					
Word list memory	−0.149**	0.601	(0.517, 0.686)		
Letter–number sequences forward	−0.084	0.589	(0.506, 0.671)		
Letter–number sequences backward	−0.184***	0.627	(0.541, 0.714)		
Linguistic competences					
Vocabulary	−0.184***	0.615	(0.523, 0.706)		
Sentence repetition	−0.189***	0.642	(0.556, 0.727)	−0.040 (0.961)	
Intelligence					
Subtest A	−0.072	0.581	(0.504, 0.658)		
Subtest B	−0.095	0.566	(0.477, 0.654)		
Intercept				−1.675	−1.567

***p < 0.01, ***p < 0.001.*

the model identifies five non-zero predictors: initial phoneme detection ($b = -0.061$), RAN objects ($b = 0.312$), RAN digits ($b = 0.098$), letter knowledge ($b = -0.380$), and sentence repetition ($b = -0.04$). Remember that all subtests were z-scored, thus the strength of the regression coefficients could be directly compared. Applying the one standard error rule for the selection of λ (Model 2) results in two non-zero predictors. RAN objects ($b = 0.174$) and letter knowledge ($b = -0.179$)—the strongest predictors of Model 1—were selected as meaningful predictors.

The screening score based on LASSO Model 1 yields an AUC of 0.783 (DeLong 95%-CI [0.713, 0.852]), and the AUC for the screening score based on LASSO Model 2 is 0.773 (DeLong 95%-CI [0.704, 0.843]). A bootstrapped test for paired ROC curves indicates that the AUC-Difference is statistically not significant ($\Delta AUC = 0.01$, $D = 1.317$, $p = 0.188$). However, as AUC-difference tests are known to be plagued with low power [e.g., (73)], we decided to further evaluate the screening scores based on LASSO Models 1 and 2.

Table 3 reports the results for the comparison of ROC curves between groups. The tests for unpaired ROC curves show no significant differences between the groups. Thus, the screenings based on the LASSO selected predictors show no differences in diagnostic accuracy between German and non-German-speaking children, children with German-language exposure ≤2 years, children with German-language exposure >2 years, and girls and boys.

Finally, we estimated cutoffs for both screening scores by setting the sensitivity equal to 0.80. This cutoff was chosen to achieve acceptable sensitivity while holding the rate of positive screens as low as possible. The diagnostic accuracy statistics are reported in **Table 4**. Both screening scores yield a sensitivity of 0.808 (i.e., the cutoff value that achieves a sensitivity closest to 0.8). For Model 1, the cutoff is 0.195. This cutoff results in 36.3% of screening fails. Notably, given that the screening scores based on Model 1 and Model 2 achieve an identical sensitivity, the other diagnostic accuracy statistics are favoring Model 1. Importantly, as indicated by a significant McNemar Test [$\chi^2(1) = 4.923$, $p < 0.05$], the Model 1 screening turns out to be significantly more specific than the Model 2 screening (Model 1: Sp = 0.733, 95%-CI [0.672, 0.787], Model 2: Sp = 0.695, 95%-CI [0.633, 0.753]).

For Model 2, the cutoff of 0.186 results in 39.3% of screening fails. Given that the screening scores based on Model 1 and Model 2 achieve an identical sensitivity, the other diagnostic accuracy statistics of Model 1 are better than those of Model 2. Notably, as indicated by a significant McNemar Test [$\chi^2(1) = 4.923$, $p < 0.05$], the Model 1 screening turns out to be significantly more specific (Model 1: Sp = 0.733, 95%-CI [0.672, 0.787], Model 2: Sp = 0.695, 95%-CI [0.633, 0.753]).

DISCUSSION

Constructing a New Screening Battery

The analysis of a broad battery of subtests aiming to predict word reading deficiencies at the end of first grade resulted in two models: a short one consisting of two subtests (AUC = 0.77) and a broader one with five subtests (AUC = 0.78). Both models include a task for rapid naming and letter knowledge; the broader version additionally includes two language subtests (vocabulary and grammar) and a short assessment of phonological awareness (first phoneme detection).

There is the general consensus about the acceptable test accuracy of developmental screenings, namely, a sensitivity of 0.80 and specificity of 0.70 (74, 75). Thus, moderate specificity values may be acceptable, but high sensitivity is demanded for universal screening (76, 77). Setting a sensitivity range of 0.80, the specificity of the short version with two predictors is 0.69, and the

TABLE 3 | Comparing receiver operating characteristics (ROC) curves between subsamples.

		LASSO Model 1			LASSO Model 2		
		AUC	95%-CI	AUC-Difference (2)	AUC	95%-CI	AUC-Difference (2)
First language	German (1)	0.768	(0.671, 0.866)		0.761	(0.664, 0.857)	
	Non-German (2)	0.786	(0.680, 0.892)	$E = 0.038$, $p = 0.502$ $D = 0.248$, $p = 0.804$	0.776	(0.667, 0.885)	$E = 0.041$, $p = 0.417$ $D = 0.202$, $p = 0.8401$
German language exposure	≤2 years (1)	0.767	(0.681, 0.856)		0.760	(0.674, 0.846)	
	>2 years (2)	0.810	(0.687, 0.934)	$E = 0.015$, $p = 0.780$ $D = -0.550$, $p = 0.582$	0.790	(0.659, 0.920)	$E = 0.014$, $p = 0.808$ $D = -0.385$, $p = 0.700$
Gender	Boys (1)	0.755	(0.643, 0.866)		0.744	(0.632, 0.856)	
	Girls (2)	0.809	(0.721, 0.897)	$E = 0.017$, $p = 0.695$ $D = -0.754$, $p = 0.451$	0.803	(0.716, 0.890)	$E = 0.02$, $p = 0.576$ $D = -0.834$, $p = 0.405$

E-value for AUC-Difference refers to the Venkatraman test that compares ROC curves and the D-value refers to the bootstrapped test for paired ROC curves that compares AUCs.

TABLE 4 | Diagnostic accuracy statistics.

	Se (95%-CI)	Sp (95%-CI)	PPV (95%-CI)	NPV (95%-CI)	DLR+ (95%-CI)	DLR− (95%-CI)
LASSO Model 1	0.808 (0.675, 0.904)	0.733 (0.672, 0.787)	0.393 (0.326, 0.591)	0.947 (0.898, 0.960)	3.012 (2.360, 3.865)	0.263 (0.150, 0.461)
LASSO Model 2	0.808 (0.675, 0.904)	0.695 (0.633, 0.753)	0.362 (0.300, 0.559)	0.944 (0.893, 0.957	2.652 (2.104, 3.344)	0.277 (0.157, 0.486)

specificity of the extended model with three additional subtests is 0.74. Yielding significantly higher specificity, the extended model is thus the preferred model.

The positive predictive value of Model 1 is 0.36; that is, 36% of the children with low screening results are in the slow readers' group at the end of Grade 1. The preferable five-variable model identifies 39% of the children with low reading results at the end of first grade, identifying 42 children at risk as low readers correctly. The achieved predictive values represent an improvement compared to the only recent assessment for preschoolers in German, the "LRS-Screening" (50) which uses a range of 14 subtests in the year prior to starting school to predict word reading deficits at the end of first grade, with a sensitivity of 0.74, a specificity of 0.68, and a PPV of 0.27. The "LRS-Screening" is presented in a one-to-one setting in a paper–pencil version. It lasts for a duration of about 30 min, requires additional scoring time, and does not provide a cover story that would be assumed to make the assessment more appealing. In comparison, the five-component model of the newly developed screening tool can now be administered in about 15 min, including a shortened cover story, and is the only app-based screening tool for the early identification of reading problems in German. The "Bielefelder Screening" (78) is widely used in the year prior to school entry. In the manual, good predictive values (as high as 50%) are quoted, which could not be replicated by independent studies (79). Another screening, designed for group assessment during the last year of kindergarten, is the Phonological Awareness-Reading and Spelling Screening [PB-LRS; (80)]. The authors reported a sensitivity of 63%, specificity of 87%, and PPV of 36%. The duration of this screening tool with acceptable predictive quality is about 60 min.

Another established screening tool is called "Tour through Hörhausen" (81), which provides a phonological assessment of children in one-to-one settings at the beginning and the midpoint of the first grade. Prognostic validity was analyzed using a sample of 375 children, focusing on word reading speed at first grade. The authors described its specificity as over 80%, whereas the sensitivity varies between 38 and 48%. The assessment time is about 40 min. In summary, established screenings to predict reading difficulties in German require a long administration time and demonstrate low predictive power [for an overview, see (82)].

With the five-component model, there is no significant difference in the prediction of reading deficits according to gender or first language (German or non-German). Therefore, no specific cutoffs for gender or first language are needed.

Phonological awareness is one component of phonological information processing that is highly significant for the prediction of reading deficits in the international English-dominated literature [e.g., (83) for an overview]; in more consistent orthographies, such as Italian or German, word reading deficits are primarily predicted by the measures of letter knowledge and RAN (54, 84, 85). The prediction of reading performance at word level by vocabulary and grammar, summarized as linguistic competencies (34), was confirmed in the present study for the German language.

Interestingly, the factor of non-verbal intelligence plays a subordinate role; in the statistical model, it does not attain significance. In German-language longitudinal studies, the predictive quality of non-verbal intelligence on reading fluency was minimal (86, 87). As in the present study, factors specific to reading and writing, such as RAN and letter knowledge, showed higher predictive power for reading difficulties than the general factor of non-verbal intelligence.

Family history of reading problems was not included in our analysis, although prior studies found some contribution to a prediction model (36). However, a recent longitudinal study with a representative, epidemiological sample did not report acceptable AUC values for predicting reading problems by eliciting family risk factors (88), therefore diminishing the predictive value of family risk factors. This effect is expected especially for a German-speaking country because there are usually reservations about reporting family predispositions, and therefore no or unreliable information is provided.

School Use of the Screening Battery

For use in schools, screening tools should not only have high predictive power and reliability but also should have applicability with limited resources. In addition, screenings should be highly motivating for children. Children indicated that they experienced the assessment as a game and were able to stay with it well over a median duration of about 38 min. Not a single child had to stop for reasons of motivation or declining attention. The identification figure SCHWUPP provided continuous positive feedback after each completed subtest, and the frame story between the tasks could be used for relaxation. The new screening tool can be administered in about 15 min, making it shorter than any other screening tool available for the German language.

For the testers themselves, a high degree of objectivity was ensured because all instructions were played as audio files and important additional information (e.g., when naming the letters) was documented in the app. Due to the high degree of standardization, the training effort was low. Given that the one-to-one test setting remains necessary since some tasks require a paper target (such as letter knowledge or RAN) and the screening tool is for young children, a contact person is important in stressful situations. For teachers, rapid feedback of the results through automatic uploading of the results and further evaluations by the project team was important. Furthermore, for teachers, automatic scoring is regarded as a key feature of a feasible instrument.

Strengths of the New Screening Tool

The comprehensive sample is representative of Austria and the German language. A five-variable screening for surveillance in a community school setting showed good predictive power to detect slow readers at the end of first grade. For children, a motivating cover story presented interactively through tablets helps to maintain their motivation through a series of tasks. Advantages for screeners are short administration time, objectivity, and a quick computation of results. In order to make the whole screening tool available to primary schools without licensing costs, sub-tests had to be newly designed. There is now a screening tool that meets these requirements.

Limitations

Although a PPV of 39% is good compared to given screening tools, it does not cover all the children with slow reading at the end of first grade. Some data about environmental factors have been captured, but a big amount of variance is still to be detected: reading socialization *via* parents, school, friends, and at a macro level, society. With regard to the predictors, it must be noted that only language-related variables were included. Evidence on preschool visual processing has also recently been shown to be predictive of the reading process. Visual predictors were unfortunately not collected in the present study. Finally, continuous surveillance of reading is required because there might be different pathways to reading difficulties (many children with early difficulties do not develop later reading problems, and many children who do not fail the initial screening demonstrate reading difficulties later on).

Implications for Research and Practice

The first steps for a new screening on reading deficits have been implemented. Further validation of the newly constructed screening is needed, the next steps include a bigger normative sample and comparisons with screening tools already in use. Feasibility data for school usage must be gathered from children and teachers in order to enhance the screening and support a broader and well-accepted rollout.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Regional School Board for Upper Austria (Bildungsdirektion Upper Austria). Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

MS and GS: conceptualization, investigation, data curation, and project administration. MS, GS, and CW: methodology. CW: formal analysis. MS, GS, CW, and DH: writing original draft preparation and reviewing and editing. All authors contributed to the article and approved the submitted version.

FUNDING

All costs related to this research project were borne by the University of Education Upper Austria and Research Institute for Developmental Medicine (RID), Johannes Kepler University Linz. The tablets were provided by the project team with funds from the University of Education Upper Austria. Article processing charge was funded by the Johannes Kepler University Linz Open Access Publishing Fund.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fped.2022.863477/full#supplementary-material

REFERENCES

1. Ehri LC. Phases of development in learning to read words by sight. *J Res Read.* (1995) 18:116–25. doi: 10.1111/j.1467-9817.1995.tb00077.x
2. Stanovich K, West R. *Read Res Quart.* (1989) 24:747605. doi: 10.2307/747605
3. Esser G, Wyschkon A, Schmidt M. Was wird aus achtjährigen mit einer lese- und rechtschreibstörung: ergebnisse im alter von 25 jahren. *Zeitschrift Für Klinische Psychologie und Psychotherapie Forschung Praxis.* (2002) 31:235–42. doi: 10.1026//1616-3443.31.4.235
4. Gasteiger-Klicpera B, Klicpera C, Schabmann A. Der Zusammenhang zwischen Lese-, Rechtschreib- und Verhaltensschwierigkeiten. *Kindheit und Entwicklung.* (2006) 15:55–67. doi: 10.1026/0942-5403.15.1.55
5. Kohn J, Wyschkon A, Ballaschk K, Ihle W, Esser G. Verlauf von umschriebenen entwicklungsstörungen: eine 30-monats-follow-up-studie. *Lernen Lernstörungen.* (2013) 2:77–89. doi: 10.1024/2235-0977/a000032
6. Landerl K. Lese-/Rechtschreibstörung. In: Schneider S, Margraf J editors. *Lehrbuch der Verhaltenstherapie: Band 3: Störungen im Kindes- und Jugendalter*. Berlin: Springer (2009). p. 395–410. doi: 10.1007/978-3-540-79545-2_25
7. Moll K, Hutzler F, Wimmer H. Developmental dyslexia in a regular orthography: a single case study. *Neurocase.* (2005) 11:433–40. doi: 10.1080/13554790500263537
8. Moll K, Kunze S, Neuhoff N, Bruder J, Schulte-Körne G. Specific learning disorder: prevalence and gender differences. *PLoS One.* (2014) 9:e103537. doi: 10.1371/journal.pone.0103537
9. Schulte-Körne G. Diagnostik und Therapie der Lese-Rechtschreib-Störung. *Deutsches Ärzteblatt Int.* (2010) 107:718–27. doi: 10.3238/arztebl.2010.0718
10. Roth E, Warnke A. Diagnose und Therapie der Lese-Rechtschreibstörung. In: Koletzko B, Reinhardt D, Stöckler-Ipsiroglu S editors. *Pädiatrie upgrade 2002: Weiter- und Fortbildung*. Berlin: Springer (2002). p. 166–75. doi: 10.1007/978-3-642-56123-8_17
11. Perfetti CA. The representation problem in reading acquisition. In: Gough PB, Ehri LC, Treiman R editors. *Reading Acquisition*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc (1992). p. 145–74. doi: 10.4324/9781351236904-6
12. Perfetti CA, Hart L. The lexical bases of comprehension skill. In: Gorfien D editor. *On the Consequences of Meaning Selection*. Washington, DC: American Psychological Association (2001). doi: 10.1037/10459-004
13. McElvany N, Schneider C. Förderung von Lesekompetenz. In: Lenhard W, Schneider W editors. *Diagnostik und Förderung des Leseverständnisses*. Göttingen: Hogrefe (2009). p. 151–84.
14. Duncan GJ, Dowsett CJ, Claessens A, Magnuson K, Huston AC, Klebanov P, et al. School readiness and later achievement. *Dev Psychol.* (2007) 43:1428–46. doi: 10.1037/0012-1649.43.6.1428
15. Jordan N, Kaplan D, Locuniak M, Ramineni C. Predicting First-Grade Math Achievement from Developmental Number Sense Trajectories. *Learn Disabil Res Pract.* (2007) 22:36–46. doi: 10.1111/j.1540-5826.2007.00229.x
16. Moffitt T, Caspi A, Harrington H, Milne B. Males on the life-course-persistent and adolescence-limited antisocial pathways: follow-up at age 26 years. *Dev Psychopathol.* (2002) 14:179–207. doi: 10.1017/S0954579402001104

17. Nordström T, Jacobson C, Söderberg Juhlander P. Early word decoding ability as a longitudinal predictor of academic performance. *Eur J Psychol Educat.* (2015) 2015:1–17. doi: 10.1007/s10212-015-0258-5
18. McLaughlin MJ, Speirs KE, Shenassa ED. Reading disability and adult attained education and income: a 30-year longitudinal study of a population-based sample. *J Learn Disabil.* (2014) 47:374–86. doi: 10.1177/0022219412458323
19. Klicpera C, Gasteiger-Klicpera B, Schabmann A. *Lesen und Schreiben—Entwicklung und Schwierigkeiten. Die Wiener La?ngsschnittuntersuchungen u?ber die Entwicklung, den Verlauf und die Ursachen von Lese- und Schreibschwierigkeiten in der Pflichtschulzeit*. Bern: Huber (1993).
20. Moll K, Gangl M, Banfi C, Schulte-Körne G, Landerl K. Stability of Deficits in Reading Fluency and/or Spelling. *Sci Stud Read.* (2020) 24:241–51. doi: 10.1080/10888438.2019.1659277
21. Livingston EM, Siegel LS, Ribary U. Developmental dyslexia: emotional impact and consequences. *Austral J Learn Difficult.* (2018) 23:107–35. doi: 10.1080/19404158.2018.1479975
22. Fischbach A, Schuchardt K, Mähler C, Hasselhorn M. Zeigen Kinder mit schulischen Minderleistungen sozio-emotionale Auffälligkeiten? *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychol.* (2010) 42:201–10. doi: 10.1026/0049-8637/a000025
23. Mammarella IC, Ghisi M, Bomba M, Bottesi G, Caviola S, Broggi F, et al. Anxiety and depression in children with nonverbal learning disabilities, reading disabilities, or typical development. *J Learn Disabil.* (2016) 49:130–9. doi: 10.1177/0022219414529336
24. Galuschka K, Von Wirth E, Krick K, Schulte-Körne G. Effectiveness of treatment approaches for children and adolescents with reading disabilities: a meta-analysis of randomized controlled trials. *PLoS One.* (2014) 9:e89900. doi: 10.1371/journal.pone.0089900
25. Amelang M, Schmidt-Atzert L. *Psychologische Diagnostik und Intervention*. Berlin: Springer (2006).
26. Paleczek L, Seifert S, Schöfl M. Comparing digital to print assessment of receptive vocabulary with GraWo-KiGa in Austrian kindergarten. *Br J Educat Technol.* (2021) 52:13163. doi: 10.1111/bjet.13163
27. Thompson P, Hulme C, Nash H, Gooch D, Hayiou-Thomas E, Snowling M. Developmental dyslexia: predicting individual risk. *J Child Psychol Psychiat Allied Discipl.* (2015) 56:12412. doi: 10.1111/jcpp.12412
28. Grainger J, Dufau S, Ziegler JC. A Vision of Reading. *Trends Cognit Sci.* (2016) 20:171–9.
29. Vernet M, Bellocchi S, Leibnitz L, Chaix Y, Ducrot S. Predicting future poor readers from pre-reading visual skills: a longitudinal study. *Appl Neuropsychol Child.* (2021) 2021:1895790. doi: 10.1080/21622965.2021.1895790
30. Landerl K, Castles A, Parrila R. Cognitive precursors of reading: a cross-linguistic perspective. *Sci Stud Read.* (2021) 2021:1983820. doi: 10.1080/10888438.2021.1983820
31. Ptok M, Berendes K, Gottal S, Grabherr B, Schneeberg J, Wittler M. Phonologische Verarbeitung. *Monatsschrift Kinderheilkunde.* (2008) 156:860–6. doi: 10.1007/s00112-008-1770-3
32. Schnitzler CD, Springer L, Schrey-Dern D. *Phonologische Bewusstheit und Schriftspracherwerb*. Stuttgart: Thieme Verlag (2008). doi: 10.1055/b-002-21540
33. Pinto G, Bigozzi L, Tarchi C, Accorti Gamannossi B, Canneti L. Cross-lag analysis of longitudinal associations between primary school students' writing and reading skills. *Read Writ.* (2015) 28:9. doi: 10.1007/s11145-015-9569-9
34. Snowling M, Moll K, Hulme C. Language difficulties are a shared risk factor for both reading disorder and mathematics disorder. *J Exp Child Psychol.* (2021) 202:105009. doi: 10.1016/j.jecp.2020.105009
35. Price KM, Wigg KG, Misener VL, Clarke A, Yeung N, Blokland K, et al. Language Difficulties in School-Aged Children With Developmental Dyslexia. *J Learn Disabil.* (2021) 2021:2221942110062. doi: 10.1177/00222194211006207
36. Ozernov-Palchik O, Norton E, Sideridis G, Beach S, Wolf M, Gabrieli J, et al. Longitudinal stability of pre-reading skill profiles of kindergarten children: implications for early screening and theories of reading. *Dev Sci.* (2016) 20:12471. doi: 10.1111/desc.12471
37. Hurrelmann B, Becker S, Nickel-Bacon I. *Lesekindheiten. Familie und Lesesozialisation im historischen Wandel*. Weinheim: Juventa (2006).
38. Rosebrock C, Nix D. *Grundlagen der Lesedidaktik und der systematischen schulischen Leseförderung*. Hohengehren: Baltmannsweiler: Schneider-Verl (2008).
39. Rosebrock C, Nix D, Rieckmann C, Gold A. *Leseflüssigkeit fördern. Lautleseverfahren für die Primar- und Sekundarstufe*. Seelze: Kallmeyer (2011).
40. Ibáñez Alfonso JA, Hernández Cabrera JA, Duñabeitia JA, Monzó A, Macizo Soria P, Bajo Molina MT, et al. Socioeconomic status, culture, and reading comprehension in immigrant students. *Front Psychol.* (2021) 12:752273. doi: 10.3389/fpsyg.2021.752273
41. Lervåg A, Aukrust VG. Vocabulary knowledge is a critical determinant of the difference in reading comprehension growth between first and second language learners. *J Child Psychol Psychiat.* (2010) 51:612–20. doi: 10.1111/j.1469-7610.2009.02185.x
42. Leikin M, Schwartz M, Share DL. General and specific benefits of bi-literate bilingualism: a Russian–Hebrew study of beginning literacy. *Read Writ.* (2010) 23:269–92. doi: 10.1007/s11145-009-9210-x
43. Lipka O, Siegel LS. The development of reading comprehension skills in children learning English as a second language. *Read Writ.* (2012) 25:1873–98. doi: 10.1007/s11145-011-9309-8
44. Shany M, Geva E. Cognitive, language, and literacy development in socio-culturally vulnerable school children – the case of Ethiopian Israeli children. In: Leikin M, Schwartz M, Tobin Y editors. *Current Issues in Bilingualism*. Dordrecht: Springer Netherlands (2012). p. 77–117. doi: 10.1007/978-94-007-2327-6_5
45. Proctor CP, Silverman RD, Harring JR, Montecillo C. The role of vocabulary depth in predicting reading comprehension among English monolingual and Spanish–English bilingual children in elementary school. *Read Writ.* (2012) 25:1635–64. doi: 10.1007/s11145-011-9336-5
46. Geva E, Zadeh Z. Reading efficiency in native english-speaking and english-as-a-second-language children: the role of oral proficiency and underlying cognitive-linguistic processes. *Sci Stud Read.* (2006) 10:31–57. doi: 10.1207/s1532799xssr1001_3
47. Farnia F, Geva E. Growth and predictors of change in English language learners' reading comprehension. *J Res Read.* (2013) 36:389–421. doi: 10.1111/jrir.12003
48. Duzy D, Ehm J-H, Souvignier E, Schneider W, Gold A. Prädiktoren der Lesekompetenz bei Kindern mit Deutsch als Zweitsprache. *Zeitschrift Für Entwicklungspsychologie Und Pädagogische Psychol.* (2013) 45:173–90. doi: 10.1026/0049-8637/a000093
49. Statistik Austria. *Bildung in Zahlen. 2019/20. Schlüsselindikatoren und Analysen*. Wien: Statistik Austria (2021).
50. Endlich D, Lenhard W, Marx P, Scheider W. *LRS-Screening Laute, Reime, Sprache - Würzburger Screening zur Früherkennung von Lese-Rechtschreibschwierigkeiten*. Göttingen: Hogrefe (2019).
51. Schroeder S, Würzner KM, Heister J, Geyken A, Kliegl R. childLex–eine lexikalische datenbank zur schriftsprache für kinder im Deutschen. *Psychol Rundsch.* (2015) 66:155–65. doi: 10.1026/0033-3042/a000275
52. Muthén LK, Muthén BO. *Mplus user's guide. Eighth edition*. Los Angeles, CA: Muthén & Muthén (1998-2017).
53. Denckla MB, Rudel R. Rapid "automatized" naming of pictured objects, colors, letters and numbers by normal children. *Cortex J Devoted Study Nerv Syst Behav.* (1974) 10:186–202. doi: 10.1016/S0010-9452(74)80009-2
54. Landerl K, Ramus F, Moll K, Lyytinen H, Leppanen P, Lohvansuu K, et al. Predictors of developmental dyslexia in European orthographies with varying complexity. *J Child Psychol Psychiat Allied Discipl.* (2012) 54:12029. doi: 10.1111/jcpp.12029
55. Grob A, Meyer CS, Hagmann-von Arx P. *Intelligence and Development Scales (IDS). Intelligenz- und Entwicklungsskalen für Kinder von 5-10 Jahren*. Bern: Verlag Hans Huber (2009).
56. Marinis T, Armon-Lotem S. *Sentence Repetition*. Bristol: Multilingual Matters (2015). p. 95–124. Available online at: https://kops.uni-konstanz.de/handle/123456789/42352.
57. Ibrahim L, Hamann C, Öwerdieck D. *Identifying Specific Language Impairment (SLI) across Different Bilingual Populations: German Sentence Repetition Task (SRT)*. (2018). Available

