

Manuscript accepted for publication in Early Human Development (2021).
<https://doi.org/10.1016/j.earlhumdev.2021.105345>
[https://authors.elsevier.com/sd/article/S0378-3782\(21\)00044-X](https://authors.elsevier.com/sd/article/S0378-3782(21)00044-X)

EARLY RECEPTIVE AND EXPRESSIVE LEXICONS AND LANGUAGE AND PRE-
LITERACY SKILLS AT 5;0 YEARS

– A longitudinal study

SUVI-MARIA VEHKAVUORI^{a, b, c*},

MAIJU KÄMÄRÄINEN^d

AND

SUVI STOLT^b

^aFaculty of Humanities, Logopedics, Child Language Research Center, University of Oulu,
Finland

^bFaculty of Medicine, Department of Psychology and Logopedics, Unit of Logopedics,
University of Helsinki, Finland

^cCoronaria Oy Ltd, Therapy Center, Turku, Finland

^dCogniMed Oy Ltd, Helsinki, Finland

*Corresponding author at: Faculty of Humanities, Logopedics, Child Language Research
Center, P.O. Box 8000, 90014 University of Oulu, Finland; Coronaria Oy Ltd, Therapy
Center, Turku, Finland.

Email-addresses: suvi-maria.vehkavuori@student oulu.fi; suvi.vehkavuori@helsinki.fi (S.-M.
Vehkavuori), maiju.susanna@hotmail.fi (M. Kämäräinen), suvi.stolt@helsinki.fi (S. Stolt)

Abstract

Background: The long-term associations between early receptive/expressive lexical skills and later language/pre-literacy skills require clarification.

Aims: To study the association between and predictive values of early receptive/expressive lexical skills and language/pre-literacy skills at 5;0 years, and to examine the language profiles at 5;0 years of children with weak receptive language/expressive lexical skills at 2;0 years.

Participants and methods: The participants were 66 monolingual children. Their lexical skills were measured using the Finnish short-form version of the MacArthur–Bates Communicative Development Inventories at 1;6 and 2;0 years. Receptive language skills were measured at 2;0 years using the Reynell Developmental Language Scales III. A broader assessment at 5;0 years measured lexical, phonological, morphological and pre-literacy skills.

Results: Significant associations between receptive/expressive lexical skills at 1;6 years and language and pre-literacy skills at 5;0 years were found. Both receptive language and expressive lexical development measured at 2;0 years were greatly and relatively evenly associated with language and pre-literacy skills at 5;0 years. Lexicon/language variables at 1;6 years and 2;0 years had statistically significant predictive values for general language and pre-literacy scores at 5;0 years. The best models that included early lexical predictors explained 20–34% of later language/literacy outcome. Weak skills at 2;0 years proposed vulnerability in language and pre-literacy skills at 5;0 years.

Conclusions: Language and pre-literacy skills at 5;0 years can to some extent be explained by early receptive language and/or expressive lexical development. Further assessment and/or follow-up is important for children who have had weak language/lexical skills at 2;0 years.

Keywords: receptive lexicon; expressive lexicon; language skills; pre-literacy skills; longitudinal study; MacArthur–Bates CDI

Abbreviations: BNT, Boston Naming Test; CDI, MacArthur–Bates Communicative Development Inventories; FinCDI-SF, the Finnish short-form version of the CDI; FMT, Finnish Morphology Test; FPT, Finnish Phonology Test; GLS, general language score; P-LS, pre-literacy score; RDLS, Reynell Developmental Language Scales III.

Highlights

- Significant associations between early lexical skills and different language and pre-literacy skills at 5 years were found.
- The best models that included early lexical predictors explained 20–34% of later language/literacy outcome.
- Most children with weak lexical/language skills at 2 years had weak skills in at least one language domain at 5 years.
- The findings emphasize the importance of screening for both early receptive and expressive lexical/language skills.

1. Introduction

The acquisition of the first basic lexicon, both receptive and expressive, is one of the most important linguistic milestones during the second year of life. The important issue as regards early lexical skills is their possible role in predicting later language and language-related skills. The main aim of the present study was to investigate the longitudinal associations between early lexical skills and language and pre-literacy skills at five years of age.

The size of the receptive lexicon is often beyond measurement at 1;6 years and especially at 2;0 years of age. The expressive lexicon grows rapidly from the typical 30–90 words at 1;6 years to 200–350 words at 2;0 years in different languages [1-3], including Finnish [4-6]. Representative estimations of children's early receptive and expressive lexical skills may be made using structured, normed and validated parental assessment tools, such as the widely used and adapted long- and short-form versions of the MacArthur–Bates Communicative Development Inventories (CDI) [2,7,8]. An approximately 50-word expressive lexicon or performance at or below the lowest 10th percentile in parental assessment measures of lexical skills is often used as the limit to differentiate between typical and weak lexical skills at 2;0 years. Children who score below but have no developmental disorders or sensory deficits are traditionally called late talkers [9-11]. Although many children with only weak expressive lexical skills at the age of two usually catch up with their peers by the age of five in many language domains [9,10,12-14], a history of being a late talker or having weak expressive lexical skills may still negatively influence later language abilities, such as reading and writing skills, throughout childhood and adolescence (see [10,15-18] for reviews). Late talkers who have also had weak receptive lexical skills around 2;0 years are particularly at risk of persistent language difficulties [12,16].

Five year-olds are typically versatile language users with large vocabularies. They are in the finalizing phase of phonological acquisition, during which the most difficult consonants and consonant clusters are acquired [19]. The basic knowledge of inflectional morphological rules in the native language has already been acquired at 5;0 years [20,21]. Children are typically proficient language users, and their language development can be assessed by a broad formal test at 5;0 years. Pre-reading skills, such as letter naming and phonological awareness, have also begun to develop at this age [22,23].

Studies investigating longitudinal associations between early receptive lexical and later language skills are scarce [22,24-27]. Specific information on what language domains (i.e. lexicon, phonology, morpho-syntax) early receptive lexical skills are associated with beyond the first years of life is especially lacking. In terms of expressive lexicon, continuity between early expressive and later lexical skills is intuitively expected and often found [18,22,25,26,28,29]. However, as is the case with receptive lexicon, the question of what other language domains early expressive lexical skills are specifically associated with remains unanswered. More is known about the simultaneous associations between early expressive lexical skills and phonology [30,31]. Studies assessing longer-term associations are scarce, but shorter-term associations have been found [24,32,33]. Further, early receptive lexical skills may have weaker associations with later phonological skills in comparison to early expressive language skills [24,34]. Continuity between early receptive and/or expressive lexical skills and later morphological skills has also been reported [21,22,35]. Early receptive lexical skills may have even clearer or earlier associations with morphological skills than early expressive lexical skills [22,24]. Furthermore, difficulties in morphological skills may be more prevalent among children at risk of persistent language difficulties, such as late talkers [10,13,21,36,37].