online at: https://www.semanticscholar.org/paper/Identifying-Specific-Language-Impairment-(-SLI-)-%3A-Ibrahim-Hamann/48cf396157502fa5835f8bc09c180086a30945ac (accessed November 24, 2021).
58. Seifert S, Paleczek L, Schwab S, Gasteiger-Klicpera B. *GraWo: Grazer Wortschatztest*. Göttingen: Hogrefe (2017).
59. Wyschkon A, Esser G. *PITVA: Potsdamer Intelligenztest für das Vorschulalter: Manual*. Göttingen: Hogrefe (2019).
60. Lenhard W, Lenhard A, Schneider W. *Ein Leseverständnistest für Erst- bis Siebtklässler (ELFE II)*. Göttingen: Hogrefe (2017).
61. Swets JA. Measuring the accuracy of diagnostic systems. *Science.* (1988) 240:1285–93. doi: 10.1126/science.3287615
62. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformat.* (2011) 12:77. doi: 10.1186/1471-2105-12-77
63. McNeish D. Using lasso for predictor selection and to assuage overfitting: a method long overlooked in behavioral sciences. *Multivar Behav Res.* (2015) 50:471–84. doi: 10.1080/00273171.2015.1036965
64. Gareth J, Witten D, Hastie T, Tibshirani R. *An Introduction to Statistical Learning*. Heidelberg: Springer (2013).
65. Silverman RD, McNeish D, Speece DL, Ritchey KD. Early screening for decoding-and language-related reading difficulties in first and third grades. *Assess Effect Interv.* (2021) 46:99–109.
66. Friedman J, Hastie T, Tibshirani R, Narasimhan B, Tay K, Simon N, et al. *Package 'glmnet'. CRAN R Repository*. Vienna: R Core Team (2021).
67. Youngstrom EA. A primer on receiver operating characteristic analysis and diagnostic efficiency statistics for pediatric psychology: we are ready to ROC. *J Pediatr Psychol.* (2014) 39:204–21. doi: 10.1093/jpepsy/jst062
68. Venkatraman ES. A Permutation Test to Compare Receiver Operating Characteristic Curves. *Biometrics.* (2000) 56:1134–8. doi: 10.1111/j.0006-341x.2000.01134.x
69. López-Ratón M, Rodríguez-Álvarez M, Cadarso-Suárez C, Gude F. OptimalCutpoints: an R package for selecting optimal cutpoints in diagnostic tests. *J Statist Softw.* (2014) 61:i08. doi: 10.18637/jss.v061.i08
70. Greenslade K, Plante E, Vance R. The diagnostic accuracy and construct validity of the structured photographic expressive language test—preschool: second edition. *Lang Speech Hearing Serv Sch.* (2008) 40:150–60. doi: 10.1044/0161-1461(2008/07-0049
71. Pepe MS. *The Statistical Evaluation of Medical Tests for Classification and Prediction*. Oxford: University Pres (2003).
72. Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature: III. How to use an article about a diagnostic test: A. Are the results of the study valid? *J Am Med Assoc.* (1994) 271:389–91. doi: 10.1001/jama.1994.03510290071040
73. Seshan VE, Gönen M, Begg CB. Comparing ROC curves derived from regression models. *Statist Med.* (2013) 32:1483–93. doi: 10.1002/sim.5648
74. Catts H, Nielsen D, Bridges M, Liu Y-S. Early Identification of Reading Comprehension Difficulties. *J Learn Disabil.* (2014) 49:22219414556121. doi: 10.1177/0022219414556121
75. Kilgus S, Eklund K. Consideration of base rates within universal screening for behavioral and emotional risk: a novel procedural framework. *School Psychol Forum.* (2016) 10:120–30.
76. Compton DL, Fuchs D, Fuchs LS, Bryant JD. Selecting at-risk readers in first grade for early intervention: a two-year longitudinal study of decision rules and procedures. *J Educat Psychol.* (2006) 98:394–409. doi: 10.1037/0022-0663.98.2.394
77. Jenkins J, Hudson R, Johnson E. Screening for At-Risk Readers in a Response to Intervention Framework. *School Psychol Rev.* (2007) 36:12087919. doi: 10.1080/02796015.2007.12087919
78. Jansen H, Mannhaupt G, Marx H, Skowronek H. *Bielefelder Screening zur Früherkennung von Lese-Rechtschreibschwierigkeiten (BISC)*. Göttingen: Hogrefe (1999).
79. Marx P, Weber J. Vorschulische Vorhersage von Lese-undRechtschreibschwierigkeiten. *Zeitschrift Für Pädagogische Psychol.* (2006) 20:251–9. doi: 10.1024/1010-0652.20.4.251
80. Barth K, Gomm B. *Gruppentest zur Früherkennung von Lese- und Rechtschreibschwierigkeiten (PB-LRS)*. München: Reinhardt (2004).
81. Martschinke S, Kirschhock EM, Frank A. *Diagnose und Förderung im Schriftspracherwerb. Band 1. Der Rundgang durch Hörhausen. Erhebungsverfahren zur phonologischen Bewusstheit*. Hamburg: Auer-Verlag (2001).
82. Marx P, Lenhard W. Diagnostische Merkmale von Screeningverfahren. In: Hasselhorn M, Schneider W editors. *Fru?hprognose schulischer Kompetenzen*. Go?ttingen: Hogrefe (2010).
83. Rose J. *Identifying and teaching children and young people with dyslexia and literacy difficulties*. Nottingham: DCSF (2009).
84. Lohvansuu K, Torppa M, Ahonen T, Eklund K, Hämäläinen J, Leppanen P, et al. Unveiling the Mysteries of Dyslexia—Lessons Learned from the Prospective Jyväskylä Longitudinal Study of Dyslexia. *Brain Sci.* (2021) 11:427. doi: 10.3390/brainsci11040427
85. Huschka S, Georgiou G, Brandenburg J, Ehm J-H, Hasselhorn M. Examining the Contribution of RAN Components to Reading Fluency, Reading Comprehension, and Spelling in German. *Read Writ.* (2021) 34:2317–36. doi: 10.1007/s11145-021-10145-6
86. Ennemoser M, Marx P, Weber J, Schneider W. Spezifische vorläuferfertigkeiten der lesegeschwindigkeit, des leseverständnisses und des rechtschreibens: evidenz aus zwei längsschnittstudien vom kindergarten bis zur 4. klasse. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychol.* (2012) 44:53–67. doi: 10.1026/0049-8637/a000057
87. Landerl K, Wimmer H. Development of Word Reading Fluency and Spelling in a Consistent Orthography: an 8-Year Follow-Up. *J Educat Psychol.* (2008) 100:150–61. doi: 10.1037/0022-0663.100.1.150
88. Ferrer E, Shaywitz B, Holahan J, Shaywitz S. Family History Is Not Useful in Screening Children for Dyslexia. *J Pediatr Neuropsychol.* (2021) 8:15–21. doi: 10.1007/s40817-021-00110-0
89. Stanovich K. Matthew effects in reading: some consequences of individual differences in the acquisition of literacy. *Read Res Q.* (1986) 21:360–407. doi: 10.1598/RRQ.21.4.1

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.

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.

frontiers | Frontiers in Pediatrics

BRIEF RESEARCH REPORT
published: 02 June 2022
doi: 10.3389/fped.2022.878163

In Bilinguals' Hands: Identification of Bilingual, Preverbal Infants at Risk for Language Delay

Carina Lüke[1]*, *Ute Ritterfeld*[2] *and Ulf Liszkowski*[3]

[1] Special Education and Therapy in Language and Communication Disorders, Faculty of Human Sciences, University of Würzburg, Würzburg, Germany, [2] Department of Language and Communication, Faculty of Rehabilitation Sciences, TU Dortmund University, Dortmund, Germany, [3] Department of Developmental Psychology, Institute of Psychology, University of Hamburg, Hamburg, Germany

Studies with monolingual infants show that the gestural behavior of 1–2-year-olds is a strong predictor for later language competencies and, more specifically, that the absence of index-finger pointing at 12 months seems to be a valid indicator for risk of language delay (LD). In this study a lack of index-finger pointing at 12 months was utilized as diagnostic criterion to identity infants with a high risk for LD at 24 months in a sample of 42 infants growing up bilingually. Results confirm earlier findings from monolinguals showing that 12-month-olds who point with the extended index finger have an advanced language status at 24 months and are less likely language delayed than infants who only point with the whole hand and do not produce index-finger points at 12 months.

Keywords: gesture, pointing, language delay, language acquisition, bilingualism

OPEN ACCESS

Edited by:
Daniel Holzinger,
Hospitaller Brothers of Saint John of God Linz, Austria

Reviewed by:
Mila Vulchanova,
Norwegian University of Science and Technology, Norway
Natalia Meir,
Bar-Ilan University, Israel

Correspondence:
Carina Lüke
carina.lueke@uni-wuerzburg.de

Specialty section:
This article was submitted to Children and Health, a section of the journal Frontiers in Pediatrics

Received: *17 February 2022*
Accepted: *11 May 2022*
Published: *02 June 2022*

Citation:
Lüke C, Ritterfeld U and Liszkowski U (2022) In Bilinguals' Hands: Identification of Bilingual, Preverbal Infants at Risk for Language Delay. Front. Pediatr. 10:878163. doi: 10.3389/fped.2022.878163

INTRODUCTION

Gestures are one of the most important precursors of linguistic skills in young children. Toward the end of the first year of life infants use gestures to communicate intentionally with others. They initiate situations of joint attention and direct the attention of their caregivers to something they are interested in or want to communicate about. Colonnesi et al. (1) showed with their meta-analysis that especially the early use of pointing gestures is predictive for later language skills. Infants who produced between 10 and 20 months a high rate of pointing gestures had better linguistic skills between 12 and 54 months compared to infants who did not use as many pointing gestures at this early age.

Further research confirmed the predictive value of pointing gestures for later language skills (2–5). For example, Kuhn et al. (2) revealed with an epidemiological sample of over 1,000 infants studied within a prospective longitudinal study, that individual differences in communicative gestures at 15 months predict language skills at 2 and 3 years of age. Similarly, a recent meta-analysis by Kirk et al. (6) found a predictive function of early pointing for later language abilities, but with a considerably smaller effect size compared to Colonnesi et al.'s meta-analysis (1). Moreover, differences in the gestures between children with typical language development and children with language delay (LD) were identified (5, 7–10). About 20% of 2 year old children have been reported to have significant delays in their language acquisition (11). Especially children whose LD manifests in a developmental language disorder (DLD), which is the case for about 40% of the children (12), face negative and long-term effects on academic achievement or mental health, and consequently, in social participation (13).

Early pointing gestures in infants take on distinct hand shapes. Liszkowski and Tomasello (14) found in a sample of monolingual 12-month-olds that the canonical shape of index-finger pointing, as opposed to whole-hand pointing, revealed advanced prelinguistic communicative competencies. Infants who pointed with the extended index finger had a better understanding of communicative intentions of their communication partner, pointed more frequently and accompanied their pointing more often with vocalizations than infants who only pointed with the whole hand. Early pointing behavior, especially in the later emerging canonical form of index-finger pointing, is a milestone in social-cognitive development [for an overview see (15)] and is predicted by infants' prior social-cognitive ability to follow the gaze direction of a communication partner (16).

Subsequent studies explored the prediction of these distinct pointing shapes to language competence. Lüke et al. found in a monolingual sample that infants at 12 months, who only pointed with the whole-hand shape but not yet canonically with the extended index finger, were more likely to have a language delay (LD) at 24 months and to have lower language skills up to the age of 6 years compared to infants who produced index-finger points at the early age of 12 months (5, 10, 17). In two high risk populations—siblings of children with an Autism Spectrum Disorder diagnosis and preterm born infants with an extremely low gestational age—Sansavini et al. (9) established that those children who turned out to have a LD between 24 and 36 months produced fewer index-finger points at 18 months of age than typical developing (TD) children. The predictive value of index-finger pointing for later language skills in children with Autism Spectrum Disorder was also found by Ramos-Cabo et al. (18).

The absence of index-finger points at a certain age (i.e., 12 months), the age of onset of index-finger pointing (19), and the number of pointing gestures at a certain age might be valid criteria for an early and cost-effective identification of infants with a high risk of LD. The usage of these criteria within a screening would be of high clinical importance since children with LD develop lower language and literacy skills than TD children throughout school years and into adolescence [e.g., (20, 21)].

Many studies indicate that parent guided intervention programs are effective to support language acquisition in children with diagnosed LD [for a meta-analysis see (22)]. We argue that such secondary preventive intervention trainings could also be facilitative for even younger children with a high risk for LD within the second year of life. To provide such secondary preventive interventions, a screening tool for identifying infants with a high risk for LD is needed. Especially in bi- and multilingual children such a screening tool would be of tremendous value as the identification of a LD or a DLD is particularly challenging given the heterogeneity of the developmental pathways of bi- and multilingual children (23, 24). Building on the data on gesture development, we argue that preverbal communication such as early gestures could be an innovative, language independent indicator of LD, also valid in bilingual children.

However, in order to establish this argument, it needs to be proven that infant pointing is similar across monolingual and bilingual samples in terms of frequency, use, and shape. So far, Liszkowski et al. (25) demonstrated that prelinguistic gesturing can be considered a universal part of communication: 10–14-month-old infants from seven very different cultures and languages (from Papua New Guinea, Indonesia, Japan, Peru, Mexico (Tzeltal and Yucatec), and Canada) used in a standardized setting pointing gestures with the same canonical form, i.e., index-finger points, and with a similar frequency, and it correlated with their caregivers' pointing for them. While Tamis-LeMonda et al. (26) found differences in the number of gestures produced by Mexican American, Dominican American and African American mothers in interaction with their 14- and 24-months-old children, no differences were found between the number of gestures produced by the children from these three different cultural backgrounds. Germain et al. (27) verified this finding and found no differences in gestural development in French and English mono- and bilinguals living in Canada, assessed with different versions of the MacArthur-Bates Communicative Development Inventories CDI (28). Cameron-Faulkner et al. (29) compared gestural behavior in 10–12-months-olds and their mothers from three culturally distinct groups (Bengali, Chinese, and English) in the United Kingdom. They also found no differences in infants' gestural development, including index-finger pointing between these groups, and a positive correlation between gestures productions at 10–12 months and lexical skills at 18 months of age. Salomo and Liszkowski (30) on the other hand, analyzed index-finger pointing during different, natural daily activities in 8–16 months old Yucatec Mayan, Dutch, and Chinese infants and found differences in the frequency of index-finger pointers. While only 27% of the Mayan infants produced index-finger points, 72% of the Dutch and 88% of the Chinese infants did. In their recent meta-analysis, Kirk et al. (6) criticized that "none of the studies included samples of bilingual infants" (p. 2). Apart from the recently published study by Cameron-Faulkner et al. (29) no data on pointing behavior in bilingual infants and its predictive value for later language competencies is currently available.

If non-verbal communication ought to be used for screening in bilinguals the parallelism of gesture communication within both language situations must be assured. A study with a sample of five French-English bilingual boys between 2.0 and 3.6 years of age suggested that the frequency of gesture-speech productions was similar in the two languages (31). In a recent study, Limia et al. (32) found comparable results in 34 Spanish-English bilingual children within nearly the same age range (2.6–3.6 years). Moreover, the authors extended the meaningfulness of their insights from this bilingual sample by comparing them with 34 monolingual children growing up with either Spanish or English: The bilingual children referred uniquely gesturally to items in their surrounding with a similar proportion as the monolingual children. Also, their proportion of unique gestures was comparable in both languages even if one language was stronger. However, insights about the very early use of pointing gestures—around 12 months of age—and their predictive value for later language competencies are still missing for bilingual children [c.f. (6)].

In the current study we took a first step toward closing this gap and collected gestural behavior of 12-month-old infants growing up bilingually and assessed their language competencies a year later, at 24 months of age. We employed the standardized paradigm from Liszkowski et al. (25) to see whether the findings on pointing of monolingual infants pertain to bilingual infants; and to assess whether the relation between index-finger pointing and language delay found by Lüke et al. (5) in monolingual infants would also pertain to a bilingual sample. Specifically, we expect higher language skills at 24 months in children who produced index-finger points 1 year before and a lower rate of children with LD in this group of children compared to children who did not use index-finger points at 12 months. If infants growing up bilingually would show a similar developmental pattern as monolingual infants, this would enable practitioners in applied settings to use similar criteria to predictively assess mono- and bi-lingual infants' language development. It would also suggest that at this initial stage of development, language exerts less influence and instead rather "piggy-backs" on its gestural and social-cognitive infrastructure (33).

METHODS

Participants

Fifty-one 12-months-olds and their primary caregivers (90% mothers) were recruited for this longitudinal study. Data of nine children had to be excluded from analyses because they did not participate in both test sessions. The final sample included 42 infants (25 girls, 17 boys). At the first measurement, mean age of infants was 12 months and 9 days ($SD = 10$ days) and at the second measurement, mean age was 24 months and 17 days ($SD = 14$ days). All children were living together with their mothers and fathers and raised as bi- or multilingual speakers since birth. Using the graphical parent questionnaire *Input Contexts in Multilingualism* [ICOM; (34)] the language use of mothers and fathers in interaction with their infant was measured in detail. All children had exposure to German and at least to one other language on regular basis since birth. The majority of the children (76%) grew up with two languages, while 24% of the children grew up with three or more languages. Besides German, twenty different languages were spoken by the children's families with Turkish (29%), Spanish (14%), English (9.5%), and Chinese (7%) as the most prevalent languages. Albanian, Amharic, Arabic, Bulgarian, Dutch, French, Hindi, Italian, Korean, Macedonian, Pashto, Persian, Romanian, Russian, Urdu, and Vietnamese were spoken in one or two families each.

According to the pediatricians of the infants and a standardized test of global development (*Entwicklungstest für Kinder von 6 Monaten bis 6 Jahren -ET 6-6*) (Developmental test for children between 6 months and 6 years) (35) all children were typically developing. They grew up in families with a rather high level of education ($M_{\text{maternal years of education}} = 16$, $SD_{\text{maternal years of education}} = 2$; $M_{\text{paternal years of education}} = 16$, $SD_{\text{paternal years of education}} = 3$) and a household disposable income per month ($Md =$ € 1.944, $IQR =$ € 1.865) comparable to the German median in the same year (36).

Procedure

Eliciting and Coding of Pointing Gestures

Infants and their primary caregiver took part in two sessions of data collection, one at the age of 12 months, one at 24 months. At 12 months the pointing behavior of the children was captured, using the same setting as in 5 (5), the so called "decorated room" (14). In this room 19 interesting objects and pictures are placed to elicit natural, multimodal interaction between caregivers and their infants. Caregivers were asked to carry their child for 6 min and to look at the items without touching them, if possible. Caregivers were not informed that gestures were analyzed. Four cameras recorded the scene from the four corners of the room.

A research assistant unaware of the research question coded the videos in both real-time and frame-by-frame analyses for the occurrence of pointing gestures using the annotation tool ELAN (37). Pointing was coded when the infant extended the hand and the arm toward an object or a picture without grabbing or touching it. The gestures were coded either as index-finger points—when the index finger was clearly extended relative to all other fingers—or as hand points—when the index finger was not clearly extended relative to the other fingers. To assess inter-rater reliability a second research assistant coded independently a random 10% of the sample data. Inter-rater reliability for infants' pointing was very good (Krippendorff's $\alpha = 0.967$).

Based on the reliable coding of the hand shape, infants were classified as *index-finger pointers* if they pointed at least once with the index finger. Infants who only pointed with their whole hand were classified as *hand pointers* (14). Validity for group assignment was assured by an additional procedure: In all cases where infants only pointed with the whole hand or pointed just once or twice with the index finger (20 cases), the video was coded by the second independent coder. Comparisons revealed that the two codings did overlap in all but one case. In this case of non-agreement, the vote of an additional independent third coder decided upon the final group assignment. All coders were uninformed about the research question and the language status of the infants.

Language Testing and Diagnosis of LD

Verbal skills of the children were assessed at 24 months using standardized German language measures: *Sprachentwicklungstest für zweijährige Kinder (SETK-2)* [test of language acquisition for 2-year-old children] (38) and the *Fragebogen zur frühkindlichen Sprachentwicklung (FRAKIS)* [German equivalent of the standardized parent questionnaire CDI (28)] (39). The SETK-2 (38) consists of four subtests assessing comprehension and production of words and sentences. Results are presented in standard T-scores. Since the children in this study were raised as bilingual speakers some adaptions within the language testing were made. For the three subtests *word comprehension*, *sentence comprehension*, and *word production* the child got the instruction in German as well as in their other language. The test items were first presented by the experimenter in German. If the child did not react or gave a wrong answer, the caregiver provided the item again in the other language spoken within the family. This procedure was feasible because the presented items were basic words or sentences which could be easily translated by the caregivers. Beyond that, the majority of the presented words have a comparable age of acquisition in many different languages [c.f. (40)]. The procedure was explained to the caregivers before testing. The procedure of the fourth subtest *sentence production* was too complex to be translated and presented by the caregivers so that this subtest was only administered in German with those children who were able to comply ($n = 24$).

From the German parental questionnaire FRAKIS (39) the vocabulary checklist of 600 words was used. Parents were asked to indicate which of the presented 600 words were spoken by their child in either of the child's active languages. The parents marked the spoken words in the different languages by using different colored pencils, one for each language. For analyses the conceptual vocabulary as a composite of all items spoken in at least one language (41) was calculated for each child. This procedure was used since there were not for all languages in our sample adaptions of the CDI (28) available, all parents in our sample had so much German proficiency that they could answer the questionnaire easily in German and to avoid an influence by different word lists.

Paralleling the procedure with a monolingual sample (5), we defined a 2-year-old child as language delayed if s/he scored in at least one of the three standardized language subtests (*word comprehension, sentence comprehension, word production*) of the SETK-2 (38) or in the vocabulary checklist 1½ standard deviation below the mean (i.e., T-score of ≤ 35) and in at least one additional subtest or the vocabulary checklist 1 standard deviation below the mean (i.e., T-score of < 40). By this definition 24% of the children in the sample were classified as language delayed at 24 months. The gestural and language skills of children with and without a LD are presented in **Table 1**.

Data Analysis

Apart from the variables *word comprehension* and *word production* none of the language measures were normally distributed as indicated by the Shapiro-Wilk-test. Therefore, we used the non-parametric Mann-Whitney-U test for group comparisons and report Median (*Md*) and interquartile range (*IQR*) for distributions. For better interpretation of the results we report Cohen's *d* as effect size.

For classification of children as either LD or TD at 24 months using their pointing behavior at 12 months as criterium, we calculated the relative improvement over chance [RIOC; (42)]. RIOC incorporates the base rate and the selection rate of a developmental disorder and a screening. The rationale behind is that screenings for developmental disorders with a low base rate which randomly identifies all children as TD would reach high scores in specificity and accuracy and would thus be invalid. As the RIOC responds much more sensible in this case it is considered a superior indicator for the prognostic validity of a screening tool (42).

RESULTS

At 12 months, all 42 infants produced pointing gestures. Twenty-seven infants pointed at least once with the index finger and were therefore classified as index-finger pointers (64%), while the remaining 15 solely used whole-hand points and no index-finger points and were consequently classified as hand pointers (36%). The vast majority of the index-finger pointers (89%) pointed more than once with the index finger within a range from 2 to 45 index-finger points. The index-finger pointers produced more pointing gestures in total ($Md = 27.0$, $IQR = 20.0$) compared to the hand pointers ($Md = 12.0$, $IQR = 13.0$, $U = 76.0$, $p = 0.001$, $d = 1.193$). Moreover, index-finger pointers combined their pointing gestures more often with a vocalization (proportionately to the number of pointing gestures; $Md = 0.47$, $IQR = 0.41$) than hand pointers ($Md = 0.24$, $IQR = 0.25$, $U = 107.5$, $p = 0.013$, $d = 0.834$). The number of index-finger points produced by the caregivers of index-finger pointers ($Md = 14.0$, $IQR = 18.0$) and hand pointers ($Md = 18.0$, $IQR = 25.0$) did not differ ($U = 201.0$, $p = 0.969$, $d = 0.012$).

Comparing the language development of infants who were able to point with the extended index finger at 12 months to those infants who did not use the index finger for pointing at this young age resulted in differences between the two groups. Index-finger pointers showed an advanced language status at 24 months compared to hand pointers; they had a greater conceptional vocabulary and a better sentence comprehension (**Table 2**). Moreover, most of the index-finger pointers were typically developed at 24 months (89%) while nearly half of the hand pointers (47%) were language delayed. **Table 3** summarizes the results of classifying index-finger pointers and hand pointers as either language delayed or typically developed. In other words, using the criterion of index-finger pointing at 12 months for identifying children with LD at 24 months reflects a sensitivity of 70%, a specificity of 75%, an accuracy of 74%, and a RIOC of 53% (see **Table 4** for all commonly used values of quality criteria for screening tools).

When directly comparing the results of the bilingual sample to those of a monolingual sample (5), there were no statistical significant differences in the number of index-finger pointers [$\chi^2_{(1)} = 1.72$, $p = 0.189$, *Cramer's* $V = 0.131$] or the number of index-finger points produced ($Md_{mono} = 6.0$, $IQR_{mono} = 16.0$, $Md_{bi} = 4.0$, $IQR_{bi} = 15.0$, $U = 1{,}071.0$, $p = 0.241$, $d = 0.232$), although the effect sizes indicate small effects. There were no differences found in any gestural or language measures between children growing up with two or more languages.

DISCUSSION

The current study reproduced the pattern of earlier findings from a monolingual sample (5) on the relation between early pointing and language development and extended it to bilingual infants. Current main findings were that the absence of index-finger pointing at 12 months in a bilingual sample, like in a monolingual sample, indicates a higher risk for being language delayed at 24 months. These findings are in line with the current state of research showing that index-finger pointing reflects advances in children's communication, which leads to an earlier achievement of linguistic competencies (1, 4, 8). Bilingual 12-month-old infants in the current study who used index-finger points to communicate with their caregivers demonstrated better

TABLE 1 | Descriptive statistics of gestural and linguistic skills in TD children and children with LD at 24 months.

	TD ($n = 32$)		LD ($n = 10$)	
	M (SD)	*Md (IQR)*	*M (SD)*	*Md (IQR)*
Hand points	15.31 (13.72)	11.50 (14.75)	17.20 (12.07)	14.50 (15.75)
Index-finger points	10.94 (13.46)	5.50 (18.50)	1.60 (2.63)	0 (4.50)
Vocabulary size in German[a]	154.81 (124.33)	105.0 (206.0)	25.10 (17.39)	26.0 (30.0)
Vocabulary size in the other language[a]	137.42 (126.98)	107.0 (195.0)	22.00 (24.06)	13.5 (42.0)
Conceptual vocabulary size[a]	233.90 (115.07)	225.0 (157.0)	36.50 (23.25)	29.5 (50.0)
Word comprehension[b]	50.66 (7.89)	51.0 (10.0)	40.70 (9.14)	39.5 (16.0)
Sentence comprehension[b]	49.29 (9.56)	54.0 (13.0)	35.40 (7.59)	35.0 (15.0)
Word production[b]	43.61 (7.85)	43.0 (10.0)	31.78 (3.73)	33.0 (7.0)
Sentence production[b]	44.00 (7.00)	42.0 (8.0)	34.40 (4.62)	35.0 (9.0)

[a] *Number of spoken words, measured with the parent questionnaire FRAKIS (39).*
[b] *Standard T-scores, measured with the language test SETK-2 (38).*

TABLE 2 | Comparison of language skills at 24 months between index-finger pointers and hand pointers.

	Index-finger pointers		Hand pointers				
	Md	*IQR*	*Md*	*IQR*	*U*	*p*	*d*
Conceptional vocabulary	225.0	199.0	105.0	132.0	102.0	0.12	0.891
Word comprehension	48.0	13.0	48.0	13.0	169.0	0.371	0.274
Sentence comprehension	54.0	14.0	41.0	13.0	114.0	0.026	0.768
Word production	39.5	12.0	37.0	15.0	136.5	0.196	0.555
Sentence production*	41.5	8.0	38.5	11.0	25.0	0.052	0.859

**Since this subtest was only done with those children who were able to do it solely in German ($n = 24$) the group of index-finger pointers consisted of 18 children and the hand-pointers of 6.*

TABLE 3 | Classification of children as being language delayed at 24 months based on their ability to produce index-finger points at 12 months.

		LD at 24 months		
		Yes	No	Total
Index finger pointing at 12 months	No	7 (70%)	8 (25%)	15 (36%)
	Yes	3 (30%)	24 (75%)	27 (64%)
Total		10 (100%)	32 (100%)	42 (100%)

$\chi^2_{(1)} = 6.72$, $p = 0.010$, Cramer's $V = 0.40$.

TABLE 4 | Quality criteria of index-finger pointing at 12 months as screening tool for LD at 24 months.

Criterion	Value
Sensitivity	0.70
Specificity	0.75
Positive predictive value	0.47
Negative predictive value	0.89
Accuracy	0.74
Selection rate	0.58
Relative improvement over chance (RIOC)	0.53

language skills at 24 months than infants who did not use index-finger points at 12 months. Based on the finding that just very few index-finger pointers were identified as having a LD at 24 months, while nearly half of the hand pointers had a LD at 24 months, it seems appropriate to consider index-finger pointing as a sign of TD and its absence as a risk factor for language acquisition.

In line with other studies reporting no differences in gestural development based on different cultural backgrounds or bilingualism (25–27, 29), we found neither differences in the number of index-finger pointers and hand pointers nor in the number of index-finger points produced between our presented bilingual sample and a monolingual sample using the identical procedures (5). Nevertheless, the screening criteria, specificity and RIOC of index-finger pointing, seem less robust in this bilingual sample compared to the monolingual sample (5). In the current bilingual sample 36% of the infants were classified as hand pointers while only 20% of the monolingual infants, investigated by Lüke et al. (5), did not produce index-finger points at 12 months. Possibly, this could be the result of the time point of data collection in some infants. In the monolingual sample eight infants could only be tested comparably late (16–37 days after their first birthday). Two of these slightly older infants did not point with the index finger and were later identified as being language delayed while the other six infants produced index-finger points and were not language delayed at 24 months. In the bilingual sample, a reverse pattern had occurred: The three bilingual infants who had been tested comparably late (23–30 days after their first birthday) did produce index-finger points and were identified as having a LD at 24 months. It remains unknown whether the results would have been identical if these infants had been observed 1 or 2 weeks earlier. These observations demonstrate the highly dynamic developmental pathways at this young age which may have affected the findings. Since the first productions of index-finger points occur between 10 and 12 months (43), further research with infants between 9 and 12 months of age is needed, so that the onset of index-finger pointing as predictor of later language skills can be analyzed and might be more robust as diagnostic tool compared to the absence of index-finger pointing at 12 months. These slightly differences reveal very clearly that the development of early communicative gestures is occurring at a rapid pace, resulting in a sudden change in categorization of a child as index-finger vs. hand pointer from 1 day to the next.

Beyond that, the sample sizes of both samples, the monolingual as well as the bilingual, are with 59 or 42 too small and not appropriate to prove any ability or tool as a prognostic valid screening instrument. The presented values of quality criteria for screening tools in this study as well as in the study with the monolingual sample (5) can only serve as orientation. Nevertheless, these orientating values with, for example good accuracies between 74 and 85%, the predictive value of pointing gestures found in many studies [for meta-analysis see (1, 6)], and the language and cultural universal occurrence of index-finger pointing (25–27, 29) are encouraging to further investigate the use of index-finger pointing as an early indicator of LD in a population-based study with children between 10 and 12 months during pediatric service. This would be especially important for children growing up with two or more languages since the identification of bilingual children with LD or even DLD is particularly challenging (23, 24).

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available on request to the corresponding author.

ETHICS STATEMENT

The study was reviewed and approved by the Internal Review Board of Bielefeld University (EUB 2015-079). Written informed consent to participate in this study was provided by the participants' legal guardians.

AUTHOR CONTRIBUTIONS

CL was responsible for data collection, data analysis, and drafted the first manuscript. All authors designed the study, developed the coding system, interpreted the data, and revised the manuscript critically.

FUNDING

This project was funded by the Deutsche Forschungsgemeinschaft, LI 1989/2-2 to UL, LU 2116/1-2 to CL, and RI 898/6-2 to UR. This publication was supported by the Open Access Publication Fund of the University of Wuerzburg.

ACKNOWLEDGMENTS

We are grateful to all families for participating in the study and our partners in cooperation, Katharina J. Rohlfing and Angela Grimminger.

REFERENCES

1. Colonnesi C, Stams GJ, Koster I, Noom MJ. The relation between pointing and language development: a meta-analysis. *Dev Rev.* (2010) 30:352–66. doi: 10.1016/j.dr.2010.10.001
2. Kuhn LJ, Willoughby MT, Wilbourn MP, Vernon-Feagans L, Blair CB. Early communicative gestures prospectively predict language development and executive function in early childhood. *Child Dev.* (2014) 85:1898–914. doi: 10.1111/cdev.12249
3. Beuker KT, Rommelse NNJ, Donders R, Buitelaar JK. Development of early communication skills in the first two years of life. *Infant Behav Dev.* (2013) 36:71–83. doi: 10.1016/j.infbeh.2012.11.001
4. Murillo E, Belinchón M. Gestural-vocal coordination: longitudinal changes and predictive value on early lexical development. *Gesture.* (2012) 12:16–39. doi: 10.1075/gest.12.1.02mur
5. Lüke C, Grimminger A, Rohlfing KJ, Liszkowski U, Ritterfeld U. In infants' hands: identification of preverbal infants at risk for primary language delay. *Child Dev.* (2017) 88:484–92. doi: 10.1111/cdev.12610
6. Kirk E, Donnelly S, Furman R, Warmington M, Glanville J, Eggleston A. The relationship between infant pointing and language development: a meta-analytic review. *Dev Rev.* (2022) 64:101023. doi: 10.1016/j.dr.2022.101023
7. L010 C, Ritterfeld U, Grimminger A, Liszkowski U, Rohlfing KJ. Development of pointing gestures in children with typical and delayed language acquisition. *J Speech Lang Hear Res.* (2017) 60:3185–185 doi: 10.1044/2017_JSLHR-L-16-0129
8. Bello A, Onofrio D, Remi L, Caselli C. Prediction and persistence of late talking: a study of Italian toddlers at 29 and 34 months. *Res Dev Disabil.* (2018) 75:40–8. doi: 10.1016/j.ridd.2018.02.006
9. Sansavini A, Guarini A, Zuccarini M, Lee JZ, Faldella G, Iverson JM. Low rates of pointing in 18-month-olds at risk for autism spectrum disorder and extremely preterm infants: a common index of language delay? *Front Psychol.* (2019) 10:2131. doi: 10.3389/fpsyg.2019.02131
10. Lüke C, Leinweber J, Ritterfeld U. Walking, pointing, talking alThe predictive value of of early walking and pointing behavior for later language skills. *J Child Lang.* (2019) 46:1228–37. doi: 10.1017/S0305000919 000394
11. Reilly S, Wake M, Bavin EL, Prior M, Williams J, Bretherton L, et al. Predicting language at 2 years of age: a prospective community study. *Pediatrics.* (2007) 120:e1441–1449. doi: 10.1542/peds.2007-0045
12. Dale PS, Price TS, Bishop DV, Plomin R. Outcomes of early language delay: I. Predicting persistent and transient language difficulties at 3 and 4 years. *J Speech Lang Hear Res.* (2003) 46:544–60. doi: 10.1044/1092-4388(2003/044)