Longitudinal studies on the association between early receptive and/or expressive lexical skills and later pre-literacy skills, such as letter identification and phonological awareness, are scarce (see, however, [22]). The size of early expressive lexicon, measured using the long-form version of the CDI, has been associated with pre-literacy skills at the age of five and literacy skills up to the age of 16 [22,29,38,39]. To our knowledge, only one study has used the short-form version of the CDI to investigate the association between early lexical skills and later pre-reading skills. It found only a few significant correlations between early expressive lexical skills and pre-literacy skills four years later and did not assess early receptive lexical skills at all [28]. Assessing both early receptive and expressive skills together may deepen our understanding of their roles in pre-literacy skills [22]. Moreover, the risk of reading- and writing-related skill difficulties increases if the child has multiple risk factors, such as weak receptive and expressive early lexical development [12,38,40].

The possible effect of demographic factors such as gender and maternal education on lexical and language skills has been of wide interest [9,18,41-44]. Girls have often been reported as outperforming boys, especially in expressive lexical and language skills, across early and middle childhood, and gender-specific norms have been provided [2,41]. However, gender differences in receptive language have been studied less, and further investigation is thus needed [2,5,41,44]. Furthermore, less is known about whether these differences can be found in other language domains and beyond the second year [42,45]. Lastly, although demographic factors tend to explain only a small amount of the variance in lexical/language skills, their possible contribution has been discussed repeatedly and is hence worthy of attention [2,6,9,27].

The main aim of the present study was to investigate the longitudinal associations between early lexical ability at the end of the second year, and language and pre-literacy skills at the age of five. The following research questions were addressed:

1) Is there an association between early lexical (receptive and expressive) skills and language skills at five years of age? If so, what language domains (lexicon, phonology, morphology, pre-literacy skills) are early lexical skills associated with?

2) To what extent do early lexical skills (receptive and expressive) measured at 1;6 and at 2;0 years explain language and/or pre-literacy skills at 5;0 years, with and without the possible effect of gender?

3) Do children with weak receptive language and/or weak expressive lexical skills at 2;0 have weak language and pre-literacy skills at 5;0 years? If so, which language/pre-literacy skills are weak?

2. Participants and methods

This study is part of the adaptation and norming study of the Finnish short-form version of the MacArthur–Bates Communicative Development Inventories (FinCDI-SF; so-called Sanaseula Study, project leader: Prof. Suvi Stolt). The FinCDI-SF study has been approved by the Ethical Committee at the University of Turku. Parents received written feedback on their child's language skills at 2;0 and 5;0 years. If their child had delayed language skills, the families were advised to contact their local child health center for detailed guidance.

2.1. Participants

The development in language skills of 66 (34 girls, 39 firstborns) monolingual Finnish-speaking children was followed at 1;6, 2;0 and 5;0 years. The participants were not known to

have any neurological disorders, such as hearing impairments, cognitive delay, cerebral palsy or autism spectrum disorder, when they were invited to participate in the study during a normal follow-up visit to a child health center at the end of their first year of life. The parents were not known to have alcohol or drug abuse or mental health issues at the time the families were invited to join the study. All parents had completed more than nine years of compulsory schooling; 11% (n=7) of mothers had completed high school or vocational school and 24% (n=15) of fathers had completed high school or vocational school or had achieved other qualifications after compulsory schooling; and 89% (n=59) of mothers and 76% (n=50) of fathers had more than 12 years of basic education. The education level of the parents in this study is representative of the high education levels of young adults in Finland in general [46].

2.2. Methods

The following methods were used at 1;6 and 2;0 years. The parents completed the two different versions of the FinCDI-SF, which measured early lexical skills. The Infant version of the FinCDI-SF, used at 1;6 years, consists of an 89-word checklist with separate columns for receptive and expressive lexical skills [47]. The Toddler version, used at 2;0 years, consists of a 100-word checklist and assesses only expressive lexical skills, with an additional question on the emergence of word combinations. The checklists contain words from different semantic categories. The parents completed the forms within two weeks of their child turning 1;6 and 2;0 years. The receptive part of the Reynell Developmental Language Scales III (RDLS III [48]) was used at 2;0 years to assess receptive language skills. The RDLS is a well-known, formal, structured language test, which has been adapted in Finnish and has Finnish norms. The first parts measure lexical skills, such as understanding and naming objects and actions. The later sections also assess morphological and syntactical skills, and are completed if the child is skillful. Testing continues until the child makes three errors in a row in two consecutive sections

(max. 62 points). The raw and standardized scores (mean=100, +/-1SD=15) of the receptive part were used.

The language domains assessed at 5;0 years were lexicon, phonology, morphology and pre-literacy skills. The Boston Naming Test (BNT [49]) was used to measure lexical skills. It consists of 60 pictures, presented to the child in an order of increasing difficulty, which the child is asked name (max. 60 points). The child continues naming the pictures until they are unable to name six pictures in a row. The Finnish Phonology Test (FPT [50]) measures phonological skills. The phonotactic score (max. 127 points), which measures the length of phonemes and syllables (max. 26 points), word length in syllables (max. 18 points) and phoneme combinations (max. 83 points), was used. The test consists of 90 pictures, all of which the child is asked to name. The Finnish Morphology Test (FMT [20,21]) assesses the ability to generalize the knowledge of six different inflection types (adverbs; comparative and superlative for adjectives; relative for nouns; present and past tense for verbs; five words for each inflection, altogether 30 words, max. 90 points). A word derived from old, obsolete Finnish is presented orally together with a picture, and the child is instructed to use the same word with an appropriate inflection based on the information presented in the picture.

Information on pre-literacy skills was gathered using the following methods at 5;0 years: letter identification (derived from the Lukiva screening test [23]) and two phonological awareness tasks (derived and adapted from the Armi 1 test [51]). In the letter identification task, the children are asked to name as many as they can of the shown 23 capital letters as they could (max. 23 points). A point is given from each correct answer, either letter name or phoneme. The phonological awareness tasks were the construction of words from syllables uttered to the child and initial phoneme recognition (total 12 points).