13. Law J, Rush R, Schoon I, Parsons S. Modeling developmental language difficulties from school entry into adulthood: literacy, mental health, and employment outcomes. *J Speech Lang Hear Res.* (2009) 52:1401–16. doi: 10.1044/1092-4388(2009/08-0142)
14. Liszkowski U, Tomasello M. Individual differences in social, cognitive, and morphological aspects of infant pointing. *Cogn Dev.* (2011) 26:16–29. doi: 10.1016/j.cogdev.2010.10.001
15. Liszkowski U. Three lines in the emergence of prelinguistic communicsation and social cognition. *J Cognitive Educ Psychol.* (2011) 10:32–43. doi: 10.1891/1945-8959.10.1.32
16. Matthews D, Behne T, Lieven EV, Tomasello M. Origins of the human pointing gesture: a training study. *Dev Sci.* (2012) 15:817–29. doi: 10.1111/j.1467-7687.2012.01181.x
17. Lüke C, Ritterfeld U, Grimminger A, Rohlfing KJ, Liszkowski U. Integrated communication system: gesture and language acquisition in typically developing children and children with LD and DLD. *Front Psychology.* (2020) 11:118. doi: 10.3389/fpsyg.2020.00118
18. Ramos-Cabo S, Acha J, Vulchanov V, Vulchanova M. You may point, but do not touch: impact of gesture-types and cognition on language in typical and atypical development. *Int J Lang Commun Disord.* (2022) 57:324–39. doi: 10.1111/1460-6984.12697
19. Liszkowski U, Rüther J. Ontogenetic origins of infant pointing. In: Gontier N, Lock A, Sinha C, editors. *Oxford Handbook of Human Symbolic Evolution.* (2021). doi: 10.1093/oxfordhb/9780198813781.013.31
20. Paul R. Patterns of development in late talkers: preschool years. *Commun Disord Q.* (1993) 15:7–14. doi: 10.1177/152574019301500103
21. Rescorla L. Age 17 language and reading outcomes in late-talking toddlers: support for a dimensional perspective on language delay. *J Speech Lang Hear Res.* (2009) 52:16–30. doi: 10.1044/1092-4388(2008/07-0171)
22. Roberts MY, Kaiser AP. The effectiveness of parent-implemented language interventions: a meta-analysis. *Am J Speech Lang Pathol.* (2011) 20:180–99. doi: 10.1044/1058-0360(2011/10-0055)
23. Dollaghan CA, Horner EA. Bilingual language assessment: a meta-analysis of diagnostic accuracy. *J Speech Lang Hear Res.* (2011) 54:1077–88. doi: 10.1044/1092-4388(2010/10-0093)
24. Armon-Lotem S, Jong Jd, Meir N, editors. *Assessing Multilingual Children: Disentangling Bilingualism from Language Impairment*. Bristol: Multilingual Matters (2015). xi, 364.
25. Liszkowski U, Brown P, Callaghan T, Takada A, de Vos C. A prelinguistic gestural universal of human communication. *Cogn Sci.* (2012) 36:698–713. doi: 10.1111/j.1551-6709.2011.01228.x
26. Tamis-LeMonda CS, Song L, Leavell Smith A, Kahana-Kalman R, Yoshikawa H. Ethnic differences in mother-infant language and gestural communications are associated with specific skills in infants. *Dev Sci.* (2012) 15:384–97. doi: 10.1111/j.1467-7687.2012.01136.x
27. Germain N, Gonzalez-Barrero AM, Byers-Heinlein K. Gesture development in infancy: effects of gender but not bilingualism. *Infancy.* (2022). doi: 10.1111/infa.12469
28. Fenson L, Dale PS, Reznick JS, Thal D, Bates E, Hartung JP, et al. *MacArthur-Bates Communicative Development Inventories: User's guide and technical manual.* Baltimore, Md: Paul H. Brookes Pub. Co. (2007). xviii, 188. doi: 10.1037/t11538-000
29. Cameron-Faulkner T, Malik N, Steele C, Coretta S, Serratrice L, Lieven E, et al. cross-cultural analysis of early prelinguistic gesture development and its relationship to language development. *Child Dev.* (2021) 92:273–90. doi: 10.1111/cdev.13406
30. Salomo D, Liszkowski U. Sociocultural settings influence the emergence of prelinguistic deictic gestures. *Child Dev.* (2013) 84:1296–307. doi: 10.1111/cdev.12026
31. Nicoladis E, Mayberry RI, Genesee F. Gesture and early bilingual development. *Dev Psychol.* (1999) 35:514–26. doi: 10.1037/0012-1649.35.2.514
32. Limia V, Özçalişkan S, Hoff E. Do parents provide a helping hand to vocabulary development in bilingual children? *J Child Lang.* (2019) 46:501–21. doi: 10.1017/S0305000918000594
33. Tomasello M. *Origins of Human Communication*. Cambridge, Mass: MIT Press (2008). 1 online resource, xiii, 393.
34. Ritterfeld U, Lüke C, Schnöring A. *ICOM – Input Contexts in Multilingualism* (2015). Available online at: http://www.sk.tu-dortmund.de/cms/de/materialien/ICOM.pdf (accessed April 12, 2022).
35. Petermann F, Stein IA, Macha T. *ET 6-6. Entwicklungstest sechs Monate bis sechs Jahre.* Frankfurt am Main: Pearson (2008).
36. Kott K. *Einkommensverteilung* (2021). Available online at: https://www.bpb.de/nachschlagen/datenreport-2021/private-haushalte-einkommen-und-konsum/329921/einkommensverteilung (accessed December 22, 2021).
37. Max Planck Institute for Psycholinguistics. *ELAN*. Nijmegen: Max Planck Institute for Psycholinguistics (2020).
38. Grimm H. *SETK-2. Sprachentwicklungstest fpr zweijhentwi Kinder: 2 üinder:ntwick und neunormierte Auflage.* Guflageie: Hogrefe (2016).
39. Szagun G, Stumper B, Schramm SA. *FRAKIS. Fragebogen zur frrgebogenteel Sprachentwicklung.* Frankfurt am Main: Pearson (2009).
40. Łuniewska M, Haman E, Armon-Lotem S, Etenkowski B, Southwood F, Anđelković D, et al. Ratings of age of acquisition of 299 words across 25 languages: is there a cross-linguistic order of words? *Behav Res Methods.* (2016) 48:1154–77. doi: 10.3758/s13428-015-0636-6
41. Pearson BZ, Fernández SC, Oller DK. Lexical development in bilingual infants and toddlers: comparison to monolingual norms. *Lang Learn.* (1993) 43:93–120. doi: 10.1111/j.1467-1770.1993.tb00174.x
42. Farrington DP, Loeber R. Relative improvement over chance (RIOC) and phi as measures of predictive efficiency and strength of association in 2×2 tables. *J Quant Criminol.* (1989) 5:201–13. doi: 10.1007/BF01062737
43. Lock A, Young A, Service V, Chandler P. Some observations on the origins of the pointing gesture. In: Volterra V, Erting C, editors. *From Gesture to Language in Hearing and Deaf Children*. Berlin: Springer (1990). p. 42–55.

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.

frontiers | Frontiers in Public Health

ORIGINAL RESEARCH
published: 22 June 2022
doi: 10.3389/fpubh.2022.866598

Check for updates

Development and Validation of a Language Screening for Implementation in Pre-School Settings

Daniel Holzinger[1,2,3], Christoph Weber[2,4] and Bettina Diendorfer[1]*

[1] *Institute of Neurology of Senses and Language, Hospital of St. John of God, Linz, Austria,* [2] *Research Institute of Developmental Medicine, Johannes Kepler University Linz, Linz, Austria,* [3] *Institute of Linguistics, University of Graz, Graz, Austria,* [4] *Department for Inclusive Education, University of Education Upper Austria, Linz, Austria*

Background: To prevent or mitigate long-lasting learning problems and emotional, behavioral, and social-adaption difficulties associated with language disorders, age-appropriate German language competence at school entry level is essential. Therefore, universal screening of children in their penultimate year of pre-school has been established in Upper Austria. So far, the screenings administered by speech and language pathologists to identify risk of language disorder (LD) were not based on standardized materials.

Objective: To develop a screening instrument to identify increased risk of LD and to evaluate its validity and feasibility within the constraints of regular universal pre-school language screening.

Design: A two-component screening instrument including direct assessment of expressive and receptive grammar was used in a sample of 374 children with German as their dominant language attending a public pre-school in their penultimate year (age 4-5 ½ years) in the state of Upper Austria. Assessment by use of standardized German language tests including a variety of linguistic domains was considered reference standard for diagnosing LD. Feasibility was assessed by a self-developed questionnaire completed by the administrators of the screening.

Results: The combination of the expressive and receptive grammar scales demonstrated excellent accuracy (area under the curve score 0.928). A cut-off of 18 resulted in a failing rate of 21.8% and showed good sensitivity (84.2%) and specificity (85.3%). Acceptance by children and testers, time-economy and sustainability of the screening were mostly rated as high.

Keywords: language screening, language disorder, LOGiK-S, validity, feasibility, pre-school

Edited by:
Mary A. Khetani, University of Illinois at Chicago, United States

Reviewed by:
Tanya Serry, La Trobe University, Australia
Sarah Verdon, Charles Sturt University, Australia

****Correspondence:***
Christoph Weber
christoph.weber@ph-ooe-at

Specialty section:
This article was submitted to Children and Health, a section of the journal Frontiers in Public Health

Received: *31 January 2022*
Accepted: *25 May 2022*
Published: *22 June 2022*

Citation:
Holzinger D, Weber C and Diendorfer B (2022) Development and Validation of a Language Screening for Implementation in Pre-School Settings. Front. Public Health 10:866598. doi: 10.3389/fpubh.2022.866598

INTRODUCTION

The international CATALISE consortium (1) recently addressed the issue of terminology and definition of problems with language development, by defining diagnostic criteria for the newly termed Developmental Language Disorder (DLD) by the CATALISE Consensus. The new term DLD refers to a language disorder (LD) that emerges during development and is not associated with known biomedical conditions. DLD is a heterogeneous condition, which can affect language production and/or comprehension and different linguistic domains (lexical, morpho-syntactic, pragmatic). The new definition of DLD does not preclude the co-occurrence with other neurodevelopmental conditions, the presence of environmental risk factors or require a mismatch between verbal and non-verbal cognition. In addition, the consensus statement agreed on the serious nature of language problems with a significant impact on everyday social interactions or educational progress and poor prognosis of LD. What the consensus statement did not define is the extent of language difficulties in mode (receptive and/or expressive) and linguistic dimension (phonology, vocabulary, morphology, syntax, pragmatic s). Therefore, DLD remains a clinical diagnosis, where professionals need to be able to recognize language deficits associated with functional impairment and the potential of these conditions to become chronic with an increased risk of learning and mental health problems.

With language abilities at least 1.5 SD under those of peers in at least two of the five linguistic domains, Norbury et al. (2) found a prevalence of LD of any origin of about 10% (7.58% specific with unknown origin and 2.34% non-specific with medical diagnosis), which makes LD one of the most common developmental problems in childhood. Similarly, earlier studies that assume language abilities around 1.25 SD below the norm in two linguistic domains, expect a prevalence rate of 5-8% of specific LD in children speaking English (3, 4), English or French (5) or German (6).

Children with LD are at high risk of difficulties in academic and vocational qualification (7, 8), mental health problems and social adaptation difficulties (9–11). Early identification of LD may help children to access specialized educational (9), therapeutic (12) and parent-implemented (13) intervention to support them to improve their language skills by school entry and to reduce the risk of neuropsychological sequelae. As a consequence, a system for a universal language check-up has been established in the State of Upper Austria since the mid 90's administered by speech and language pathologists. In Upper Austria, a federal state with a population of 1.45 million inhabitants, all children (about 14.000/year at the time of data collection) are assessed in their penultimate year of pre-school for speech and language development every year. Up to this point, speech and language pathologists are faced with the challenge of accurately identifying the children with the highest risk of persisting language difficulties and need of language intervention. The challenges concern the lack of a generally accepted definition of what constitutes a LD and the lack of a standardized and feasible procedure for language screening. Another challenge concerns the high variability of language development during the early years with a high proportion of children with initially poor language catching up before school entry (14–16) and others manifesting deterioration in the trajectory of language development over time. Whereas some studies have demonstrated relatively stable trajectories of language development from the age of 5-6 years (2, 17), more recent population cohort studies have shown that the degree of variability in child language pathways even after the age of 4 or 5 years might have been underestimated suggesting the necessity of continuous surveillance of language development and environmental risk factors (18, 19).

In 2006, Nelson et al. concluded their review for the US Preventive Services Task Force advising against universal language screenings because of many methodological problems they had identified in language intervention and outcome studies (20) provided an update to the (21) systematic review reporting sufficient accuracy of some screening tools for the identification of children with LD but highlighting a lack of studies demonstrating their feasibility in primary care settings. They also reported that some treatments for children 5 years and younger might be effective but criticized the lack of well-conducted studies. As consequence, the US Preventive Service Task Force continued not to recommend universal language screenings for language delay (22). For the German speaking community, the German Institute for Quality and Efficiency in Health Care (IQWIG; Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen) also criticized the lack of evidence for long-term outcomes of language therapy. Following international systematic reviews (21, 23) the implementation of universal language screenings in Germany was not recommended (24).

So far, no standardized language screening instrument validated for use in Austrian pre-schools has been available. In Germany, several federal states commissioned research institutes to generate standardized language measures for the identification of language delayed children [i.e., Sismik & Seldak in Bavaria from (25, 26); HASE in Baden-Wuerttemberg from (27); KiSS in Hessia from (28) or Delfin 4 in North Rhine-Westfalia from (29) to name some]. Nevertheless, an analysis of the German Mercator- Institute for language promotion and German as second language ascertained insufficient quality and efficacy for all the screenings, mainly because of lack of sufficient validity and objectivity and the exclusion of multilingualism (30).

In Upper Austria, the request for a standardized procedure to be used within the regular universal check-ups in pre-schools, led to the LOGiK-S (Logopädie im Kindergarten—Screening) project. The new measure assesses language skills in Standard Austrian German, the variety of Standard German spoken in Austria in more formal situations (eg in schools and in the media) and with the highest sociolinguistic prestige. In less formal situations most Austrians use dialectal variations of German (Bavarian and Alemannic). The minor differences between Austrian German and Standard German spoken in Germany relate particularly to vocabulary and idiomatic expressions and less to language structure.

Our aim was to develop an accurate screening tool for the identification of high risk of LD (of unknown origin or associated with other biomedical conditions) in Austrian children and to evaluate its feasibility in the pre-school community setting.

METHODS AND PROCEDURES

Participant Recruitment

In summer 2012 and summer 2013, the public pre-schools in the city of Linz and in the whole state of Upper Austria were invited to participate in the project LOGiK-S (logopedics in kindergarten–screening) with the aim to develop a standardized instrument for language screening. In total, 31 pre-schools (14 of them well spread over different districts of Linz and 17 in the districts of Upper Austria) agreed to participate in the study. The recruitment of pre-schools in two consecutive years was due to limited human resources in the research team and to avoid overburdening the collaborating pre-schools. The managers of the pre-schools disseminated information about the project to all parents of children in their penultimate year of pre-school (age of 4-5 ½ years; Children attending their penultimate year of pre-school in the school year 2012/2013 are hereafter labeled as Cohort A and children attending their penultimate year of pre-school in the school year 2013/2014 are labeled as Cohort B) and asked for written consent for their children's participation. Overall, 423 monolingual children with German as their only language (as reported by the pre-school teachers) were eligible to participate. 97.9% of the parents (total $n = 414$, $n = 208$ in Cohort A and $n = 206$ in Cohort B) gave their written permission for inclusion in the research study. Testing was conducted in the first half of the school year (October 2012 to April 2013 for Cohort A and September 2013–March 2014 for Cohort B). We excluded children with incomplete data on the screening and reference tests ($n = 13$ in Cohort A and $n = 16$ in Cohort B), children with a time interval between screening and reference test of more than 60 days ($n = 7$ in Cohort B) and children outside the target age range ($n = 1$ in Cohort A and $n = 3$ in Cohort B). The remaining $n = 374$ children ($n = 194$ in Cohort A and $n = 180$ in Cohort B) were included in this study. **Table 1** provides an overview of the sample characteristics. Half of the children were girls (50.0%). The mean age was 55.66 months (SD = 4.01), whereas Cohort B was about 1 month older than Cohort A (t = 2.100, $p < 0.05$). Compared to the Upper Austrian parent population (29), the share of parents with university degree was overrepresented in the sample [36.1% vs. 25%; $\chi^2(3) = 28.725$, $p < 0.001$], which can be probably be explained in part by the exclusion of children with first languages other than German, whose parents are less likely to have a university degree [(29) Population data on parental education are not available for German-speaking children]. Moreover, there were some differences in parental education between Cohort A and Cohort B (see **Table 1**), most likely due to different catchment areas of pre-schools. However, these differences were not significant [$\chi^2(3) = 7.604$, $p > 0.05$]. For the analyses of this paper, we used pooled data (i.e. we analyzed cohort A and cohort B together) to maximize statistical power. Data pooling would also increase external validity, as the pooled sample is likely to be more heterogeneous in terms of individual characteristics than the single cohorts.

The study project (cohorts A and B) was approved by the hospital's ethic commission "Ethikkommission Barmherzige Schwestern und Barmherzige Brüder".

TABLE 1 | Participant characteristics.

	Cohort A (2012) ***n* = 194**	**Cohort B (2013)** ***n* = 180**	**Total** ***n* = 274**
Number of pre-schools	17	18	31
Females %	50.5%	49.9%	50.0%
Age M (SD)	55.24 (4.02)	56.11 (3.94)	55.66 (4.01)
Highest parental education[a] %			
Compulsory education (or below)	6.4%	2.5%	4.6%
Vocational education	33.5%	39.9%	36.4%
University entrance level	19.7%	26.%	22.8%
University degree	40.4%	31.0%	36.1%

[a] *the highest education of the two parents was used.*

Measures

Construction of the Screening Measures

At the age range relevant for the current study (4 ½ to 5 years) the primary markers of LD in German are deficits in morphosyntax, such as lacking or incorrect inflection of verbs (31), subject-verb-agreement (30) or use of function words (31). In addition, clinical experience shows that the valid assessment of grammatical skills is less time-consuming than the assessment of vocabulary. An expressive and receptive screening scale was developed because LD can affect the production and comprehension of language structures. In addition, assessments of language reception do not require the child's active production of language and therefore, higher acceptance of the receptive language assessment was anticipated. For both screening scales, grammatical structures that are usually acquired at pre-school age were selected. Based on the available literature on acquisition of German grammar (32–37), morphosyntactic structures with different degrees of complexity were selected. Children in their penultimate year of pre-school were chosen as the target group by request of the public authorities, following the tradition of universal language screening before the final year of pre-school, when—if necessary—intervention can be implemented before school entry.

Expressive Grammar Screening

The expressive grammar (EG) scale includes sentence completion tasks eliciting spoken phrases from the child with the help of predetermined sentence patterns. The scale includes 17 items. The tester successively presents two pictures, separated by a dividing line. The grammatical pattern structure is introduced with reference to the first picture (e.g., "Look! This is Tobias. He drinks juice."). After that, the child completes the sentence presented along with the second picture eliciting the same grammatical target structure (e.g. "*And this is Maria. She …*"– target structure: verb second position). Child utterances are scored as correct, when the child is able to produce the target grammatical structure. Errors beyond the targeted grammatical structure are negligible. To facilitate the scoring (0/1 points) of the expressive language items, a collection of correct and incorrect answers is provided. Notably, in cohort A, the screening scale comprised a total of 27 items. The final set of 17 items for measuring EG was selected based on the item statistics (difficulties, item-scale correlation) and the feedback of speech therapists who administered the screenings.

Receptive Grammar Screening

The receptive grammar (RG) scale includes 14 items, again ranked by anticipated increase in complexity, following German language acquisition research. Single sentences are read aloud by the administrator of the test and the child is asked to point to the corresponding picture from a selection of four with well-chosen semantic and grammatical distractors. The test items assess comprehension of different syntactic (e.g. "*The boy slides and the girl swings*"—coordination) or morphological structures (e.g. "*He gives her the book.*"—pronouns). Similar to the EG scale development, an initial number of 20 items was reduced to 14 items based on results of cohort A.

Reference Language Tests

Without an accurately defined gold standard for LD in the literature, LD was operationalized by significant deficits (-1.25 standard deviations below the norm) in at least two of the three linguistic dimensions of EG, RG and expressive vocabulary (compare 2–4).

(1) Expressive grammatical skills were assessed by the plural and case (accusative and dative) marking subtests of the PDSS [(38); *Patholinguistische Diagnostik bei Sprachentwicklungsstörungen*] as well as the subtests for comparatives, superlatives and participle perfect formation of the ETS 4-8 [(39); *Entwicklungstest Sprache für Kinder von 4 bis 8 Jahren*]. Following the results of a principal component analysis (PCA; one component with an eigenvalue of 2.27, 57% explained variance; loading between .71 and .83), we saved the component score (z-score with $M = 0$ and $SD = 1$) to be used as a single EG measure. Internal consistency (Cronbach's α) was good at .73. Children were classified as atypical in EG, if they scored in the bottom 10% (-1.25 SD) of the component score.

(2) The TROG-D [German version of the Test the Reception of Grammar; (40)] assesses the understanding of German grammar. Although the TROG-D provides norm values for German-speaking children, these norms are based on a substantially smaller number of children than contained in this study and do not include children speaking Austrian varieties of German. Therefore, we used the sample percentiles to identify the bottom 10% (-1.25 SD) of the TROG-D scores. Based on three age groups (48–50 months, 51–56 months, and 57–62 months), percentiles were estimated using a continuous norming approach as implemented in the Cnormj package (41) in jamovi 1.6 (42).

(3) The AWST-R [Revised Active Vocabulary Test for 3- to 5-year-old children, Aktiver Wortschatztest für 3- bis 5-Jährige, Revision; (43)] is a standardized picture-naming test for the age range from 3;0 to 5;5 years. The items are ordered by increasing difficulty. To reduce the length of the assessment, we only used the first of the two picture folders (35 items) for the assessment of expressive vocabulary. As the AWST-R lacks norm values for the reduced version of 35 items, we again estimated norm values based on the study data. We once more applied a continuous norming approach. Screening scores in the bottom 10% were considered atypical.

Based on our definition, children with atypical scores (≤ -1.25 SD) in at least two of the reference tests were classified as LD. This applies to 38 children (10.2%).

Feasibility

A short questionnaire (7 items) was developed for screeners to assess time economy, acceptance of the screening materials by children and test administrators, practicability of LOGiK-S within the constraints of the universal screening procedure in the pre-school setting, ease of administration and estimation of sensitivity. Finally, testers were asked whether they would recommend the screening to others. All items were coded by use of three-point Likert scales, except the last one (yes-no answer). Due to the high similarity of the materials and procedures for children with German as their dominant language and children with a first language other than German no separate versions of the feasibility questionnaire were completed by the screeners. Only for information on screening time specific information relating exclusively to the LOGiK-S version for children speaking dominantly German was collected.

Procedures

The screening procedures for both cohorts (A and B) were carried out by the speech and language pathologists, who usually conduct the annual universal language screening for children in their penultimate year in pre-school. The assessments were performed with each child individually in a separate room of their pre-school. The RG scale was introduced by a practice item to ensure the child's comprehension of the task and it was administered first, because it is usually perceived as less demanding or threating as no language production by the child is required. Within a maximum of 90 days, language development of the children was tested by use of standardized reference tests. The tests were administered in the pre-schools by experienced language experts from the Institute of Neurology of Senses and Language, who were blinded to the screening results.