2.3. Data analyses

Spearman's correlation co-efficient values were used to investigate the associations between early (measured at 1;6 and 2;0 years) and later (measured at 5;0 years) language and pre-literacy variables. Unique variances (*R* squares) were calculated to assess the predictive value for each early language variable and gender separately. Furthermore, four regression models were run in order to obtain information on the extent to which early receptive and expressive lexical skills together, when measured at 1;6 or at 2;0 years, may explain variance in later language or pre-literacy skills. In Models 1 and 3, the outcome variable was a composite variable, called the general language score (GLS; based on the BNT, FPT, FMT) at 5;0 years. In Models 2 and 4, the outcome variable was a composite score, called the pre-literacy score (P-LS), based on the two pre-literacy skill tasks at 5;0 years. To create the composite scores, the values of each language and pre-literacy variable at 5;0 years were first converted to *z*-scores (mean=0; $\pm 1 = \pm 1$ SD etc.). Then the mean values of the *z*-scores were calculated for GLS and P-LS. The variables of early receptive and expressive lexical skills and gender were entered to the models at the same time. Finally, using the *z*-scores, the language profiles at 5;0 years of the children who had had weak receptive and/or expressive lexical/language skills at 2;0 years were examined in detail. Weak receptive language skills at 2;0 years were defined as < 85 SD in the receptive part of the RDLS III. Weak lexical skills were defined as performing within the weakest 10th percentile according to the normative data of the FinCDI-SF at 1;6 years (receptive and expressive) and 2;0 years (expressive).

3. Results

3.1. Data description

Individual children's language skills varied greatly in all the language variables and at all the measured age points (Table 1). The mean standard score of the receptive part of the RDLS at 2;0 years was 107 (SD 17, range 73–142). At 1;6 years, six children (9%; 3 boys) had weak receptive skills and eight children (12%; 6 boys) had weak expressive lexical skills. At 2;0 years, seven children (11%; 6 boys) had weak receptive language skills, two of whom had had weak lexical skills at 1;6 years. Five children (8%; 3 boys) had weak expressive lexical skills at 2;0 years; none of these had started combining words, and four had had weak expressive lexical skills at 1;6 years. One child had weak receptive language and expressive lexical skills at 2;0 years. The results of the tests conducted at 5;0 years were roughly in parallel with Finnish normative data.

(insert Table 1 about here)

At 1;6, the girls scored statistically significantly higher than the boys in expressive lexical skills [$t(63) = -2.55, p < .05, \eta^2 = .09$; mean(*SD*) for girls: 31(20) and boys: 19(18)] but not in receptive lexical skills [$t(63) = -1.44, p = .16$]. At 2;0 years the girls outperformed the boys in both receptive language [$t(64) = -2.34, p < .05, \eta^2 = .08$; mean(*SD*) for girls: 26(8) and boys: 21(9)] and expressive lexical skills [$t(64) = -2.19, p < .05, \eta^2 = .07$; mean(*SD*) for girls: 63(22) and boys: 49(29)]. The effect sizes were small, explaining 9% of the variance in expressive lexical skills at 1;6 years, and 8% and 7% of the variance in receptive language and expressive lexical skills, respectively at 2;0 years, and a statistically significant difference was found. At 5;0 years, the mean performance of the girls was higher than that of the boys in all the measured variables, but the differences were not statistically significant. No significant differences were found between children with higher and lower maternal or paternal education (more than 12 years; 12

years or less), or between children born first or later in terms of any measured language variable at any age point.

3.2. Associations between the early lexicon and language skills at 5;0 years

Receptive lexical skills at 1;6 years were only associated with expressive lexical skills at 5;0 years (Table 2). Correlations with phonology, morphology and pre-literacy skills did not reach the statistically significant level (p -values=.10–.13). Expressive lexical skills at 1;6 years correlated statistically significantly and relatively evenly with all the measured variables at 5;0 years. Furthermore, receptive language skills at 2;0 years correlated moderately and significantly with all the measured language and pre-literacy skills at 5;0 years. Expressive lexical skills at 2;0 years had significant, clear correlations with lexical, phonological and letter naming skills and phonological awareness at 5;0 years, but the correlation with morphology did not reach a statistically significant level (p =.08).

(insert Table 2 about here)

3.3. Predictive values of early lexical skills and later language and pre-literacy skills

Based on unique variances, which can be interpreted as independent predictive values, all the early lexicon/language variables had statistically significant predictive values for the GLS (Figure 1). All the lexicon/language variables, excluding receptive lexical skills at 1;6 years, also had statistically significant predictive values for the P-LS. At 1;6 years, receptive lexical skills significantly predicted a small amount of variation only in the GLS. Expressive lexical skills predicted 22% of variation in both the GLS and P-LS. At 2;0 years, receptive language skills predicted 35% of variation in the GLS but only 10% of variation in the P-LS. Expressive

lexical skills predicted variation in the GLS and P-LS equally well, that is, 16%. Gender and maternal education were not significant predictors of the GLS or P-LS.

(Insert Figure 1 about here)

Regression models were used to examine the predictive value of four models with lexicon/language and background variables at 1;6 and 2;0 years for the GLS and P-LS at 5;0 years. Regarding predictors at 1;6 years (Models 1 and 2), both models explained 19% of the variation in the GLS and P-LS (Table 3). Expressive lexicon was the only significant predictor in both models. Regarding predictors at 2;0 years (Models 3 and 4), the models explained 33% of variation in the GLS and 14% of variation in the P-LS at 5;0 years. The only statistically significant predictors in these models were receptive language skills for the GLS, and expressive lexical skills for the P-LS. Model 3 was the best predictor of the GLS and Model 2 the best predictor of the P-LS.

(Insert Table 3 about here)

3.4. Language skills at 5;0 years of children with weak receptive language and/or expressive lexical skills at 2;0 years

The children who had had weak receptive language skills at 2;0 years (n=7, 11%) had varying language profiles at 5;0 years (Figure 2). Of these seven children, all but one (Child 2, dashed line with +) had weak language skills ($<-1SD$) in at least one language outcome at 5;0 years. The children with weak expressive lexical skills (n=5, 8%) also had varying language profiles at 5;0 years (Figure 3). Of these five children, one had a steady and age-appropriate profile in all measured domains at 5;0 years (Child 11, line with squares). The rest of the children with

weak early expressive lexical skills had weak skills in at least two language domains at 5;0 years (Figure 3). The only child with weak receptive language and expressive lexical skills at 2;0 years (Child 4, black line with circles) had weak language skills ($<-1SD$) in all of the measured language outcomes at 5;0 years.

4. Discussion

The main aim of this study was to analyze the longitudinal associations between early receptive and expressive lexical skills measured at 1;6 and 2;0 years and language and pre-literacy skills measured at 5;0 years. Continuity between early receptive and expressive lexical skills, and lexical, phonological, morphological and pre-literacy skills at 5;0 years was found. Early receptive and expressive lexical skills significantly predicted the GLS and P-LS at 5;0 years. The model that best predicted the GLS at 5;0 years had predictors at 2;0 years and explained 33% of the variation in the GLS. The model that best predicted the P-LS had predictors at 1;6 years and explained 19% of the variance in the P-LS. The girls had an advantage over the boys at 1;6 years (expressive lexical skills) and 2;0 years (receptive language and expressive lexical skills) but not at 5;0 years. Among most of the children, both weak receptive language and expressive lexical skills at 2;0 years were associated with weak language skills at 5;0 years in at least one language domain.

Significant associations between early receptive and expressive lexical skills and language and pre-literacy skills at 5;0 years were found already at 1;6 years and also at 2;0 years. Receptive lexical skills at 1;6 years associated significantly with lexical skills at 5;0 years. This finding is in accordance with those of previous studies that have reported long-term associations between receptive lexical skills measured before 2;0 years and later expressive lexical skills [25,27]. The

associations with other language domains were not significant. However, receptive language skills measured at 2;0 years associated significantly with all variables measured at 5;0 years. This may be due to maturation in receptive language skills, but also to the change in the assessment method between these age points. Expressive lexicon, on the other hand, measured at both 1;6 and 2;0 years, associated in a rather parallel manner across all the measured language and pre-literacy domains at 5;0 years. These findings add to those of previous studies that show that direct associations between early lexical skills and later language skills may extend beyond three years of age [22,24,26,27]. The association between expressive lexical skills at 2;0 years and morphology at 5;0 years was evident but only almost significant. This is an interesting finding, as previous studies have reported significant associations between early expressive lexicon and later morphology around the age of five [35]. Thus, this association may have reached statistical significance if the sample had been larger. Our findings also revealed associations between early receptive language and expressive lexical skills and later pre-literacy skills. To our knowledge, no significant associations between these variables have found earlier using the short-form version of the CDI (compare to [28]). The association between very early lexical ability and pre-literacy skills at 5;0 years may extend even further to literacy skills at school age, up to the age of 16 [22,29,38,39]. Thus, our finding underlines the importance of very early screening of early receptive and expressive lexical skills.

Our findings revealed that receptive lexical skills at 1;6 years, expressive lexical skills at 1;6 years, and expressive lexical skills at 2;0 years predicted the GLS and P-LS in a comparable manner. The exception to this trend was receptive language skills at 2;0 years, which predicted them differently: 35% of variation in the GLS but only 10% of that in the P-LS. The strong relationship between receptive language skills at 2;0 years and general language skills at 5;0 years could be partly explained by the use of different testing methods at 2;0 years. The RDLS

also measures comprehension of morphology and syntax if a child is skillful, whereas the FinCDI-SF only measures lexical skills. This may strengthen the predictive value of receptive language skills for the GLS. In addition to the possible impact of different assessment methods, previous literature has also shown that receptive lexical/language skills at about 2;0 years may have good predictive value for different expressive language skills later in life [18], but that they make no direct contributions to the P-LS [22]. Expressive lexical skills, on the other hand, seem to predict language and pre-literacy skills across a wider range of skills and relatively more evenly, as reported earlier [22,28,29]. Thus, our results support these previous findings.

Including all the predictors of interest in the same model did not lead to a particularly higher predictive value. The finding that only one variable was significant in each model may be explained by the overlapping variance between receptive and expressive language. However, neither the correlations between nor the variance inflation factor values of these skills at 1;6 and 2;0 years were too large to prevent their simultaneous usage in the analyses. Although early lexical skills were able to explain a part of the later GLS and P-LS, the other factors, child and/or environmental, that could help better predict later language and pre-literacy skills are still not known thoroughly (compare to [9,18,22,52]).

The present findings show that weak early receptive/expressive lexical skills indicate vulnerability to weak skills in at least one language or pre-literacy domain at 5;0 years. This is in accordance with the results of previous studies [10,11,39]. A positive outcome regarding weak skills at 2;0 years is that most children with only weak early expressive lexical skills catch up with their peers by 5;0 years in many language domains, which has also been reported previously [15,16]. The one child with both weak receptive and expressive language/lexical skills at 2;0 years in the present sample had considerably weak language skills in all measured

domains at 5;0 years. Although the number of children with weak receptive language/expressive lexical skills was small (11%, 8%), it was representative in regard of late talkers [9,10]. A greater number of larger-scale studies are needed to verify these results. To conclude, in line with previous studies [16,17], although catch-up is often achieved, weak early receptive/expressive lexical skills still seem to present a risk of weaker language skills in later years.

The present results raise the question of the more preferable age for screening, 1;6 or 2;0 years. This issue has previously received relatively little attention [18,27,53]. The more frequently used age has been 2;0 years, due to the typically highly active phase of lexical acquisition during the second half of the second year and the presumed emergence of word combinations [1,2,9,54]. In particular, 2;0 years has been frequently used for defining late talkers [9,10,15,16]. Therefore we also chose 2;0 years for the further analysis of the language profiles of children with weak receptive language and/or expressive lexical skills at 2;0 years. Although our results show clear associations and predictive values already at 1;6 years, suggesting that it may be a promising age for screening, 1;6 years may be too early an age for reliable screening, as discussed in previous studies [27,53]. Some individual fluctuation between typical and weak outcomes between 1;6 years and 2;0 years is possible and was also evident in the present study. Therefore, 2;0 years seems to be the more reliable age for screening. It might also be a better age for the assessment of receptive language. However, methodological issues are a problem. Quick, easy and valid methods for screening receptive lexical skills at 2;0 years are greatly needed.

Girls were ahead of boys in expressive lexical skills at 1;6 years and receptive language and expressive lexical skills at 2;0 years, but not at 5;0 years. This is in accordance with previous

findings in which girls have tended to be at least slightly more advanced than boys, especially in expressive lexical skills during the second year of life [2,3,27,41,42]. However, boys seem to catch up during the later years of middle childhood [13,42], especially if they are from higher socioeconomic status families [45]. Despite this, allowing boys with weak early lexical skills more time to catch up should be considered with great caution, for several reasons. Firstly, boys are at a higher risk of persistent language delays [9]. Secondly, according to both the present and previous studies, early lexical skills overlap considerably between genders [2,41,42]. Moreover, as the present study also found, when examining both unique variances and regression models, gender explains at best only a small amount of the variability in early and especially later language skills. Finally, waiting may result in delays in possibly essential intervention [17,41].