Statistical Analyses

First, we report descriptive statistics for the subscales. Second, we report reliability estimates (Kuder-Richardson KR-20) for the screening scales. Third, to evaluate construct validity of the screening scales, we applied confirmatory factor analysis (CFA) for binary items using a weighted least squares estimation (WLSMV) in Mplus 8 (44). Following the guidelines proposed by (41, 45) a good model fit is indicated by $\chi^2/df \leq 2$, $CFI \geq 0.97$, $RMSEA \leq 0.05$. An acceptable fit is indicated by $\chi^2/df \leq 3$, $CFI \geq 0.95$, $RMSEA \leq 0.08$. Fourth, to evaluate criterion validity, receiver operator characteristic (ROC) analyses were used to evaluate the diagnostic accuracy of the subscales. Following Swets (46), AUCs $\geq$ 0.9 are regarded as excellent, AUCs $\geq$ 0.8 and $<$ 0.9 as good, AUCs $\geq$ 0.7 and $<$ 0.8 as fair, and tests with AUCs $<$ 0.7 as poor. To compare AUCs of the subtests, we used a bootstrapped test for paired ROC curves—as implemented in the pROC package (47) in R. Fifth, we applied logistic regression using Jamovi 1.6 (42) to investigate whether both subscales independently contribute to the prediction of LD. Sixth, to evaluate the generalizability of the screening results we compared

ROC curves between subsample (Cohort A vs. cohort B, boys vs. girls, age groups). As noted by Youngstrom (48) significant differences between subsamples would indicate variations in the diagnostic accuracy and thus, limit the generalizability of the screening results. A bootstrapped test for unpaired ROC curves was used to compare the AUCs between subgroups. Additionally, the Venkatraman permutation test (49) was used that compares actual ROC curves—not AUCs. If two ROC curves do not differ significantly, each cutoff values would result in the same sensitivity and specificity for the subsamples and therefore, a single cutoff would be appropriate for both subsamples. Finally, we used the R-OptimalCutpoints package (50) to determine appropriate cutoff scores. Cutoff scores are evaluated using the following diagnostic accuracy statistics: sensitivity (Se), specificity (Sp), positive predictive values (PPV), negative predictive values (NPV), and diagnostic likelihood ratios for positive and negative screening results (DLR+ and DLR–, respectively). Se and Sp $\geq$ 0.90 indicate good diagnostic accuracy, and Se and Sp $\geq$ 0.80 are regarded as fair. Values below.80 indicate an unacceptably high rate of misclassification (51). DLR+ indicates the multiplicative change in the pre-screening odds of having an LD given a positive screening result (i.e., post-screening odds = DLR+ $\times$ pre-screening odds) and DLR– is the change in the pre-screening odds of having an LD given a negative screening result (post-screening odds = DLR– $\times$ pre-screening odds). DLR+ values $\geq$ 10 and DLR– $\leq$ 0.1 indicate large changes in pre-screening odds, DLR+ $\leq$ 10 and $>$ 5, and DLR– $>$ 0.1 and $\leq$ 0.2 indicate moderate changes, DLR+ $\leq$ 5 and $>$ 2, and DLR– $>$ .2 and $\leq$ .5 indicate small changes. DLR+ $<$ 2 and DLR– $>$ 0.5 are rarely important (52).

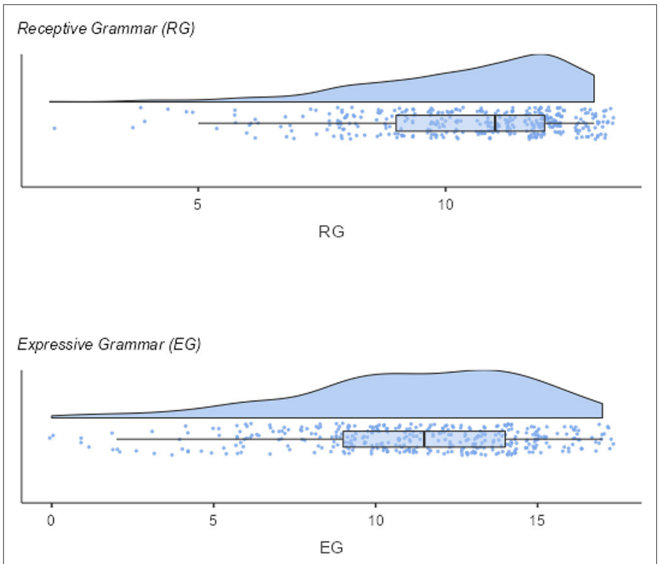


FIGURE 1 | Distribution of the screening scales.

RESULTS

Descriptive Statistics

Figure 1 shows the distribution of the RG and EG screening subscales. As exactable for an LD screening, items are rather easy and thus, just a few children score in the bottom range of the screening scales. Consequently, the empirical means (MRG = 10.5, SDRG = 2.01; MEG = 11.2, SDEG = 3.57) are higher than the midpoints of the scales (RG = 6.5, EG = 8.5).

Reliability

The internal consistency (KR-20) for the RG scale was rather low at .60. The internal consistency of the EG scale was of moderate size (KR-20 = .74).

Construct Validity

We performed separate CFAs for the screening subscales. Overall, the CFAs for RG and EG yielded an acceptable fit (RG: $\chi^2(65) = 90.046$, $p = 0.022$, RMSEA = 0.032, CFI = 0.927, EG: $\chi^2(119) = 245.049$ ($p < 0.001$), RMSEA = 0.053, CFI = 0.898). However, for EG the CFI was quite low but also near the cutoff of 0.90 what is also sometimes considered as acceptable [e.g., (53)]. Next, we compared a two-factor model (EG and RG) with a one-factor model (i.e., all EG and RG items load on a single factor). The two-factor model yielded an acceptable to good fit ($\chi^2(404) = 528.532$, $p < 0.001$, RMSEA = 0.029, CFI = 0.921). The fit for the one-factor model was somewhat worse ($\chi^2(405) = 568.600$, RMSEA = 0.033, $p < 0.001$, CFI = 0.896). Notably, a χ^2-difference tests indicated that the two-factor fits the data significantly better than the one-factor model [$\Delta\chi^2(1) = 18.406$, $p < 0.001$]. Overall, these results indicate that EG and RG are distinct but highly correlated constructs (latent correlation = 0.740, $p < 0.001$).

Criterion Validity—Diagnostic Accuracy of Subscales

The EG subscale yielded an excellent AUC of .918 (Delong 95%-CI [0.881, 0.954]). The AUC for the RG subscale (AUC = 0.826; Delong 95%-CI [0.749, 0.902]) was good, but—as indicated by a bootstrapped test for AUC-differences—significantly smaller than the AUC for EG ($D = -2.567$, $p < 0.05$).

Logistic Regression

A logistic regression showed that both subscales independently contribute to the prediction of LD (EG: $b = -0.430$, $p < 0.001$; OR = 0.650. RG: b = - 0.412, $p < 0.001$, OR = 0.662). McFadden's R^2 was .433. Notably, as coefficients (bs and odds ratios) for EG and RG were quite equal, an increase of 1 in both subscales is associated with a similar increase in the risk for LD. Thus, a simple sum of RG and EG is an appropriate and easy to calculate (and thus, feasible) total screening score. The AUC for the total screening score was excellent (AUC = 0.928, DeLong 95%-CI = [0.888, 0.976]).

Diagnostic Accuracy Differences Between Subgroups

The results of the comparisons of unpaired ROC curves (based on the total screening score) between subsamples are shown in **Table 2**. AUCs were generally excellent in all subsamples (only in the group of children younger than 56 months, the AUC was just

TABLE 2 | Tests for unpaired ROC curves.

	AUC	95%-CI (DeLong)	Comparision[a]
B—comparing cohorts			
(1) Cohort A (2012)	0.945	[0.900, 0.991]	
(2) Cohort B (2013)	0.905	[0.841, 0.958]	E = 0.006, $p > 0.05$; D = 0.883, p > 0.05
C—comparing age-groups			
(1) younger than 56 months	0.899	[0.822, 0.977]	
(2) 56 months and older	0.967	[0.939, 0.996]	E = 0.013, p >0 .05; D = −1.618, p > 0.05
D–comparing boys and girls			
(1) boys	0.913	[0.856, 0.970]	
(2) girls	0.945	[0.893, 0.997]	E = 0.007, p >.05; D = − 0.833, $p > 0.05$

[a] *The first test statistic E refers to the Venkatraman test for unpaired ROC curves. The second test statistic D refers to a bootstrapped test for unpaired ROC curves.*

below the limit of 0.90). Moreover, as indicated by insignificant group differences in AUCs (bootstrapped test for unpaired ROC curves) as well as in actual ROC curves (Venkatraman test), the diagnostic accuracy did not differ between the subsamples.

Cut-Off Estimation

Finally, to determine an optimal cut-off, we used the "SpEqualSe" criterion (i.e., specificity equals sensitivity) in the Optimal Cutoff R-Package (50). At a cut-off of 18 (21, 8% screening fails) yielded a sensitivity of 0.842 (95%-CI = [0.687, 0.940]) and a specificity of 0.853 (95%-CI = [0.810, 0.889]). The PPV was 0.395 (95%-CI = [0.325, 0.656]) and the NPV was 0.979 (95%-CI = [0.951, 0.985]). DLR+ and DLR– were of moderate size. DLR+ was 5.722 (95%-CI = [4.270, 7.671]) and DLR– was 0.185 (95%-CI = [0.089, 0.386]).

Feasibility

The 7-item questionnaire on feasibility of the LOGiK-S language screening, including both versions for children with German and Non-German as their dominant language and a phonology scale was returned by 39 (93%) from a total of 42 speech-language-therapists.

The average screening time was 9.49 min (SD 3.49). Screening materials were rated as very appealing by 44% and as appealing by 54%. Similarly, practicability within the constraints of universal language screening in the pre-school setting was rated as very good by 49% and as good by 46% of the respondents. Sensitivity (ie correct identification of children with LD) of LOGiK-S was assessed as very good by 15% and good by 80%. Thirty-nine percent described no personal effort in administering LOGiK-S, and another 90% stated low effort. As compared to the former screening without standardized measures 74% did not feel stressed at all by the new procedure whereas the rest reported minimal strain. Ninety-two percent would recommend the new measure to others.

DISCUSSION

This study investigated the performance (accuracy and feasibility) of the new screening measure LOGiK-S in a sample of two cohorts of 374 children in total, having German as their only or dominant language and attending the penultimate year of a public pre-school in Upper Austria. To avoid bias, the whole study sample that had been screened underwent testing by use of standardized language tests by speech-language experts blinded for the screening results. Screening results of the first cohort and practical experiences of the screeners were used to systematically reduce the number of screening items. Finally, all available data for the final selection of screening items were analyzed.

The EG scale of LOGiK-S demonstrated excellent accuracy (AUC = 0.918). The AUC of the RG scale was significantly smaller but still good (0.826). As indicated by logistic regression, both scales independently predict LD. A total screening score (combining EG and RG) showed excellent accuracy (AUC = 0.928). Using a cut-off of 18, the rate of screening fails was 21.8 %. Sensitivity (0.842) and specificity (0.853) were found to be good. As predictive values depend on the prevalence of the disorder under investigation (48), the rather low PPV (0.395) is not surprising given only 10.2 % of LD in our sample. Diagnostic likelihood ratios for positive and negative screening results (DLR+ and DLR–) of moderate size were found. Even though a higher PPV would be desirable as it leads to an overreferral of children, the dimensional nature of LDs must be taken into account. Children with false-positive screening scores have been shown to perform significantly lower on subsequent standardized measures than children with true-negative results (54) linked with a higher risk for language, psycho-social and cognitive delay. Therefore, follow-up diagnostic testing should be regarded as an opportunity to identify children with unmet needs for interventions (educational language and social support).

Tests for comparing unpaired ROC curves demonstrated no significant difference in screening accuracy (AUC and actual ROC curves) between both cohorts, despite some diversity in the study characteristics. Similarly, AUCs and ROC curves did not significantly differ between boys and girls and between younger and older children. Therefore, age related norms or sex related norms are not required. Overall, the independence of screening accuracy between groups (cohorts, sex, age) can be regarded as strengths of the screening instrument, as it supports its generalizability and therefore implementation with a variety of children and pre-schools can be recommended.

Feasibility of the new screening procedure was mostly rated as good or very good. Average screening time was below 10 min, materials were reported to be appealing to the children. Practicability within the constraints of the universal pre-school screenings was rated as very good and good. No or minimal personal effort involved in the administration of the new standardized instrument was described, and more than 90% of the screeners, who had to adapt their screening procedure to the new instrument, would recommend LOGiK-S to others.

Due to the lack of an accurately defined gold standard for LD in the literature we operationalized LD by language skills of at least 1.25 standard deviations below the norm in at least two of three linguistic dimensions following common practice in the field. Nevertheless, uncertainties of definition of the reference criterion must be considered a limitation. Moreover, a slight overrepresentation of children of parents with a university degree cannot be ruled out since population data for the specific target group (i.e., parents of children growing up monolingually in Upper Austria) are not available. Finally, the socioeconomic description of the sample is limited to parental education, because it was not possible to collect data on family income.

CONCLUSION

The LOGiK-S is the first validated language screening measure that identifies increased risk of LD in children with Austrian German as their first or dominant language in their penultimate year of pre-school. Accuracy of LOGiKS was found to be high. ccuracy. Implementation with a variety of screeners and in a variety of pre-schools confirms high feasibility of the new measure. Consequently, the implementation of LOGiK-S for universal language screening can be recommended in Austria.

DATA AVAILABILITY STATEMENT

The dataset presented in this article is not readily available because parents have not given their consent to data sharing. Requests to access the dataset should be directed to daniel.holzinger@bblinz.at.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethikkommission Barmherzige Schwestern und Barmherzige Brüder. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

DH: conceptualization, funding acquisition, and supervision. DH and CW: methodology. BD: validation, investigation, and data curation. CW: formal analysis. DH, BD, and CW: writing—review and editing. DH and BD: project administration, and writing—original draft preparation. All authors have read and agreed to the published version of the manuscript.

FUNDING

This work was supported by the Department of Social Affairs of the Upper Austrian Government. Article processing charge is funded by the Johannes Kepler University Open Access Publishing Fund.

ACKNOWLEDGMENTS

We express our gratitude to the Upper Austrian speech-language pathologists cooperating with us in the current project.

REFERENCES

1. Bishop D, Snowling MJ, Thompson PA, Greenhalgh T, the CATALISE-2 consortium. Phase 2 of CATALISE: a multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *J Child Psychol Psychiatry Allied Disciplines*. (2017) 58:1068–80. doi: 10.1111/jcpp.12721
2. Norbury CF, Vamvakas G, Gooch D, Baird G, Charman T, Simonoff E, et al. Language growth in children with heterogeneous language disorders: a population study. *J Child Psychol Psychiatry.* (2017) 58:1092–105. doi: 10.1111/jcpp.12793
3. Tomblin JB, Smith E, Zhang X. Epidemiology of specific language impairment: prenatal and perinatal risk factors. *J Commun Disord.* (1997) 30:325–44. doi: 10.1016/S0021-9924(97)00015-4
4. Leonard LB. Children with specific language impairment and their contribution to the study of language development. *J Child Lang.* (2014) 41 (Suppl. 1):38–47. doi: 10.1017/S0305000914000130
5. Thordardottir E. Proposed diagnostic procedures for use in bilingual and cross-linguistic contexts. *In:* Armon-Lotem S, Jong J, Meir N, editors. *Assessing Multilingual Children: Multilingual Matters.* (2015). p. 331–58.
6. Kiese-Himmel C. Entwicklung sprach-und kommunikationsgestörter Kinder, am Beispiel von≫ Late Talkers ≪sowie Kindern mit spezifischen Sprachentwicklungsstörungen. In: Marcus Hasselhorn and Rainer K Silbereisen (Hg), editors. *Enzyklopädie der Psychologie, Serie V (Entwicklungspsychologie), Bd.* (2008). p. 693–730.
7. Johnson CJ, Beitchman JH, Brownlie EB. Twenty-year follow-up of children with and without speech-language impairments: family, educational, occupational, and quality of life outcomes. *Am J Speech Lang Pathol.* (2010) 19:51–65. doi: 10.1044/1058-0360(2009/08-0083)
8. Durkin K, Mok PL, Conti-Ramsden G. Core subjects at the end of primary school: identifying and explaining relative strengths of children with specific language impairment (SLI). *Int J Lang Commun Disord*. (2015) 50:226–240. doi: 10.1111/1460-6984.12137
9. St Clair MC, Pickles A, Durkin K, Conti-Ramsden G. A longitudinal study of behavioral, emotional and social difficulties in individuals with a history of specific language impairment (SLI). *J Commun Disord.* (2011) 44:186–99. doi: 10.1016/j.jcomdis.2010.09.004
10. Yew SG, O'Kearney R. Emotional and behavioural outcomes later in childhood and adolescence for children with specific language impairments: meta-analyses of controlled prospective studies. J Child Psychol Psychiatr. (2013) 54:516–24. doi: 10.1111/jcpp.12009
11. Conti-Ramsden G, Mok PL, Pickles A, Durkin K. Adolescents with a history of specific language impairment (SLI): strengths and difficulties in social, emotional and behavioral functioning. *Res Dev Disabil*. (2013) 34:4161. doi: 10.1016/j.ridd.2013.08.043

12. Law J, Garrett Z, Nye C. The efficacy of treatment for children with developmental speech and language delay/disorder: a meta-analysis. *JSLHR.* (2004) 47:924–43. doi: 10.1044/1092-4388(2004/069)
13. Roberts MY, Kaiser AP. The effectiveness of parent-implemented language interventions: a meta-analysis. *Am J Speech Lang Path.* (2011) 20:180–99. doi: 10.1044/1058-0360(2011/10-0055)
14. Reilly S, Wake M, Ukoumunne OC, Bavin E, Prior M, Cini E, et al. Predicting language outcomes at 4 years of age: findings from early language in victoria study. *Pediatrics.* (2010) 126:e1530–7. doi: 10.1542/peds.2010-0254
15. Rice ML, Taylor CL, Zubrick SR. Language outcomes of 7-year-old children with or without a history of late language emergence at 24 months. *J Speech Lang Hear Res*. (2008) 51:394–407 doi: 10.1044/1092-4388(2008/029)
16. Dale PS, Price TS, Bishop DV, Plomin, R. Outcomes of early language delay. *J Speech Lang Hear Res.* (2003) 46:544–60 doi: 10.1044/1092-4388(2003/044)
17. Law J, Tomblin JB, Zhang X. Characterizing the growth trajectories of language-impaired children between 7 and 11 years of age. *JSLHR.* (2008) 51:739–49. doi: 10.1044/1092-4388(2008/052)
18. Eadie P, Levickis P, McKean C, Westrupp E, Bavin EL, Ware RS, et al. Developing preschool language surveillance models-cumulative and clustering patterns of early life factors in the early language in victoria study cohort. *Front Pediatr.* (2022) 10:826817. doi: 10.3389/fped.2022.826817
19. McKean C, Reilly S, Bavin EL, Bretherton L, Cini E, Conway L, et al. Language outcomes at 7 years: early predictors and co-occurring difficulties. *Pediatrics.* (2017) 139:e20161684. doi: 10.1542/peds.2016-1684
20. Wallace IF, Berkman ND, Watson LR, Coyne-Beasley T, Wood CT, Cullen K, et al. Screening for speech and language delay in children 5 years old and younger: a systematic review. *Pediatrics.* (2015) 136:e448–62. doi: 10.1542/peds.2014-3889
21. Nelson HD, Nygren P, Walker M, Panoscha R. Screening for speech and language delay in preschool children: systematic evidence review for the US Preventive Services Task Force. *Pediatrics.* (2006) 117:e298–319. doi: 10.1542/peds.2005-1467
22. Siu AL, and US, preventive services task force. screening for speech and language delay and disorders in children aged 5 years or younger: us preventive services task force recommendation statement. *Pediatrics.* (2015) 136:e474–81. doi: 10.1542/peds.2015-1711
23. Law J, Harris F. The parameters of early language screening. *Child Care Health Dev.* (2000) 26:83–9. doi: 10.1046/j.1365-2214.2000.00172.x
24. IQWIG -, Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (2009). *Früherkennungsuntersuchung auf umschriebene Entwicklungsstörungen des Sprechens und der Sprache: Abschlussbericht*; Version 1.0, Nr. 57. Available online at: https://www.iqwig.de/download/s06-01_abschlussbericht_frueherkennung_umschriebener_stoerungen_des_sprechens_und_der_sprache.pdf (accessed June 2, 2022).
25. Ulich, M, Mayr, T. *Sismik*. Sprachverhalten und Interesse an Sprache bei Migrantenkindern in Kindertageseinrichtungen (Beobachtungsbogen und Begleitheft). Freiburg: Herder (2003).
26. Ulich M, Mayr T. Seldak. Sprachentwicklung und Literacy bei deutschsprachig aufwachsenden Kindern (Beobachtungsbogen und Begleitheft). Freiburg: Herder (2006)
27. Schöler H, Brunner M. *HASE – Heidelberger Auditives Screening in der Einschulungsu ntersuchung*. Wertingen: Westra (2008).
28. Neumann K, Holler-Zittlau I, van Minnen S, Sick U, Euler HA. *Das Kinder-Sprach-Screening (KiSS) – das hessische Verfahren zur Feststellung des Sprachstandes vierjähriger* Kinder (2008).
29. Bundesinstitut BIFIE (Ed.) (2019). Standardüberprüfung 2018, Mathematik, 4. Schulstufe, Landesergebnisbericht Oberösterreich. Salzburg: Bundesinstitut BIFIE.
30. Fried L, Briedigkeit E, Isele P, Schunder, R. Delfin 4 – Sprachkompetenzmodell und Messgüte eines Instrumentariums zur Diagnose, Förderung und Elternarbeit in Bezug auf die Sprachkompetenz. *Zeitschrift für Grundschulforschung*, 2/2009, 2:13–26.
31. Neugebauer U, Becker-Mrotzek, M. *Die Qualität von Sprachstandsverfahren im Elementarbereich. Eine Analyse und Bewertung*. Köln: Mercator-Institut für Sprachförderung und Deutsch als Zweitsprache (2013).
32. Chilla S, Haberzettl S, Wulff N. Dummy Auxiliaries in First and Second Language Acquisition in German. In: E. Blom et al. *(eds.): Dummy Auxiliaries in First and Second Language Acquisition. Boston-Berlin: DeGruyter*. (2013). p. 209–250. doi: 10.1515/9781614513476.209
33. Rothweiler M, Chilla S, Clahsen H. Subject verb agreement in specific language impairment: a study of monolingual and bilingual German-speaking children. *Biling: Lang Cogn.* (2012) 15:39–57. doi: 10.1017/S136672891100037X
34. Rothweiler M. Über den Zusammenhang von Lexikon, Grammatik und Mehrsprachigkeit: Was kann die Spracherwerbsforschung für die Praxis liefern? *Sprachheilarbeit.* (2009) 54:246–54.
35. Clahsen H, Richman K. *Child Language and Developmental Dysphasia*. Amsterdam: John Benjamins Publishing Company (1991).
36. Hamann C, Penner Z, Lindner K. German impaired grammar: the clause structure revisited. *Language Acquisition.* (1998) 7(2/4):193–245. doi: 10.1207/s15327817la0702-4_5
37. Hamann C. Specific language impairment in German speaking children. Specific language impairment. *Current Trends Res.* Amsterdam: Benjamins (2015) 215–251. doi: 10.1075/lald.58.10ham
38. Kauschke C, Siegmüller J. *Patholinguistische Diagnostik bei Sprachentwicklungsstörungen: PDSS. 2.* Aufl. Stuttgart: Urban and Fischer (2009).
39. Rmaier M. *Entwicklungstest Sprache für Kinder von 4 bis 8 Jahren: ETS 4 - 8; Manual. 1.* Aufl. Frankfurt am Main: Harcourt Test Service (2007).
40. Fox AV, Bäumer T. (Hg.). TROG-D: Test zur Überprüfung des Grammatikverständnisses; Handbuch. 1. Aufl. Idstein: Schulz-Kirchner (Das Gesundheitsforum) (2006).
41. Lenhard W, Lenhard A, Gary S. *cNORM: Continuous Norming.* Vienna: The Comprehensive R Network (2018).
42. The jamovi project. Jamovi (Version 1.6) [Computer Software] (2021). Available online at: https://www.jamovi.org
43. Kiese-Himmel C. *AWST-R: - Revision -: Manual*. Göttingen: Beltz Test Hogrefe Verlagsgruppe (2005).
44. Muthén LK, Muthén B. *Mplus user's Guide*. Los Angeles, CA: Muthén and Muthén (1998–2017).
45. Schermelleh-Engel K, Moosbrugger H, Müller H. Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online.* (2003) 8:23–74.
46. Swets JA. Measuring the accuracy of diagnostic systems. *Science.* (1988) 240:1285–93.
47. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics.* (2011) 12:77. doi: 10.1186/1471-2105-12-77
48. Youngstrom EA. A primer on receiver operating characteristic analysis and diagnostic efficiency statistics for pediatric psychology: we are ready to ROC. *J Pediatr Psychol.* (2014) 39:204–21. doi: 10.1093/jpepsy/jst062
49. Venkatraman ES. A permutation test to compare receiver operating characteristic curves. *Biometrics.* (2000) 56:1134–8. doi: 10.1111/j.0006-341X.2000.01134.x
50. López-Ratón M, Rodríguez-Álvarez, M, Cadarso-Suárez C, Gude-Sampedro F. OptimalCutpoints: an R package for selecting optimal cutpoints in diagnostic tests. *J Statistical Software*. (2014) 61:1–36. doi: 10.18637/jss.v061.i08
51. Plante E, Vance R. Selection of pre-school language tests: a data-based approach. *Lang Speech Hear Serv Sch.* (1994) 25:15–24. doi: 10.1044/0161-1461.2501.15
52. Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. *JAMA*. (1994) 271:703–7. doi: 10.1001/jama.271.9.703
53. Kline RB. *Principles and Practice of Structural Equation Modeling*. New York: The Guilford Press. (2016).