4.1. Strengths, limitations and further directions

The longitudinal study design and adequate number of participants are strengths of the present study. Nevertheless, a larger data set may have more effectively verified the results and would have been especially beneficial for further studying the language profiles of children who had weak lexical/language skills at 2;0 years. A comparison between 1;6 and 2;0 years, and using a larger data set on the later language skills of children with weak early lexical skills would also have provided more answers regarding the best age for screening [14,27,53]. Language and pre-literacy skills at 5;0 years were measured widely. However, a more comprehensive assessment of syntactic skills, receptive language and pragmatic skills at 5;0 years could have promoted an even broader understanding of the possible long-term associations. The analyses of the associations between early lexical skills and later language and pre-literacy skills used correlations, unique variances and regression models. The language and pre-literacy domains were analyzed independently in the correlation analyses and as combined variables for unique

variances and regression models. Together these analyses provided a coherent picture of the longitudinal relationships.

4.2. Clinical implications

The short-form versions of the CDI provided valid information on the longitudinal associations between early and later language skills. The short-form versions of the CDI are easy, fast and inexpensive tools for use in clinical contexts. Our results highlight the importance of the early assessment of both receptive and expressive lexical skills. Follow-up of children with weak early receptive or expressive lexical skills is highly recommended at the end of the second year and beyond. A broader assessment is needed at 2;0 years, especially if the child also has other risk factors for persisting language difficulties [11,17,36,49]. Knowing the history of a child's early lexical development could help professionals more readily provide later support if needed.

5. Conclusions

This study provided information on the longitudinal associations between early lexical skills, both receptive and expressive, and later language and pre-literacy outcomes. The associations between receptive language skills and expressive lexical skills were fairly parallel at 2;0 years, and also across different language domains at 5;0 years. Early receptive language and expressive lexical skills explained the variance in later language abilities both independently and as part of regression analyses models. Gender accounted for only a small amount of variation in early lexical skills and was not a significant predictor of the GLS or P-LS. Weak receptive language/expressive lexical skills at 2;0 years indicated vulnerability to difficulties in language and/or pre-literacy-related skills at 5;0 years. Follow-up and/or further assessment is recommended if weak lexical skills are identified at the end of a child's second year.

CRedit authorship contribution statement

Suvi-Maria Vehkavuori: Conceptualization, methodology, formal analysis, investigation, writing – original draft, visualization, funding acquisition.

Maiju Kämäräinen: Formal analysis – preliminary analysis, writing – review and editing.

Suvi Stolt: Conceptualization, methodology, resources, writing – review and editing, supervision, project administration.

Acknowledgements

We wish to thank all the families who participated in the study and those who helped with the data gathering. The manuscript finalizing process was supported by a grant from The University of Oulu Scholarship Foundation. The Sanaseula Study has been supported by the Jenny and Antti Wihuri Foundation.

Declaration of competing interest

The authors report no conflicts of interest.

References

- [1] D. Bleses, W. Vach, M. Slott, S. Wehberg, P. Thomsen, T.O. Madsen, et al. Early vocabulary development in Danish and other languages: A CDI-based comparison, *J. Child Lang.* 35 (2008) 619–650. <https://doi.org/10.1017/S0305000908008714>.
- [2] L. Fenson, V. Marchman, D. Thal, P. Dale, J. Reznick, E. Bates, *The MacArthur–Bates Communicative Development Inventories - User's Guide and Technical Manual*, Second Edition, Brookes, Baltimore, MD, 2007.
- [3] S. Reilly, E.L. Bavin, L. Bretherton, L. Conway, P. Eadie, E. Cini, et al. The Early Language in Victoria Study (ELVS): A prospective, longitudinal study of communication skills and expressive vocabulary development at 8, 12 and 24 months, *Int. J. Speech Lang. Pathol.* 11 (2009) 344–357. <https://doi.org/10.1080/17549500903147560>.
- [4] P. Lyytinen, *Varhaisen kommunikaation ja kielen kehityksen arviointimenetelmä* [Finnish manual of the MacArthur–Bates Communicative Development Inventories], Niilo Mäki Institute & Child Research Center, Jyväskylä, 1999.
- [5] S. Stolt, L. Haataja, H. Lapinleimu, L. Lehtonen. Early lexical development of Finnish children: A longitudinal study, *First Lang.* 28 (2008) 259–279. <https://doi.org/10.1177/0142723708091051>.
- [6] P. Korpilahti, A. Kaljonen, E. Jansson-Verkasalo. Identification of biological and environmental risk factors for language delay: The Let's Talk STEPS study, *Infant Behav. Dev.* 42 (2016) 27–35. <https://doi.org/10.1016/j.infbeh.2015.08.008>.
- [7] L. Fenson, S. Pethick, C. Renda, J.L. Cox, P.S. Dale, J.S. Reznick. Short-form versions of the MacArthur Communicative Development Inventories, *Appl. Psycholinguist.* 21 (2000) 95–116. <https://doi.org/10.3758/s13428-018-1146-0>.
- [8] J. Law, P. Roy. Parental Report of Infant Language Skills: A Review of the Development and Application of the Communicative Development Inventories, *Child Adolesc. Ment. Health.* 13 (2008) 198–206. <https://doi.org/10.1111/j.1475-3588.2008.00503.x>.
- [9] E.L. Bavin, L. Bretherton, The early language in Victoria study - Late talkers, predictors, and outcomes, in: L.A. Rescorla, P.S. Dale (Eds.), *Late Talkers: Language Development, Interventions, and Outcomes*, Paul H. Brookes Publishing Co., Baltimore, MD, 2013, pp. 3–21.
- [10] E.M. Roos, S. Ellis Weismer. Language Outcomes of Late Talking Toddlers at Preschool and Beyond, *Persp. Lang. Learn. Educ.* 15 (2008) 119–126. <https://doi.org/10.1044/ll15.3.119>.
- [11] E.M. Ellis, D.J. Thal. Early Language Delay and Risk for Language Impairment, *Persp. Lang. Learn. Educ.* 15 (2008) 93–100. <https://doi.org/10.1044/ll15.3.93>.