54. Glascoe, F. P. (2001). Are overreferrals on developmental screening tests really a problem?. *Arch. Pediatr. Adolesc. Med.* 155, 54–59. doi: 10.1001/archpedi.155.1.54

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.

frontiers | Frontiers in Pediatrics

TYPE Original Research
PUBLISHED 06 September 2022
DOI 10.3389/fped.2022.865457

OPEN ACCESS

EDITED BY
Tim S. Nawrot,
University of Hasselt, Belgium

REVIEWED BY
Tjhin Wiguna,
University of Indonesia, Indonesia
Adriana Weisleder,
Northwestern University, United States

*CORRESPONDENCE
Daniel Holzinger
daniel.holzinger@jku.at

SPECIALTY SECTION
This article was submitted to Children and Health, a section of the journal Frontiers in Pediatrics

RECEIVED 29 January 2022
ACCEPTED 12 August 2022
PUBLISHED 06 September 2022

CITATION
Holzinger D, Weber C and Fellinger J (2022) Validity and feasibility of a predictive language screening tool in 2-year-old children in primary pediatric care.
Front. Pediatr. 10:865457.
doi: 10.3389/fped.2022.865457

Validity and feasibility of a predictive language screening tool in 2-year-old children in primary pediatric care

Daniel Holzinger[1,2,3]*, Christoph Weber[2,4] and Johannes Fellinger[1,2,5]

[1]Institute of Neurology of Senses and Language, Hospital of St. John of God, Linz, Austria, [2]Research Institute for Developmental Medicine, Johannes Kepler University Linz, Linz, Austria, [3]Institute of Linguistics, University of Graz, Graz, Austria, [4]Department for Inclusive Education, University of Education Upper Austria, Linz, Austria, [5]Division of Social Psychiatry, University Clinic for Psychiatry and Psychotherapy, Medical University of Vienna, Vienna, Austria

Objective: To assess the predictive validity and feasibility of the newly developed language screening tool, SPES-2 (Sprachentwicklungsscreening), for 2-year-old children in pediatric primary care.

Methods: A prospective cohort study recruited 2,044 non-selected German-speaking children undergoing a regular well-baby check-up at the age of 2 years. Thirty primary care pediatricians spread over urban and rural areas screened the children using a short parent-reported questionnaire and direct assessment of word comprehension. To validate the screening tool, language skills were assessed using a standardized language screening tool in the complete sample 1 year later. Data of a random sample of 621 children were analyzed. Feasibility of the screening tool was evaluated using questionnaires completed by the participating pediatricians.

Results: The new screening tool, SPES-2, demonstrated good diagnostic accuracy with AUC (Area under the Roc Curve) of 0.885, a sensitivity of 0.74, and specificity of 0.86, using a parent-reported questionnaire (expressive vocabulary, two-word combinations, parental concerns) as stage 1, followed by a stage 2 direct assessment of word comprehension by the pediatrician. The second stage was restricted to children who failed the parental screening. The screening identified children with high, moderate, and low risk of significant language deficits (SLD) at the age of 3 years, permitting tailored follow-up assessment and parental counseling. Practicality and acceptability of the screening were mostly rated as high. Pediatricians regarded the availability of follow-up diagnostic services and parent guidance as most important for a general implementation of the new instrument.

Conclusion: The language screening tool, SPES-2, was valid for the identification of significant language deficits 1 year later, and considered as feasible within primary pediatric care.

KEYWORDS

language screening, predictive, two-year olds, late language emergence, late talker

Introduction

Language disorders are among the most frequently diagnosed developmental disorders among young children, with reported prevalence rates of about 10%. These disorders may be of unknown origin (7.58%) or associated with other developmental disorders, such as general cognitive delay or autism spectrum disorder (1, 2). Children with language difficulties are at increased risk of adverse outcomes, including social isolation, mental health problems (3–5), academic problems (6, 7), placement in special education, and later unemployment (8, 9). Therefore, language problems affect the functioning of individual children and represent a loss for society.

Increasing evidence shows that early intervention—particularly if family-centered—is effective in improving language outcomes among those with language (10–15) difficulties (10–15).

As early identification and subsequent intervention may significantly influence functional outcomes in children, systematic developmental surveillance has been recommended by the American Academy of Pediatrics (16, 17). However, in a systematic review, Nelson et al. (18) indicated a lack of valid language screening instruments. Almost one decade later, Robins et al. (19) reported several screening tools that could accurately identify children with language disorder, but with insufficient evidence of feasibility in the pediatric primary care setting. As a result, the U.S. Preventive Task Force did not recommend population screening for language disorders (20). Following international reviews, the German Institute for Quality and Efficiency in Health Care (21) considered the available evidence insufficient for a recommendation of universal language screening in Germany.

The development of an accurate and feasible screening tool is complicated by the instability of developmental trajectories during the early years. By the age of 3–4 years, about half of the children with late language emergence at the age of two attain age-appropriate language levels (within one standard deviation) (22–24). In a national twin study in England and Wales (25), including 4,193 twin pairs, 56% of those with late language emergence did not meet criteria for persistent language disorder at the age of 3 years. Whereas some children show late language emergence, others manifest deterioration in the trajectory of language development over time. In Poll and Miller's (26) longitudinal study ($n = 1{,}015$), more than 60% of children with weak oral language skills at 8 years did not have a history of late language emergence at the age of 2 years. In the Early Language in Victoria Study with an epidemiological focus, less than half of the children with language disorder at the age of 4 years were identified as late talkers at the age of 2 years (27). The high rate of recovery from early language delay requires the development and validation of screening instruments predicting long term language outcomes (predictive validity). The emergence of later language disorder despite earlier typical performance suggests the necessity of continuous surveillance of language development.

The use of parent-reported or direct assessment is another critical issue in the development of an instrument screening for increased risk of language difficulties. In their systematic review on preschool screening tools for language and behavioral difficulties, Sim et al. concluded that parent-reported screening tools for language in preschool aged children achieved higher sensitivity, specificity, and negative predictive value than direct child assessment (28). However, Visser-Bochane et al. (29) reported high predictive validity for a screening instrument (van Wiechenschema) based on a combination of direct child assessment and parent report. Evaluations of parent-reported tools (MCDI and LDS) showed high specificity with moderate sensitivity, whereas the combined tool had poorer specificity but better sensitivity, and thus higher rates of prediction accuracy of children with language delays. In a study to compare the validity of parental screening and direct pediatric assessment of child development (30), direct professional assessment had a higher validity than parental reports. Therefore, parental reports may not necessarily be superior to direct observation or testing. The critical factor may be the availability of valid and easy-to-use instruments for language screening within the time constraints of primary pediatric care. Pediatricians have longitudinal relationships with children and families and therefore the child's entire medical history (31), facilitating interpretation and augmentation of screening based on parent report. By completing a short, direct observation of the child's language development, pediatricians may build trust that facilitates communication with families about their child's developmental delays.

A further essential characteristic of a modern screening instrument relates to the required resources for assessment and follow-up. 2-stage screeners collecting data from parents in the first step and requiring pediatric assessment in the second step exclusively with those failing the initial screening have been shown to be effective for the identification of autism (19). The majority of available language screening instruments do not include cut-offs that allow for stepped follow-up such as referral for diagnostic assessment or parent counseling for an exception see (29).

Language screening instruments that can be used within preventive medical care have been available for decades; examples of which include the MacArthur-Bates Communicative Development Inventories (MCDI) (32) and Language Development Survey (LDS) (33). In German-speaking countries, the ELFRA (Elternfragebogen) Parent Questionnaire (34), a parent-reported language screening tool with 260 items based on the MCDI and LDS, has been normed for 2-year-old children (35, 36). In addition, a shorter parent-reported word list (SBE 2 KT; Sprachbeurteilung durch Eltern Kurztest; a short language assessment test by parents) has been developed (37) and validated. The FRAKIS (Fragebogen zur kindlichen Sprachentwicklung; Questionnaire on child language development) (38) is another more extensive parent-reported questionnaire based on the MCDI that collects information on expressive vocabulary (600-item word list) and grammatical development. A short form of the FRAKIS with a word list of 102 items and three questions referring to grammatical development is available. A systematic review on the effectiveness of population-based language screening for children at pre-school age in Germany concluded that the accuracy of the screening instruments has not been sufficiently examined (39). Within the Austrian system of well-baby checkups, the screening protocol at the age of 2 years includes several questions on language development. However, neither cutoffs for non-typical development nor guidelines for subsequent referral are available. Therefore, no evaluations of the performance of the well-baby check-up for the identification of language delay in primary pediatric care in Austria have been published. Currently, none of the above mentioned instruments fulfills all of the following criteria: validation with a representative total population sample; validation within a community setting; follow-up including referral of children according to a strict referral protocol and subsequent systematic, standardized testing of a representative sample of children, independent of screening results; reporting predictive validity of the screening measures regarding language outcomes 1 year after the screening; and inclusion of parent report and pediatric assessment.

The Federal State of Upper Austria has a population of 1.45 million inhabitants and approximately 13,000–15,000 births annually within the last 20 years. Universal, free preventive medical care is available and provided primarily by pediatricians but also by general practitioners.

This study sought to establish and validate a language screening tool for 2-year-old children (i) in a representative comprehensive population sample; (ii) with high feasibility in routine primary care pediatric settings within the regular well-baby check-ups; (iii) predicting language problems 1 year after the screening; (iv) including a two-stage procedure of parent-reported assessment followed by pediatric direct assessment only in case of atypical results in the first screening stage, with the intention to include valuable parent information and to reduce the pediatricians' time required for direct assessment; (v) with sufficient sensitivity (identification of a high proportion of children with SLD at age three) and a high proportion of true positives of the screening positives (positive predictive value) to avoid costs for follow-up and unnecessary irritation of parents; and (vi) resulting in a graduation of risk levels of SLD (low, moderate, high) that allow for well-adjusted procedures following the screening.

Methods and procedures

This study is part of a comprehensive pre-school language surveillance project with the aim to establish a language screening tool for children at the ages of 2 and 3 years in the whole State of Upper Austria. The project was implemented in close cooperation with the pediatric association of Upper Austria. An initial pilot study (2007–2008) aimed to identify screening components that predict language disorders about a year after the administration of the screening tool (predictive validity) and thus establish a screening tool. In the actual validation study (2009–2010), the predictive validity of the new screening tool was assessed and reported in this paper.

Construction of the screening tool and pilot testing

The initial screening tool was constructed and implemented within a 2007–2008 pilot study (40) as a combination of a parent-reported questionnaire, child medical data available from the well-baby check-up, and direct pediatric assessment. A representative group of pediatricians (n = 30) across the State of Upper Austria participated in the study. The pediatricians were introduced to language development and language screening, and trained in the administration of the screening instruments. In the pilot study, the parent-reported questionnaire and direct pediatric assessment were administered to all children growing up with German as their primary family language. Children having another preferred family language than German and speaking German as their best language (and thus growing up multilingually) were also excluded from the pilot study.

The parent-reported questionnaire included (i) a short form of the ELFRA, [Elternfragebogen; Parent Questionnaire; (34)], a 260 items word list of the child's expressive vocabulary, (ii) a question on the child's use of two-word combinations, (iii) sociodemographic information of the family (parental education, the child's birth order), (iv) parental concerns about language development (typical, slightly delayed, severely delayed), (v) parental estimation of language development, (vi) family predisposition for language and literacy difficulties, and (vii) a history of otitis media. In addition, (viii) information about the primary language used in the family and the child's best language were collected. Medical data extracted from the regular documentation of the well-baby check-ups (Mutter-Kind-Pass; Mother-Child-Passport) included gestational age, multiple birth, APGAR (Appearance, Pulse, Grimace, Activity, and Respiration) scores, birth weight, size, head circumference, and preceding diagnoses. The pediatricians assessed word comprehension by asking the child to identify six body parts on a doll.

For children at the age of 3 years, the pediatricians performed two standardized subtests of a comprehensive German language test [SETK-3-5 (41)], assessing noun plurals and sentence comprehension. To determine predictive validity of the screening measures for children at the age of 2 years, a validation sample of 141 children (64 with ELFRA scores below 50, 29 with ELFRA scores between 50 and 79, and a group of 48 children with scores above 79) was assessed by speech-language experts blinded for the screening results by use of standardized tests, including noun plurals, non-word repetition, sentence comprehension, and encoding of semantic relations (SETK-3-5) and expressive vocabulary [AWST-R (42)]. Cases of children with significant language delay were extrapolated from the subsample of 141 children on the total sample of children with German as their preferred family language and complete data at the ages of 2 and 3 years ($n = 1{,}543$). This extrapolation was based on the correlation between the results of the language tests performed by speech-language experts and pediatricians.

In the pilot study, parent reports on expressive vocabulary (ELFRA 2) of children at the age of 2 years were found to be the best predictors of language status 1 year later. Other factors that added significant predictive quality were (in this sequence) sentence use of two-word utterances, parental concerns, parental estimation of language development, and the short pediatric assessment of word comprehension. Based on these findings, the screening procedure was reduced to a shortened parent-reported questionnaire and direct assessment of word comprehension by the pediatricians.

Study procedures and recruitment

In 2009, the Pediatric Association of Upper Austria invited all primary care pediatric offices to participate in a longitudinal project on language screening. Thirty pediatricians (about 50% of all primary care pediatricians of Upper Austria) from across the country agreed to screen all children for language disorders at the ages of 2 and 3 years within their regular well-baby check-ups. The pediatricians and their office assistants were trained in the administration of the final screening procedures, scoring, and communication of results to parents, and provided with the screening materials.

In 2009, about 66% of the entire Upper Austrian birth cohort of 2007 ($n = 13{,}297$) were assessed during regular well-baby check-ups, 70% (6,200) of them by pediatricians and the others by general practitioners. This study used data from 2,044 children growing up with German as their primary family language screened at the ages 2 (23–25 months) and 3 years (35–7 months). This accounts for about 15% of the entire birth cohort and 33% of children examined by pediatricians. Overall, the sample was representative, regarding sex ratio and prematurity rate (see Table 1). Additionally, the distribution of maternal education was comparable to that of the population. Notably, the seeming overrepresentation of higher educated parents in the sample is likely due to the exclusion of non-German speaking children, whose parents have an lower education level than parents of German speaking children (45). Notably, the initially recorded data contained only the sum score of the 260-word expressive vocabulary list. To reduce the administration time by parents, a random sample ($n = 667$; see Figure 1) stratified by age, gender, and maternal education was drawn from the total sample; and for these children, all 260 items were separately entered in retrospect. The item reduction to a final list of 37 words was achieved by deleting items used by <25% of the children, items with considerable differences by age (23–25 months) and those with a positive predictive value below 50% without significant reduction of predictive quality with regard to a preliminary definition of SLD.

In total, 46 children were excluded due to incomplete data on the parent reports ($n = 4$) and pediatrician assessment ($n = 40$). Thus, the final sample used for the analyses in this study comprised 621 children (mean age = 23.92 months, SD = 0.972; 51.8 % were boys, 48.2% were girls). Descriptive statistics for the study variables are reported in Table 1. About 11.5% of the children were regarded as having significant language deficits (SLD) at the age of 3 years based on the reference test. Notably, the final sample ($n = 621$) used in this study did not substantially deviate from the full sample ($n = 2{,}044$; see Table 1 for details).

Measures

Screening measures

The final screening measures (SPES-2, Sprachentwicklungsscreening-2; language screening for 2-year-old children) used in the validation study included a parent-reported questionnaire with demographic information on the child (age, sex, age at completion of questionnaire, birth order, child's best language) and the family (maternal and paternal education, primary family language). Other information included parental estimation of the child's language development, parental concerns about language development, the 37 items on expressive vocabulary, and a question on two-word-combination use.

Since word comprehension was found to contribute to the prediction of SLD in the pilot study, a word comprehension subtest of a standardized German language test [SETK-2; (46)] was selected for the direct pediatric assessment. The pediatrician directs single words (9 items; Cronbach's Alpha = 0.69) referring to household items, food, animals and body parts to the child, and the child is asked to identify the corresponding picture from four options.

Reference test for significant language deficits at the age of 3 years

In the absence of a well-defined standard for diagnosing language disorder, we used the SPES-3 language screening measure (47) as reference test for SLD at the age of 3 years.

TABLE 1 Random and total sample characteristics.

	A	B	C	Difference between B and C[c]
	Population[b] (n = 13,297)	Full sample (n = 2,044)	Analysis sample (n = 621)	
Child age M (SD)		23.92 (0.993)	23.92 (0.972)	$d = 0.00, p > 0.05$
Child sex (male)[d] %	51.1%	50.9%	51.8%	$\phi = 0.01, p > 0.05$
Premature birth[e] %	8.7%	8.6%	8.7%	$\phi = 0.00, p > 0.05$
Highest parental education[a,f] %				
Compulsory education (or below)	7%	2.2%	1.4%	Cramer's V = 0.05,$p > 0.05$
Vocational education	47%	47.7%	47.0%	
University entrance	21%	23.5%	25.8%	
University	25%	26.6%	25.8%	
Parent reports				
Expressive vocabulary M (SD)		132.75 (65.617)	134.94 (64.789)	$d = -0.04, p > 0.05$
No two-word combinations %		6.9%	5.6%	$\phi = -0.02, p > 0.05$
Parental concerns (yes) %		13.2%	12.4%	$\phi = -0.02, p > 0.05$
Pediatric assessment (Age 2) M (SD)		6.91 (1.747)	6.86 (1.796)	$d = 0.05, p > 0.05$
Significant language deficit (−1.5 SD at age 3)[g] %		9.1%	11.3%	$\phi = 0.05, p < 0.05$

[a]The highest education of the two parents was used. [b]Population values for child sex and premature birth are taken from (43) and directly refer to Upper Austrian birth cohort of 2007. Due to the lack of population data for parental education directly referring to the birth cohort of 2007, we use parental education values from the parent population of Upper Austrian 4. graders of 2018 as proxies (44). Notably, these values also include Non-German speaking parents. Detailed data by language use for Upper Austria are not available. [c]p-values for categorical variables refer to χ^2-tests (effect size ϕ or Cramer's V). p-values for continuous variables refer to t-tests (effect size d). [d]Difference between A and B: $\phi = 0.00$, p > 0.05; [e]Difference between A and B: $\phi = 0.01$, p > 0.05; [f] Difference between A and B: Cramer's = 0.20, p < 0.001.