- [12] P. Lyytinen, K. Eklund, H. Lyytinen. Language Development and Literacy Skills in Late-talking Toddlers with and without Familial Risk for Dyslexia, *Ann. of Dyslexia*. 55 (2005) 166–192. <https://doi.org/10.1007/s11881-005-0010-y>.
- [13] M.L. Rice, C.L. Taylor, S.R. Zubrick, Language outcomes of 7-year-old children with or without a history of late language emergence at 24 months, *J. Speech Lang. Hear. Res.* 51 (2008) 394–407. [https://doi.org/10.1044/1092-4388\(2008/029\)](https://doi.org/10.1044/1092-4388(2008/029)).
- [14] P.S. Dale, T.S. Price, D.V.M. Bishop, R. Plomin. Outcomes of Early Language Delay: I. Predicting Persistent and Transient Language Difficulties at 3 and 4 Years, *J. Speech Lang. Hear. Res.* 46 (2003) 544–560. [https://doi.org/10.1044/1092-4388\(2003/044\)](https://doi.org/10.1044/1092-4388(2003/044)).
- [15] V.V. Hawa, G. Spanoudis. Toddlers with delayed expressive language: An overview of the characteristics, risk factors and language outcomes, *Res. Dev. Disabil.* 35 (2014) 400–407. <https://doi.org/10.1016/j.ridd.2013.10.027>.
- [16] L. Rescorla. Late Talkers: Do Good Predictors of Outcome Exist? *Dev. Disabil. Res. Rev.* 17 (2011) 141–150. <https://doi.org/10.1002/ddrr.1108>.
- [17] N. Capone Singleton, Late Talkers: Why the Wait-and-See Approach Is Outdated, *Pediatr. Clin. North Am.* 65 (2018) 13–29. <https://doi.org/10.1016/j.pcl.2017.08.018>.
- [18] E.L. Fisher. A systematic review and meta-analysis of predictors of expressive-language outcomes among late talkers, *J. Speech Lang. Hear. Res.* 60 (2017) 2935–2948. https://doi.org/10.1044/2017_JSLHR-L-16-0310.
- [19] K. Saaristo-Helin, S. Kunnari, T. Savinainen-Makkonen. Phonological development in children learning Finnish: A review, *First Lang.* 31 (2011) 342–363. <https://doi.org/10.1177/0142723710396793>.
- [20] P. Lyytinen, *Morfologiatesti - taivutusmuotojen hallinnan testi [the Finnish Morphology test]*, University of Jyväskylä, Jyväskylä, 1988.
- [21] P. Lyytinen, H. Lyytinen. Growth and predictive relations of vocabulary and inflectional morphology in children with and without familial risk for dyslexia, *Appl. Psycholinguist.* 25 (2004) 397–411. <https://doi.org/10.1017/S0142716404001183>.
- [22] M. Torppa, P. Lyytinen, J. Erskine, K. Eklund, H. Lyytinen. Language Development, Literacy Skills, and Predictive Connections to Reading in Finnish Children With and Without Familial Risk for Dyslexia, *J. Learn. Disabil.* 43 (2010) 308–321. <https://doi.org/10.1177/0022219410369096>.
- [23] A. Puolakanaho, A. Poikkeus, T. Ahonen, M. Aro, LUKIVA – Lukivalmiuksien arviointimenetelmä 4–5-vuotiaille [Pre-literacy assessment], Niilo Mäki Institute, Jyväskylä, 2011.
- [24] S. Vehkavuori, S. Stolt. Early lexicon and language skills at 42 months, *Clin. Linguist. Phon.* 33 (2019) 854–868. <https://doi.org/10.1080/02699206.2019.1584721>.

- [25] S. Stolt, A. Lind, J. Matomäki, L. Haataja, H. Lapinleimu, L. Lehtonen. Do the early development of gestures and receptive and expressive language predict language skills at 5;0 in prematurely born very-low-birth-weight children? *J. Commun. Disord.* 61 (2016) 16–28. <https://doi.org/10.1016/j.jcomdis.2016.03.002>.
- [26] T. Patrucco-Nanchen, M. Friend, D. Poulin-Dubois, P. Zesiger. Do early lexical skills predict language outcome at 3 years? A longitudinal study of French-speaking children, *Infant Behav. Dev.* 57 (2019) 1–18. <https://doi.org/10.1016/j.infbeh.2019.101379>.
- [27] J. Henrichs, L. Rescorla, J.J. Schenk, H.G. Schmidt, V.W.V. Jaddoe, A. Hofman, et al. Examining continuity of early expressive vocabulary development: The generation R study, *J. Speech Lang. Hear. Res.* 54 (2011) 854–869. [https://doi.org/10.1044/1092-4388\(2010/09-0255\)](https://doi.org/10.1044/1092-4388(2010/09-0255)).
- [28] D.D. Can, M. Ginsburg-Block, R.M. Golinkoff, K. Hirsh-Pasek. A long-term predictive validity study: Can the CDI Short Form be used to predict language and early literacy skills four years later? *J. Child Lang.* 40 (2013) 821–835. <https://doi.org/10.1017/S030500091200030X>.
- [29] J. Lee. Size matters: Early vocabulary as a predictor of language and literacy competence, *Appl. Psycholinguist.* 32 (2011) 69–92. <https://doi.org/10.1017/S0142716410000299>.
- [30] M. Kehoe, T. Patrucco-Nanchen, M. Friend, P. Zesiger. The relation between phonological and lexical development in French-speaking children, *Clin. Linguist. Phon.* 32 (2018) 1103–1125. <https://doi.org/10.1080/02699206.2018.1510984>.
- [31] C. Stoel-Gammon. Relationships between lexical and phonological development in young children, *J. Child Lang.* 38 (2011) 1–34. <https://doi.org/10.1017/S0305000910000425>.
- [32] E. Aalto, K. Saaristo-Helin, S. Stolt. Phonological development of Finnish speaking children at 3;6 and its associations to previous and simultaneous lexical ability. *Clin. Linguist. Phon.* 34 (2020) 617–633. <https://doi.org/10.1080/02699206.2019.1681517>.
- [33] A.L. Williams, M. Elbert. A prospective longitudinal study of phonological development in late talkers, *Lang. Speech. Hear. Serv. Sch.* 34 (2003) 138–153. [https://doi.org/10.1044/0161-1461\(2003/012\)](https://doi.org/10.1044/0161-1461(2003/012)).
- [34] T. Macrae, A.V. Sosa. Predictors of token-to-token inconsistency in preschool children with typical speech-language development, *Clin. Linguist. Phon.* 29 (2015) 922–937. <https://doi.org/10.3109/02699206.2015.1063085>
- [35] M. Silvén, A. Ahtola, P. Niemi. Early words, multiword utterances and maternal reading strategies as predictors of mastering word inflections in Finnish, *J. Child Lang.* 30 (2003) 253–279. <https://doi.org/10.1017/S0305000902005548>
- [36] A.M. Chilosi, L. Pfanner, C. Pecini, R. Salvadorini, C. Casalini, D. Brizzolara, et al., Which linguistic measures distinguish transient from persistent language problems in Late Talkers from 2 to 4 years? A study on Italian speaking children, *Res. Dev. Disabil.* 89 (2019) 59–68. <https://doi.org/10.1016/j.ridd.2019.03.005>.

- [37] M.J. Moyle, S. Ellis Weismer, J.L. Evans, M.J. Lindstrom. Longitudinal relationships between lexical and grammatical development in typical and late-talking children, *J. Speech Lang. Hear. Res.* 50 (2007) 508–528. [https://doi.org/10.1044/1092-4388\(2007/035\)](https://doi.org/10.1044/1092-4388(2007/035)).
- [38] M. Psyridou, K. Eklund, A. Poikkeus, M. Torppa, Reading outcomes of children with delayed early vocabulary: A follow-up from age 2–16, *Res. Dev. Disabil.* 78 (2018) 114–124. <https://doi.org/10.1016/j.ridd.2018.05.004>.
- [39] D. Bleses, G. Makransky, P.S. Dale, A. Højen, B.A. Ari. Early productive vocabulary predicts academic achievement 10 years later, *Appl. Psycholinguist.* 37 (2016) 1461–1476. <https://doi.org/10.1017/S0142716416000060>
- [40] P.A. Thompson, C. Hulme, H.M. Nash, D. Gooch, E. Hayiou-Thomas, M.J. Snowling. Developmental dyslexia: predicting individual risk, *J. Child Psychol. Psychiatr.* 56 (2015) 976–987. <https://doi.org/10.1111/jcpp.12412>.
- [41] M. Eriksson, P.B. Marschik, T. Tulviste, M. Almgren, M. Pérez Pereira, S. Wehberg, et al. Differences between girls and boys in emerging language skills: Evidence from 10 language communities, *Br. J. Dev. Psychol.* 30 (2012) 326–343. <https://doi.org/10.1111/j.2044-835X.2011.02042.x>.
- [42] M.H. Bornstein, C. Hahn, O.M. Haynes. Specific and general language performance across early childhood: Stability and gender considerations, *First Lang.* 24 (2004) 267–304. <https://doi.org/10.1177%2F0142723704045681>.
- [43] I.M. Zambrana, E. Ystrom, F. Pons. Impact of gender, maternal education, and birth order on the development of language comprehension: a longitudinal study from 18 to 36 months of age, *J. Dev. Behav. Pediatr.* 33 (2012) 146–155. <https://doi.org/10.1097/DBP.0b013e31823d4f83>.
- [44] E. Berglund, M. Eriksson, M. Westerlund. Communicative skills in relation to gender, birth order, childcare and socioeconomic status in 18-month-old children, *Scand. J. Psychol.* 46 (2005) 485–491. <https://doi.org/10.1111/j.1467-9450.2005.00480.x>.
- [45] S. Barbu, A. Nardy, J. Chevrot, B. Guellai, L. Glas, J. Juhel, et al. Sex Differences in Language Across Early Childhood: Family Socioeconomic Status does not Impact Boys and Girls Equally, *Front. Psychol.* 6 (2015) 1874–1974. <https://doi.org/10.3389/fpsyg.2015.01874>.
- [46] Official Statistics of Finland, Educational structure of population [e-publication]. http://www.stat.fi/til/vkour/2018/vkour_2018_2019-11-05_tie_001_en.html, 2018 (accessed 24th February 2020).
- [47] S. Stolt, S. Vehkavuori, SANASEULA – Lapsen varhaisen sanaston kehityksen seulontamenetelmä [Finnish short-form versions of the MacArthur–Bates Communicative Development Inventories], Niilo Mäki Institute, Jyväskylä, 2018.
- [48] M. Kortesmaa, K. Heimonen, H. Merikoski, M.-L. Warma, V Varpela, Reynell Developmental Language Scales III [Finnish scales], Psykologien kustannus Oy, Helsinki, 2001.

[49] M. Laine, P. Koivuselkä-Sallinen, R. Hänninen, J. Niemi, Bostonin nimentätesti [Finnish version of the Boston Naming Test], Psykologien Kustannus Oy, Helsinki, 1997.

[50] S. Kunnari, T. Savinainen-Makkonen, K. Saaristo-Helin, Fonologiatesti – Lapsen äänteellisen kehityksen arviointimenetelmä [the Finnish Phonology Test], Niilo Mäki Institute, Jyväskylä, 2012.

[51] M. Lerkkanen, A. Poikkeus, R. Ketonen, Armi 1. Luku- ja kirjoitustaidon arviointimateriaali 1. luokalle [Reading and writing assessment for firstgraders], WSOY pro Oy, Helsinki, 2006.

[52] S. Määttä, M. Laakso, T.A.A. Tolvanen, J. Westerholm, T. Aro. Continuity from prelinguistic communication to later language ability: A follow-up study from infancy to early school age, *J. Speech Lang. Hear. Res.* 59 (2016) 1357–1372.
https://doi.org/10.1044/2016_JSLHR-L-15-0209.

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

[54] S. Stolt, L. Haataja, H. Lapinleimu, L. Lehtonen. Associations between lexicon and grammar at the end of the second year in Finnish children, *J. Child Lang.* 36 (2009) 779–806.
<https://doi.org/10.1017/S0305000908009161>.

Table 1. Descriptive statistics of language variables at 1;6, 2;0 and 5;0 years

Age	Method	Mean (SD)	Median	Min.–Max.
1;6	Receptive/FinCDI-SF	66 (14)	68	34–89
	Expressive/FinCDI-SF	26 (20)	23	0–77
2;0	Receptive/RDLS	24 (9)	24	6–42
	Expressive/FinCDI-SF	56 (26)	59	4–100
5;0	Lexicon/BNT	36 (6)	36	22–50
	Phonology/FPT	118 (13)	125	71–127
	Morphology/FMT	58 (18)	61	17–87
	Letter identification	13 (8)	14	0–23
	Phonological awareness	9 (2)	9	5–12

Note: FinCDI-SF = Finnish Short-form version of the MacArthur–Bates Communicative Development Inventories, Infant version at 1;6 years and Toddler version at 2;0 years; RDLS = Reynell Developmental Language Scales III, raw points; BNT = Boston naming test; FPT = Finnish Phonology test; FMT = Finnish Morphology test.

Table 2. Spearman's correlations (*r*-values) between early lexical/language skills and language skills at 5;0 years

Age	Early language measure	Lexicon/ BNT	Phonology/ FPT	Morphology/ FMT	Letter identification	Phonological awareness
1;6	Receptive / FinCDI-SF	.38**	.21	.19	.20	.02
	Expressive / FinCDI-SF	.44**	.31*	.26*	.41**	.42**
2;0	Receptive / RDLS	.38**	.38**	.51**	.29*	.28*
	Expressive / FinCDI-SF	.33**	.34**	.22	.27*	.35**

Note: * = $p < .05$, ** = $p < .01$. BNT = Boston naming test, FPT = Finnish Phonology test, FMT = Finnish Morphology test, FinCDI-SF = Finnish short-form version of the MacArthur–Bates Communicative Development Inventories, RDLS = Reynell Developmental Language Scales III, raw points.