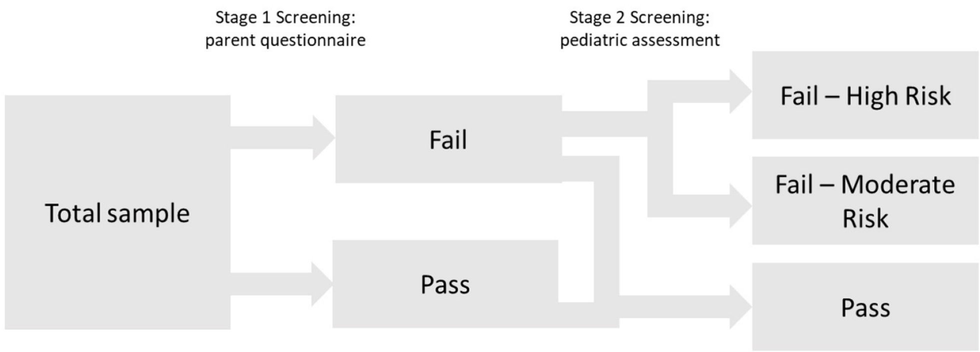


FIGURE 1
Conceptual overview of the two-stage screening tool.

The SPES-3 language screening—consisting of parent reports on expressive vocabulary and expressive grammar—is administered during regular well-baby check-ups at the age of 3 years. The SPES-3 screening tool has an excellent diagnostic accuracy (AUC = 0.946) in predicting language disorder (assessed by two experienced clinical linguists blinded to the SPES-3 result). For this study, a cutoff of 1.5 SD below the mean of the total SPES-3 screening score was used [for details see (47)].

Feasibility measures

Feasibility was measured by use of a questionnaire completed by the participating pediatricians. The questionnaire was designed following the guidelines by Bowen et al. (48). It included three components on feasibility, with all questions rated on a 4-point Likert scale (very good, good, difficult, and very difficult). Practicality concerns the extent to which administration of the screening was considered possible within regular pediatric primary care. In addition, pediatricians were asked to report on the ease of administration of the direct assessment and parental difficulties in completing the parent-reported questionnaire. Additionally, pediatricians ranked five pre-specified factors that might challenge the administration of the screening tool. Acceptability refers to the reactions to the screening tool by the children, parents, and pediatricians, including the proportion of direct screening tests that could be completed by the pediatricians. The pediatricians were asked to assess parental acceptance of the inclusion of language screening in the medical check-up of children at the age of 2 years, and assess the language screening tool's usefulness. Finally, sustainability was assessed by asking the pediatricians whether they intended to continue language screening after the research project ended.

Analytic strategy

A central aim of this study was to develop a two-stage screening procedure made up of an assessment based on parent report (stage 1) followed by pediatric direct assessment (stage 2) only for cases with atypical results in the first screening stage. To achieve this aim, we used in a first step logistic regression to predict SLD at the age of 3 years using the set of parent-reported language-related variables (expressive vocabulary list, parental concerns, two-word combinations) and child and parent-related sociodemographic variables (child age, child sex, parental education) as possible predictors.

Significant predictors were identified using a backward variable selection algorithm (Likelihood Ratio).

Subsequently, a stage 1 screening score was computed as the probability of having SLD based on the logistic regression results.

Screening failures at stage 1 that would undergo direct pediatric assessment at stage 2 were determined by selecting a cutoff point with a quite high sensitivity of 0.90. Thus, the group of stage 1 screening failures covers 90% of all children with SLD. Commonly, a sensitivity of 0.80 is regarded as acceptable for developmental screenings (49). In their meta-analysis of predictive language screening tools at the pre-school age, Sim et al. (28) reported a mean sensitivity of 0.66 (range, 0.52–0.80).

In order to improve the SLD-prediction for children who failed screening stage 1, again logistic regression was used but—in addition to the significant predictors of stage 1—the word comprehension subtest of the SETK-2 was entered as additional predictor into the model. Thus, we evaluated whether the direct pediatric assessment significantly contributes to the prediction of SLD at the age of 3 years for children who failed screening at stage 1. In the case of an incremental contribution of the pediatric assessment to the SLD-prediction, a new (stage 2) screening score (probability of having SLD) for children failing the stage 1 screening was computed.

Subsequently, the diagnostic accuracy of the 2-stage screening procedure (i.e., stage 1 score for screening passes at stage 1 and stage 2 score for screening fails at stage 1) was evaluated and compared with simple stage 1 screening by applying ROC (receiver operating characteristics) analyses and AUC difference tests for paired ROC curves. AUCs $\geq$0.9 are regarded as excellent, AUCs $\geq$0.8 and $<$0.9 as good, AUCs $\geq$ 0.7 and $<$0.8 as fair, and tests with AUCs $<$ 0.7 as poor (50).

Finally, to provide information about graduations of risk for SLD, cutoffs for the high-risk and moderate-risk groups were evaluated by estimating the following diagnostic efficiency statistics: sensitivity (Se), specificity (Sp), positive predictive values (PPV), negative predictive values (NPV), and diagnostic likelihood ratios for positive and negative screening results (DLR+ and DLR-). DLRs are alternative measures of diagnostic accuracy and display the multiplicative change in the pre-screening odds of having SLD given a positive (DLR+) or negative screening result (DLR–). DLR+ values $\geq$10 and DLR– $\leq$0.1 indicate large changes in pre-screening odds. DLR+ $\leq$10 and $>$5 and DLR– $>$0.1 and $\leq$0.2 indicate moderate changes; DLR+ $\leq$5 and $>$2 and DLR– $>$0.2 and $\leq$0.5 indicate small changes. DLR+ $<$2 and DLR– $>$0.5 are rarely important (51). To further evaluate any advantages of the 2-stage screening procedure, cutoffs were estimated for the simple stage 1 screening and the 2-stage screening, and subsequently compared in terms of their diagnostic efficiency. The cutoff for the high-risk group was determined by fixing PPV at 0.80; that is, children who failed to meet this cutoff had a probability of 80% of having SLD about 1 year after the screening. The PPV of 0.80 was selected for pragmatic reasons taking the limited availability of follow-up diagnostic services and the high probability of need for early intervention into account. For identifying a moderate-risk group, various cutoffs corresponding to sensitivities of 0.75, 0.80, and 0.85 were selected. Notably, as a sensitivity of 0.90 was chosen at stage 1, cutoffs with high sensitivity values (near at least 0.90) in the 2-stage screening process would, by design, hardly differ from stage 1 screening. Figure 1 summarizes the conceptual design of the 2-stage screening process described above.

Logistic regressions and descriptive analyses were performed using SPSS 27 (28, 52). Furthermore, the pROC package in R (53) was used to perform ROC analysis and tests for paired ROC curves. Additionally, the R-OptimalCutpoints package (54) was used to estimate cutoffs and diagnostic efficiency statistics.

Results

Before presenting the results of the two-stage screening, we compared the diagnostic accuracy of the 37 item word list with the initial 260 word list with regard to the SLD measure used in this study. Diagnostic accuracy (AUC) was 0.866 [DeLong 95%CI (0.827–0.905)] for the 260 item list and 0.857 [DeLong

95%CI (0.814–0.900)] for the 47 item list. As indicated by a bootstrapped test for comparing paired ROC curves, the difference was not significant (ΔAUC = 0.009, D = 0.721, p = 0.471).

Stage 1 screening

The results of the logistic regression to identify significant parent-reported predictors of SLD at the age of 3 years are shown in Table 2. In the first results column, the estimates for the full model are reported, i.e., the model in which all variables were entered simultaneously as predictors. The second column (Final Model) shows the estimates after excluding non-significant predictors (logistic regression-based backward elimination). Notably, as in the full model, only parental concerns [b = 1.231, $p < 0.001$, odds ratio (OR) = 3.424 95% CI (1.713, 6.843)], no two-word combinations [b = 1.268, $p < 0.01$, OR = 3.554 95% CI (1.409, 8.963)], and expressive vocabulary [b = −0.088, $p < 0.001$, OR = 0.916 95% CI (0.888, 0.944)] significantly predicted SLD. Based on these results, a stage 1 screening score (i.e. probability of SLD) was calculated.

To identify children that should also undergo screening stage 2, a cutoff was determined by fixing sensitivity at 0.90. The cutoff of 0.05 on a probability scale met this requirement and yielded the following accuracy statistics: Se = 0.900 [95% CI (0.805, 0.959)], Sp = 0.708 [95% CI (0.668, 0.745)], PPV = 0.281 [95% CI (0.668, 0.745)], NPV = 0.982 [95% CI (0.962, 0.985)], DLR+ = 3.080 [95% CI (2.646, 3.584)], and DLR– = 0.141 [95% CI (0.070, 0.286)]. Notably, as expected, the high sensitivity of 0.90 was at the expense of a quite low PPV. Given this cutoff, 224 (36%) children failed the stage 1 of the screening process and thus, are considered in stage 2 screening.

Stage 2 screening

To evaluate whether the pediatric assessment incrementally contributes to the prediction of SLD in children failing stage 1 (N = 224), a logistic regression model including all selected predictors of stage 1 and additionally the pediatric scale of word comprehension, was used. The results are reported in Table 3.

All three parent-reported stage 1 predictors and also the pediatric assessed word comprehension significantly contributed to the predicting of SLD. Notably, as indicated by the standardized coefficients, word comprehension had the greatest predictive power (stand. b = −0.590, vs. stand. b = 0.422 for two-word combinations, stand. b = 0.343 for parental concerns and stand. b = −0.365 for expressive vocabulary). Like in stage 1, the predicted probability of having SLD was calculated based on the regression results. Subsequently, a *total two-stage screening score* was computed. This score equals the stage 1 score for children who passed stage 1. For children who failed the screening at stage 1, the predicted probability of stage 2 was used.

Comparing diagnostic accuracies between the total two-stage screening and one-stage screening tool

ROC analyses yielded an AUC of 0.875 [DeLong 95% CI (0.833–0.917)] for the one-stage screening (i.e., screening score of stage 1), and a slightly better AUC of 0.885 [DeLong 95% CI (0.843–0.926)] for the total two-stage screening. A bootstrapped test for comparing paired ROC curves shows that the AUC difference (ΔAUC = 0.01) was marginally significant (D = 1.718, p = 0.086).

Cutoff estimation

Finally, various cutoffs were estimated, and the respective diagnostic efficiencies for the total two-stage screening tool and one-stage screening tool were evaluated (see Table 4). To identify a high-risk group that is characterized by a high probability of having a SLD, the cutoffs were estimated by fixing the PPV at 0.8. For the one-stage and two-stage screening tools, a cutoff of 0.758 (2.4% screening failures) and 0.594 (4% screening fails), respectively, resulted in a PPV of 0.80. Given the same PPV, the sensitivity of the total two-stage screening [0.286, 95% CI (0.092, 0.280)]tool was significantly higher than that of the one-stage screening tool [0.171, 95% CI (0.183, 0.406)]; McNemar test χ^2 (1) = 6.125, p = 0.008. Thus, the total two-stage screening tool identifies a larger proportion of children with SLD than the one-stage screening tool.

Subsequently, an additional moderate-risk group, which together with the high-risk group should cover 75% of the children with SLD (i.e., sensitivity = 0.75), was identified. Cutoff values that achieved a sensitivity closest to 0.75 were 0.089 (24.6% screening failures; Se = 0.729, 95%CI (0.609, 0.828) for the one-stage screening tool and 0.169 (20.5% screening failures; Se = 0.743, 95% CI (0.624, 0.840) for the two-stage screening tool. Notably, the total two-stage screening tool [Sp = 0.864, 95% CI (0.832, 0.891)] was more specific than the one-stage screening (Sp = 0.815, 95% CI (0.780, 0.846); McNemar Test χ^2 (1) = 11.860, $p < 0.001$.

Further cutoffs for the moderate risk group associated with higher sensitivity were also evaluated (see Table 4). Overall, the differences between the one-stage and two-stage screening tools decreased for higher sensitivity values.

Figure 2 summarizes the results of the proposed two-stage screening tool. At stage 1, the high sensitivity of 0.90 was accompanied by quite a high rate of screening failures (36.1%) and consequently with a rather large share of false positives

TABLE 2 Logistic regression predicting LD at age 3 – Stage 1.

	Full model			Final model		
	b	SE	OR (95%CI)	b	SE	OR (95% CI)
Parental highest education (reference = university)[a]						
Compulsory education	1.782*	0.895	5.940 (1.029, 34.304)			
Vocational education	0.325	0.425	1.384 (0.602, 3.182)			
University entrance	0.164	0.480	1.178 (0.460, 3.018)			
Child gender (1 = male)	0.016	0.317	1.016 (0.546, 1.892)			
Child age	0.211	0.160	1.235 (0.903, 1.690)			
Parental concerns	1.176***	0.366	3.243 (1.583, 6.645)	1.231***	0.353	3.424 (1.713, 6.843)
No two-word utterances	1.268**	0.477	3.552 (1.394, 9.055)	1.268**	0.472	3,554 (1,409, 8.963)
Expressive vocabulary	−0.091***	0.016	0.913 (0.884,0.943)	−0.088***	0.016	0.916 (0.888,0.944)
Intercept	−6.485	3.834		−1.217	0.777	
R^2 Nagelkerke	0.365			0.351		

[a]Overall p-value based on Wald test > 0.05.
The *, **, and *** symbols indicate the value of p < 0.05, p < 0.01, and p < 0.001 respectively.

(PPV = 0.281). After stage 2, a small group of 4% is classified as high-risk for SLD. Twenty out of these 25 children are classified as having SLD at the age of 3 years (PPV = 0.80). Notably, their pre-screening odds (before stage 1 screening) of having an SLD increases by about 31 times (DLR+ = 31.486), given a screening failure (high risk) at stage 2. However, the high-risk group only accounts for 20 out of 70 children with SLD at the age of 3 years. Thus, the high-risk cutoff is associated with a low sensitivity (28.6). The moderate risk group (16.5%) covers further 32 children with SLD at the age of 3 years (sensitivity = 0.457). The change in the pre-screening odds of having SLD for children in the moderate-risk group (DLR+ = 3.360) is minimal. The true positives from the high-risk and moderate-risk groups sum up to 52 (Sensitivity = 0.743) children. Finally, 476 out of 494 children with no SLD at the age of 3 years are correctly identified by the two-stage screening tool (specificity = 0.864).

Supplementary analyses

Finally, to validate the differentiation between a low, moderate and high-risk group, we performed a set of supplementary analyses. In detail, we compared the means of the two reference test scales (expressive vocabulary and expressive grammar; see measures section), between the three groups.

We found that the high-risk group [expressive vocabulary: $M = -1.73$, standard deviation (SD) = 0.945; expressive grammar: $M = -1.99$, SD = 1.020] scored about one SD lower than the moderate-risk group (expressive vocabulary: $M = -0.791$, SD = 1.030; expressive grammar: $M = -1.016$, SD = 1.193) and about two SDs lower than the remaining low-risk group (expressive vocabulary: $M = 0.255$, SD = 0.826; expressive grammar: $M = 0.260$, SD = 0.804). Thus, all groups substantially differed in their means [Expressive vocabulary: $F_{(2, 52.671)} = 88.344$, $p < 0.001$; $\eta^2 = 0.262$].

TABLE 3 Logistic regression predicting LD at age 3 – Stage 2.

	b	SE	OR (95% CI)
Parental concerns	0.721*	0.352	2.056 (1.031, 4.100)
No two-word utterances	1.159**	0.448	3.185 (1.323, 7.667)
Expressive vocabulary	−0.038*	0.018	0.962 (0.928, 0.997)
Word comprehension	−0.284***	0.084	0.753 (0.639, 0.887)
Intercept	−0.069	0.882	
R^2 Nagelkerke	0.282		

The *, **, and *** symbols indicate the value of p < 0.05, p < 0.01, and p < 0.001 respectively.

Feasibility

Twenty-three out of the 30 pediatricians (77%) who participated in the study returned completed questionnaires on feasibility of SPES-2 language screening.

Practicality

Ease of administration of the word comprehension screening was rated as "very good" and "good" by 52 and 48%, respectively, of the pediatricians. Thus, none of the pediatricians rated ease of administration as "difficult." Ease of integrating the screening within the time constraints of regular pediatric care was mostly rated as "good" (65%), as "very good" in 26%, and only one pediatrician regarded it as "difficult." Among

TABLE 4 Diagnostic accuracy statistics for different cutoffs (95% CIs).

	Cutoff	%Fails	Sens	Spec	PPV	NPV	DLR+	DLR-
High Risk								
Stage 1 screening	0.758	2.4%	0.171	0.995	0.800	0.904	31.486	0.833
			(0.092, 0.280)	(0.984, 0.999)	(0.577, 0.883)	(0.822, 0.979)	(9.107, 108.853)	(0.749, 0.927)
Total 2-stage screening	0.594	4.0%	0.286	0.991	0.800	0.916	31.486	0.721
			(0.184, 0.406)	(0.979, 0.997)	(0.630, 0.872)	(0.860, 0.971)	(12.202, 81.242)	(0.621, 0.836)
Moderate Risk & High Risk (Se = 0.75)								
Stage 1 screening	0.089	24.6%	0.729	0.815	0.333	0.959	3.936	0.333
			(0.609, 0.828)	(0.780, 0.846)	(0.287, 0.473)	(0.932, 0.967)	(3.139, 4.934)	(0.226, 0.490)
Total 2-stage screening	0.169	20.5%	0.743	0.864	0.409	0.964	5.458	0.298
			(0.624, 0.840)	(0.832, 0.891)	(0.352, 0.557)	(0.938, 0.972)	(4.244, 7.018)	(0.200, 0.444)
Moderate Risk & High Risk (Se = 0.80)								
Stage 1	0.076	28.0%	0.788	0.797	0.329	0.967	3.865	0.269
			(0.671, 0.875)	(0.761, 0.830)	(0.285, 0.483)	(0.942, 0.973)	(3.147, 4.748)	(0.171, 0.422)
Total 2-stage screening	0.128	26.7%	0.800	0.786	0.322	0.969	3.736	0.255
			(0.687, 0.886)	(0.749, 0.819)	(0.279, 0.480)	(0.944, 0.975)	(3.064, 4.555)	(0.159, 0.407)
Moderate Risk & High Risk (Se = 0.85)								
Stage 1	0.059	30.6%	0.843	0.762	0.311	0.974	3.545	0.206
			(0.736, 0.919)	(0.724, 0.797)	(0.270, 0.487)	(0.952, 0.979)	(2.960, 4.246)	(0.120, 0.355)
Total 2-stage screening	0.122	29.1%	0.843	0.779	0.326	0.975	3.807	0.202
			(0.736, 0.919)	(0.742, 0.813)	(0.283, 0.505)	(0.953, 0.980)	(3.159, 4.587)	(0.117, 0.348)

the five pre-specified factors that might possibly challenge the administration of language screening within primary pediatric care, lack of or insufficient follow-up was ranked highest (40%). Insufficient training of pediatricians and lack of or insufficient funding were equally regarded as most challenging by 20%, followed by limited meaningfulness of language screening (13%) and time constraints (4%).

Acceptability

Parental acceptance of language screening included within the regular well-baby check-up at the age of 2 years was rated as either "very good" (57%) or "good" (43%) by the pediatricians. The acceptance of the questionnaire by parents was rated as "good" by the majority (52%), as "very good" by 35%, and as "difficult" by 13% of pediatricians. Most (77%) pediatricians rated the meaningfulness of language screening within the regular preventive check-ups as "very good," 14% as" good," and only 9% expressed concerns.

Sustainability

Most of the pediatricians (83%) indicated that they would continue the SPES-2 language screening beyond the research project, whereas 17% would not. Notably, no additional funding for the language screening study was provided by the public health system.

Discussion

This study evaluated the screening accuracy and feasibility of a newly developed language screening tool (SPES-2) for 2-year-old children in a large population sample within regular well-baby check-ups in primary pediatric care in Upper Austria. The two-stage screening tool included parent-reported information on the child's expressive vocabulary, production of two-word-combinations, and parental concerns about language development during stage 1. During stage 2, results of the pediatrician's assessment of word comprehension demonstrated good diagnostic accuracy ((AUC = 0.885) for the prediction of SLD about 1 year after the screening (for children at the age of 3 years).

The two-stage screening tool identified a significantly larger proportion of children ending up with SLD at the age of 3 years than the first screening stage (parent reports) only. Defining a high-risk group by a probability of 80% of ending up with SLD at age of 3 years, 4% of the study sample were identified, representing 28.6% of the children with SLD at the age of 3 years.