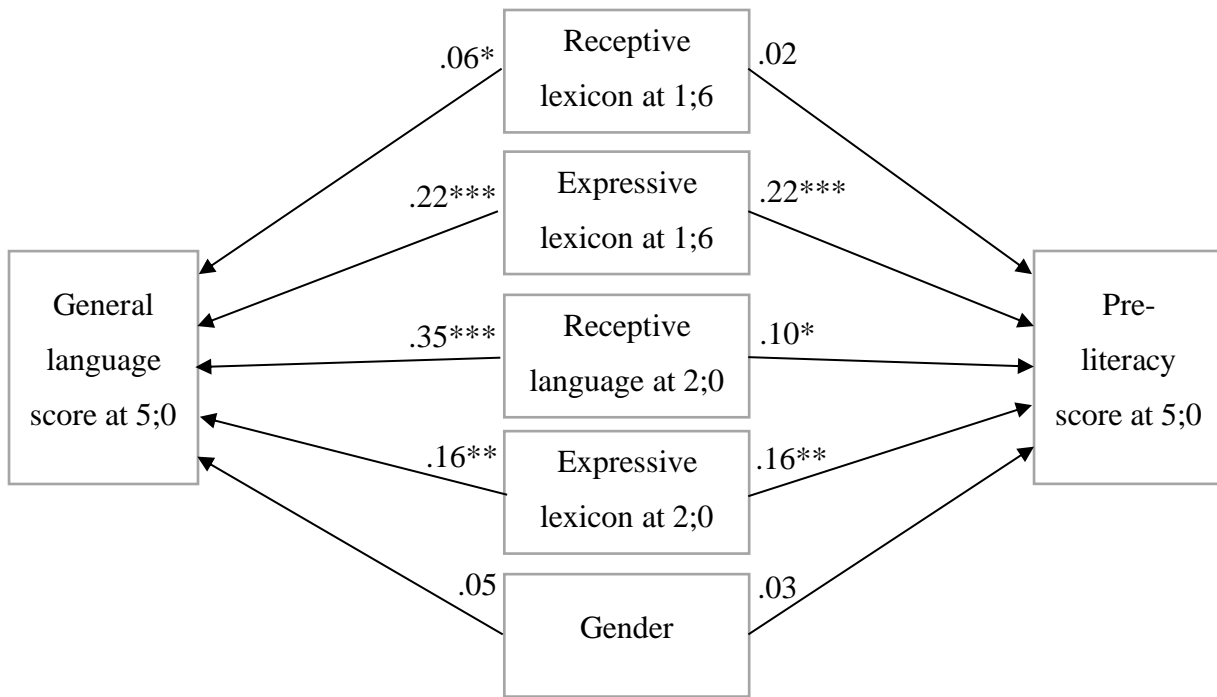


Figure 1. Unique variances (R^2) of early receptive lexical/language skills, expressive language skills and gender in general language skills and pre-literacy skills at 5;0 years. * = $p < .05$, ** = $p < .01$, *** = $p < .001$, p = model's p -value (F -test).

Table 3. Regression analyses of the general language and pre-literacy score

Model information	Predictors	β
Model 1.	At 1;6	
Outcome variable: GLS at 5;0	Receptive lexicon	.01
$F(3,60)=6.10^{**}$	Expressive lexicon	.44**
$R^2_{adj}=.19$	Gender	.09
Model 2.	At 1;6	
Outcome variable: P-LS at 5;0	Receptive lexicon	-.16
$F(3,60)=6.25^{**}$	Expressive lexicon	.53***
$R^2_{adj}=.20$	Gender	.06
Model 3.	At 2;0	
Outcome variable: GLS at 5;0	Receptive language	.51***
$F(3,62)=12.04^{***}$	Expressive lexicon	.14
$R^2_{adj}=.34$	Gender	.04
Model 4.	At 2;0	
Outcome variable: P-LS at 5;0	Receptive language	.14
$F(3,61)=4.46^{**}$	Expressive lexicon	.31*
$R^2_{adj}=.14$	Gender	.07

Note: GLS = general language score, P-LS = pre-literacy score, R^2_{adj} = adjusted R square, β = standardized beta coefficients, * = $p < .05$, ** = $p < .01$, *** = $p < .001$.

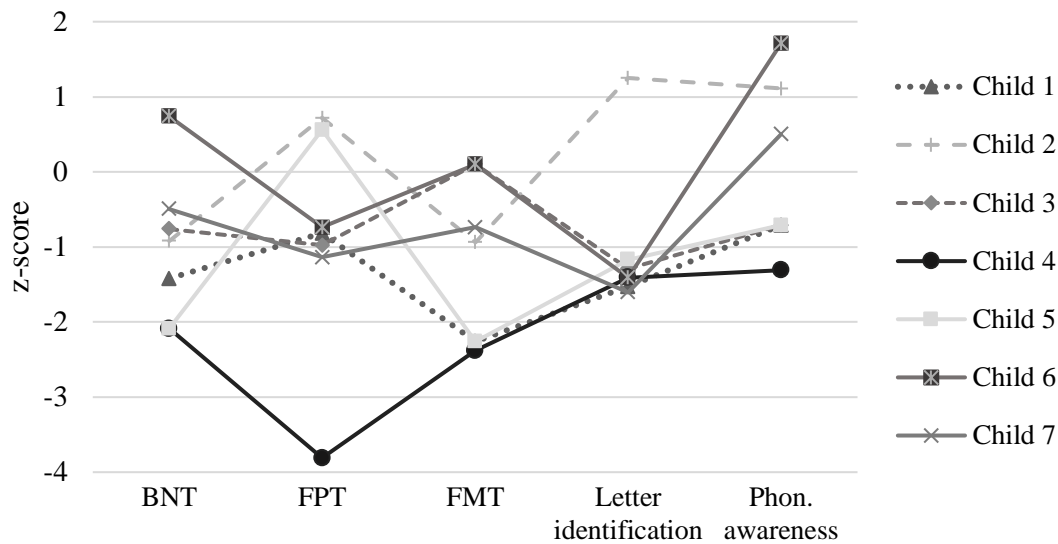


Figure 2. Z-scored language profiles at 5;0 years of children with weak receptive language skills (<-1SD in receptive score of Reynell Developmental Language Scales III) at 2;0 years. BNT = Boston naming test, FPT = Finnish Phonology test, FMT = Finnish Morphology test, Phon. = Phonological.