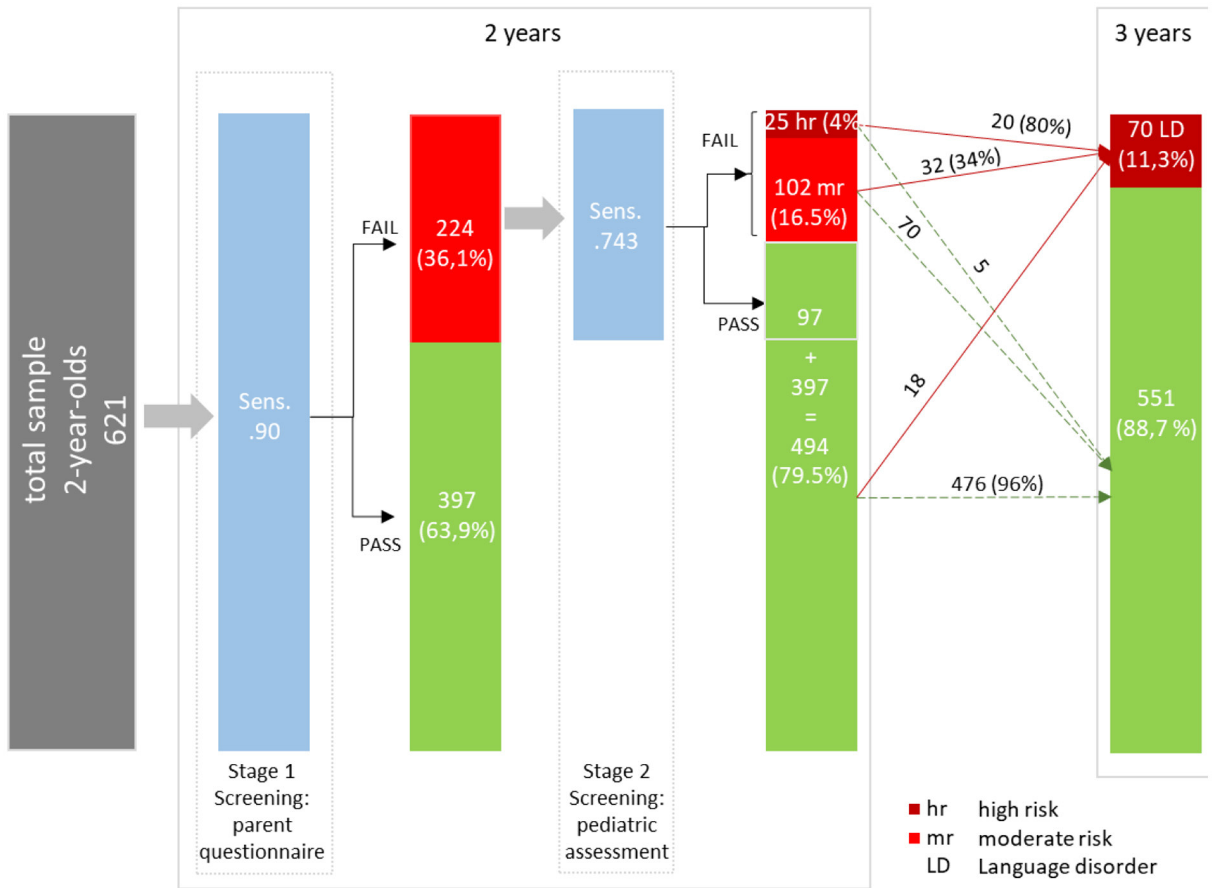


FIGURE 2
Overview of results of the two-stage screening tool.

Another 16.5% of the total sample were identified as a group with moderate risk for SLD at the age of 3 years (risk increased three-fold). The moderate-risk group comprised 45.7% of the children with SLD 1 year later. Thus, both risk groups (high and moderate: 20.5% of all children) included about 75% of all children diagnosed with SLD at the age of 3 years (sensitivity: 0.743, and specificity: 0.864). The classification of screening results by degree of risk (low-moderate-high) was also supported by significant differences (of about 1 standard deviation) among the three groups.

Our findings of high predictive validity of the screening tool after 1 year confirm those of the few language screening tools in 2-year-old children that predict later language status. The Dutch well-child language screening protocol for 2-year-old children (29), which is based on direct-child assessment of word comprehension and parent reports (word combinations and playing behavior), yielded a slightly higher sensitivity (0.82) but lower specificity (0.74) to predict language problems 1 year later. However, about half of the validation sample were screening failures, and referral bias artificially increases sensitivity. Predictive sensitivity and specificity reported for the German MCDI-based ELFRA (34) (parent questionnaire for 2-year-old children) were 0.61 and 0.94, respectively (55). However, generalizability to the total population is again limited due to a high overrepresentation of screening failures (60%) in the study sample. For the Language Development Survey (56), predictive validity was similar to that found by the ELFRA study with sensitivity of 0.67 and specificity of 0.96. However, only 15.6% of the LDS sample were included in the follow-up assessment that was performed at an average of only 23 days after the screening. Given the high variability of trajectories of early language development, a short time lag between screening and follow-up certainly contributes to its predictive validity. The screening battery approach by Stott et al. (57) applied the General Language Screen and Developmental Profile II at the age of 36 months to predict speech and language disorders at 45 months of age. Both, sensitivity (0.67) and specificity (0.68) were significantly lower than those achieved by the SPES2 measure.

Thus far, predictive validity of parental reports of expressive vocabulary, two-word-combinations, parental concerns, and language comprehension aligns with the international evidence (55, 56).

The graduation of risk (high vs. moderate) of screening failures allows for tailored follow-up and intervention, avoiding a high rate of over-referrals related to the lower specificity of SPES-2 than MCDI and LDS. Significant differences of results of standardized language tests between the group with no and moderate risk (about−1SD) and between the moderate and high-risk groups (about−1SD) confirm the validity of the three risk levels requiring different follow-up procedures. For the high-risk group, follow-up assessments of language, hearing, cognitive, and psycho-social development should be considered, because the combination of expressive and receptive language difficulties can be indicative of more comprehensive or pervasive developmental disorders, such as general developmental delay or autism spectrum disorders. For the group with moderate but still significantly increased risk of SLD, preventive parent counseling or parent training (for example, promoting a responsive interaction style with their children and facilitative language techniques within everyday family routines or dialogic book reading) seems to be indicated. Due to the instability of early language development between the ages of 2 and 3 years, continuous language surveillance (for example, language screening at the age of 3 years) should be recommended to *all* parents including those of children in the low-risk group. The severity and type of language disorders and the prevalence of other neurodevelopmental disorders in the subgroups of children with high, moderate, or low risk still need to be determined to further substantiate the risk levels resulting from the language screening.

The high feasibility of the new screening tool within primary pediatric care was not anticipated. The two-stage procedure, including an initial collection of risk indicators from parents followed by a very short and easy-to-score direct assessment of word comprehension in only about one-third of the children, undoubtedly contributes to the high practicality. For the majority of pediatricians, the implementation of language screenings within regular well-baby check-ups appeared appropriate. However, factors possibly challenging the implementation of language screenings, particularly the provision of follow-up diagnostic assessment and parental guidance, specific training of pediatricians, and funding of the screening procedures, need to be considered in the planning of a population-based realization.

Our study had some major strengths. First, the screening was designed to predict language disorders 1 year after its implementation. Second, the non-selected total population sample of 2-year-old children were assessed within regular pediatric well-baby check-ups. Third, all the children, regardless of their screening results, were followed up by use of validated language measures at the age of 3 years, which highly restricting potential bias in our study. Fourth, the inclusion of feasibility in the evaluation of the screening is, as an essential contribution to the field (19), with a demand of information about practicality and acceptability. However, the lack of multidimensional standardized follow-up diagnostic assessments on a random selection of children with high, moderate, and low risk of SLD can be considered a limitation of this study. Another limitation was the voluntary participation of pediatricians that might have positively influenced feasibility. Exclusion of bilingual children was only based on parent information excluding children with a non-German language peferrably used in their families Therefore, children growing up with another language in addition to the primary German family language were included in the study samples. The use of a second language might have an influence on their German language development that is no taken into account in the current study.

Conclusion

Our findings on predictive sensitivity and specificity of the language screening tool, SPES-2, demonstrate its validity for the early identification of SLD in 2-year-old children that persist to the age of 3 years. The two-stage procedure of parental report followed by direct pediatric assessment only in those who failed the first screening stage makes the screening time-efficient. The grading of risk levels derived from screening results supports tailored follow-up and requires further clinical validation.

Data availability statement

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

Ethics statement

The studies involving human participants were reviewed and approved by Ethikkommission Konventhospital Barmherzige Brüder Linz. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Author contributions

DH and JF: conceptualization and investigation. DH, CW, and JF: methodology, original draft preparation, review, and editing. CW: formal analysis. DH: data curation and project administration. All authors contributed to the article and approved the submitted version.

Funding

This study was financially supported by the Department of Health and Social Affairs of the Upper Austrian government. Article Processing Charge is funded by the Johannes Kepler University Open Access Publishing Fund.

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/fped.2022.865457/full#supplementary-material

References

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

2. Tomblin JB, Records NL, Buckwalter P, Zhang X, Smith E, O'Brien M. Prevalence of specific language impairment in kindergarten children. *J Speech Lang Hear Res.* (1997) 40:1245–60. doi: 10.1044/jslhr.4006.1245

3. Snowling MJ, Bishop DVM, Stothard SE, Chipchase B, Kaplan C. Psychosocial outcomes at 15 years of children with a preschool history of speech-language impairment. *J Child Psychol Psychiatry.* (2006) 47:759–65. doi: 10.1111/j.1469-7610.2006.01631.x

4. Lindsay G, Dockrell JE, Strand S. Longitudinal patterns of behaviour problems in children with specific speech and language difficulties: child and contextual factors. *Br J Educ Psychol.* (2007) 77:811–28. doi: 10.1348/000709906X171127

5. Schoon I, Parsons S, Rush R, Law J. Children's language ability and psychosocial development: a 29-year follow-up study. *Pediatrics.* (2010) 126:e73–80. doi: 10.1542/peds.2009-3282

6. Catts HW, Fey ME, Tomblin JB, Zhang X. A longitudinal investigation of reading outcomes in children with language impairments. *J Speech Lang Hear Res.* (2002) 45:1142–57. doi: 10.1044/1092-4388(2002/093)

7. Nauclér K, Magnusson E. Reading and writing development: report from an ongoing longitudinal study of language-disordered and normal groups from pre-school to adolescence. *Folia Phoniatr Logop.* (1998) 50:271–82. doi: 10.1159/000021469

8. Law J, Rush R, Schoon I, Parsons S. Modeling developmental language difficulties from school entry into adulthood: literacy, mental health, and employment outcomes. *J Speech Lang Hear Res.* (2009) 52:1401–16. doi: 10.1044/1092-4388(2009/08-0142)

9. Rosenbaum S, Simon P, editors. *Speech and Language Disorders in Children: Implications for the Social Security Administration's Supplemental Security Income Program*. Washington, DC: National Academies Press (2016).

10. Buschmann A, Jooss B, Rupp A, Feldhusen F, Pietz J, Philippi H. Parent based language intervention for 2-year-old children with specific expressive language delay: a randomised controlled trial. *Arch Dis Child.* (2009) 94:110–6. doi: 10.1136/adc.2008.141572

11. Roberts MY, Kaiser AP. The effectiveness of parent-implemented language interventions: a meta-analysis. *Am J Speech Lang Pathol.* (2011) 20:180–99. doi: 10.1044/1058-0360(2011/10-0055)

12. Roberts MY, Kaiser AP. Early intervention for toddlers with language delays: a randomized controlled trial. *Pediatrics.* (2015) 135:686–93. doi: 10.1542/peds.2014-2134

13. Baxendale J, Hesketh A. Comparison of the effectiveness of the Hanen Parent Programme and traditional clinic therapy. *Int J Lang Commun Disord.* (2003) 38:397–415. doi: 10.1080/1368282031000121651

14. Broomfield J, Dodd B. Is speech and language therapy effective for children with primary speech and language impairment? Report of a randomized control trial. *Int J Lang Commun Disord.* (2011) 46:628–40. doi: 10.1111/j.1460-6984.2011.00039.x

15. Rinaldi S, Caselli MC, Cofelice V, D'Amico S, Cagno AG de, Della Corte G et al. Efficacy of the treatment of developmental language disorder: a systematic review. *Brain Sci.* (2021) 11:407. doi: 10.3390/brainsci11030407

16. Lipkin PH, Macias MM. Promoting optimal development: identifying infants and young children with developmental disorders through developmental surveillance and screening. Pediatrics. (2020) 145:e20193449. doi: 10.1542/peds.2019-3449

17. Developmental Behavioral Pediatrics B & Medical Home Initiatives for Children With Special Needs Project Advisory Committee. Identifying infants and young children with developmental disorders in the medical home: an algorithm for developmental surveillance and screening. *Pediatrics.* (2006) 118:405–20. doi: 10.1542/peds.2006-1231

18. Nelson HD, Nygren P, Walker M, Panoscha R. Screening for speech and language delay in preschool children: systematic evidence review for the US preventive services task force. *Pediatrics.* (2006) 117:e298–319. doi: 10.1542/peds.2005-1467

19. Robins DL, Casagrande K, Barton M, Chen CMA, Dumont-Mathieu T, Fein D. Validation of the modified checklist for autism in toddlers, revised with follow-up (M-CHAT-R/F). *Pediatrics.* (2014) 133:37–45. doi: 10.1542/peds.2013-1813

20. Siu AL. Screening for speech and language delay and disorders in children aged 5 years or younger: US preventive services task force recommendation statement. *Pediatrics.* (2015) 136:e474–81. doi: 10.1542/peds.2015-1711

21. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen IQWiG. *Früherkennungsuntersuchung auf Entwicklungsstörung des Sprechens und der Sprache bei Kindern.* (2009). Available online at: https://www.iqwig.de/download/s06-01_abschlussbericht_frueherkennung_umschriebener_stoerungen_des_sprechens_und_der_sprache.pdf (accessed January 17, 2022).

22. Sachse S. *Neuropsychologische und neurophysiologische Untersuchungen bei late talkers im Quer-und Längsschnitt* (Dissertation). München: Ludwig-Maximilians-Universität (2007).

23. Bishop DV, Holt G, Line E, McDonald D, McDonald S, Watt H. 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 Dis.* (2012) 4:1–12. doi: 10.1186/1866-1955-4-3

24. Rescorla L. Late talking toddlers: a 15-year follow-up. In: Rescorla L, Philip D, editors. *Late Talkers: Language Development, Interventions, and Outcomes.* Baltimore: Brookes (2013). p. 219–39.

25. Dale PS, Price TS, Bishop DVM, Plomin R. Outcomes of early language delay. *J Speech Lang Hear Res.* (2003) 46:544–60. doi: 10.1044/1092-4388(2003/044)

26. Poll GH, Miller CA. Late talking, typical talking, and weak language skills at middle childhood. *Learn Individ Differ.* (2013) 26:177–84. doi: 10.1016/j.lindif.2013.01.008

27. Reilly S, Cook F, Bavin EL, Bretherton L, Cahir P, Eadie P, et al. Cohort profile: The Early Language in Victoria Study (ELVS). *Int J Epidemiol.* (2018) 47:11–20. doi: 10.1093/ije/dyx079

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

29. Visser-Bochane M, Luinge M, Dieleman L, van der Schans C, Reijneveld S. The Dutch well child language screening protocol for 2-year-old children was valid for detecting current and later language problems. *Acta Paediatr.* (2021) 110:556–62. doi: 10.1111/apa.15447

30. Voigt RG, Llorente AM, Jensen CL, Fraley JK, Barbaresi WJ, Heird WC. Comparison of the validity of direct pediatric developmental evaluation versus developmental screening by parent report. *Clin Pediatr.* (2007) 46:523–9. doi: 10.1177/0009922806299100

31. Voigt RG, Accardo PJ. Formal speech-language screening not shown to help children. *Pediatrics.* (2015) 136:e494–5. doi: 10.1542/peds.2015-0211

32. Fenson L. *MacArthur-Bates Communicative Development Inventories*. Brookes Publishing Company. (2007).

33. Rescorla L. The language development survey: a screening tool for delayed language in toddlers. *J Speech Hear Disord.* (1989) 54:587–99. doi: 10.1044/jshd.5404.587

34. Grimm H, Doil H. *ELFRA–Elternfragebögen für die Früherkennung von Risikokindern*. Hogrefe: 2 überarbeitete Auflage Göttingen (2006).

35. Sachse S, Pecha A, von Suchodoletz W. Früherkennung von sprachentwicklungsstörungen. *Monatsschr Kinderheilkd.* (2007) 155:140–5. doi: 10.1007/s00112-006-1314-7

36. Sachse S, Suchodoletz W. Variabilität expressiver sprachleistungen bei zweijährigen kindern erfasst mit dem ELFRA-2. *Sprache Stimme Gehör.* (2007) 31:118–25. doi: 10.1055/s-2007-982528

37. Von Suchodoletz W. Früherkennung von umschriebenen sprachentwicklungsstörungen. *Z Kinder Jugendpsychiatr Psychother.* (2011) 39:377–84. doi: 10.1024/1422-4917/a000136

38. Szagun G, Stumper B, Schramm SA. *FRAKIS: Fragebogen zur frühkindlichen Sprachentwicklung (FRAKIS) und FRAKIS-K (Kurzform). Pearson* (2009).

39. Kasper J, Kreis J, Scheibler F, Möller D, Skipka G, Lange S et al. Population-based screening of children for specific speech and language impairment in Germany: a systematic review. *Folia Phoniatr Logop.* (2011) 63:247–63. doi: 10.1159/000321000

40. Holzinger D, Fellinger J, Beitel C. *Projekt SPES (Sprachentwicklungsscreening Oberösterreich): ENDBERICHT. Linz.* (2008). Available online at: https://www.researchgate.net/publication/361663461_Projekt_SPES_Sprachentwicklungsscreening_Oberosterreich_ENDBERICHT (accessed July 1, 2022).

41. Grimm H. *Sprachentwicklungstest für drei- bis fünfjährige Kinder: SETK 3-5; Diagnose von Sprachverarbeitungsfähigkeiten und auditiven Gedächtnisleistungen*. Hogrefe: Verlag für Psyhologie (2001).

42. Kiese-Himmel C. *AWST-R-Aktiver Wortschatztest für 3-bis 5-jährige Kinder*. Göttingen: Beltz Test (2005).

43. Statistik Austria. *Lebendgeborene 2020 nach medizinischen Merkmalen.* (2021). Available online at: https://statistik.at/fileadmin/pages/427/soz_med_M_Geborene_Mappe2020.ods (accessed July 1, 2022).

44. Bundesinstitut BIFIE. *Standardüberprüfung 2018: Mathematik, 4. Schulstufe*. BIFIE (2019).

45. Oberwimmer K, Baumegger D, Vogtenhuber S. Kontext des schul- und bildungswesens. In: Oberwimmer K, Vogtenhuber S, Lassnigg K, Schreiner C, editors. *Nationaler Bildungsbericht Österreich 2018: Das Schulsystem im Spiegel von Daten und Indikatoren*. Graz: Leykam (2019).

46. Grimm H, Aktaş M. *SETK-2: Sprachentwicklungstest für zweijährige Kinder: Diagnose Rezeptiver Und Produktiver Sprachverarbeitungsfähigkeiten: Manual*. Hogrefe: Verlag für Psychologie (2000).

47. Holzinger D, Weber C, Barbaresi W, Beitel C, Fellinger J. Language screening in 3-year-olds: development and validation of a feasible and effective instrument for pediatric primary care. *Front Pediatr.* (2021) 9:752141. doi: 10.3389/fped.2021.752141

48. Bowen DJ, Kreuter M, Spring B, Cofta-Woerpel L, Linnan L, Weiner D, et al. How we design feasibility studies. *Am J Prev Med.* (2009) 36:452–7. doi: 10.1016/j.amepre.2009.02.002

49. Kilgus SP, Eklund KR. *Consideration of Base Rates Within Universal Screening for Behavioral and Emotional Risk: A Novel Procedural Framework*. School Psychology Forum (2016).

50. Swets JA. Measuring the accuracy of diagnostic systems. *Science.* (1988) 240:1285–93. doi: 10.1126/science.3287615

51. Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature. III How to use an article about a diagnostic test B. What are the results and will they help me in caring for my patients? The evidence-based medicine working group. *JAMA.* (1994) 271:703–7. doi: 10.1001/jama.271.9.703

52. IBM Corp. Released (2020). *IBM SPSS Statistics for Windows. Version 27.0*. Armonk, NY: IBM Corp61.

53. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics.* (2011) 12:77. doi: 10.1186/1471-2105-12-77

54. López-Ratón M, Rodríguez-Álvarez MX, Suárez CC, Sampedro FG. Optimalcutpoints : an R package for selecting optimal cutpoints in diagnostic tests. *J Stat Soft*. (2014) 61:i08. doi: 10.18637/jss.v061.i08

55. Sachse S, von Suchodoletz W. Early identification of language delay by direct language assessment or parent report? *J Dev Behav Pediatr.* (2008) 29:34–41. doi: 10.1097/DBP.0b013e318146902a

56. Rescorla L, Alley A. Validation of the language development survey (LDS): a parent report tool for identifying language delay in toddlers. *J Speech Lang Hear Res.* (2001) 44:434–45. doi: 10.1044/1092-4388(2001/035)

57. Stott CM, Merricks MJ, Bolton PF, Goodyer IM. (2002). Screening for speech and language disorders: The reliability, validity and accuracy of the General Language Screen. *Int J Lang Commun Disord.* 37:133–51. doi: 10.1080/13682820110116785

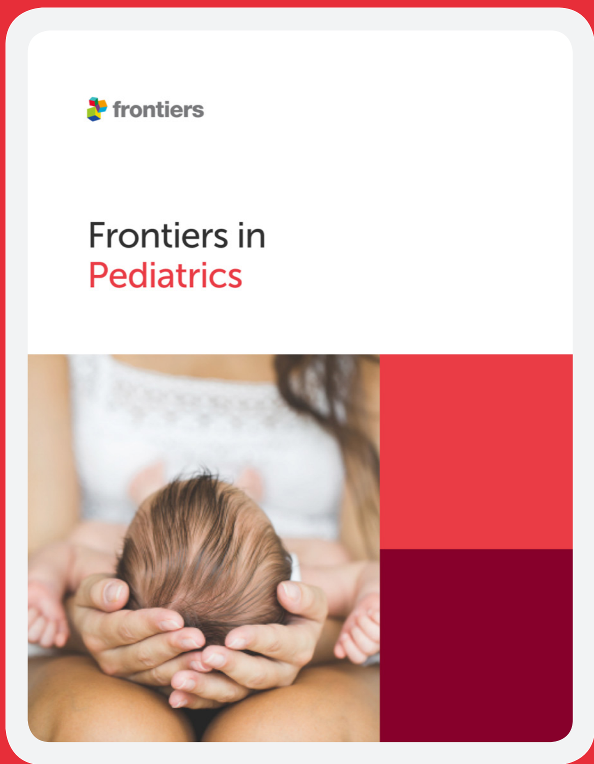
Frontiers in Pediatrics

Addresses ongoing challenges in child health and patient care

Explores research that meets ongoing challenges in pediatric patient care and child health, from neonatal screening to adolescent development.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact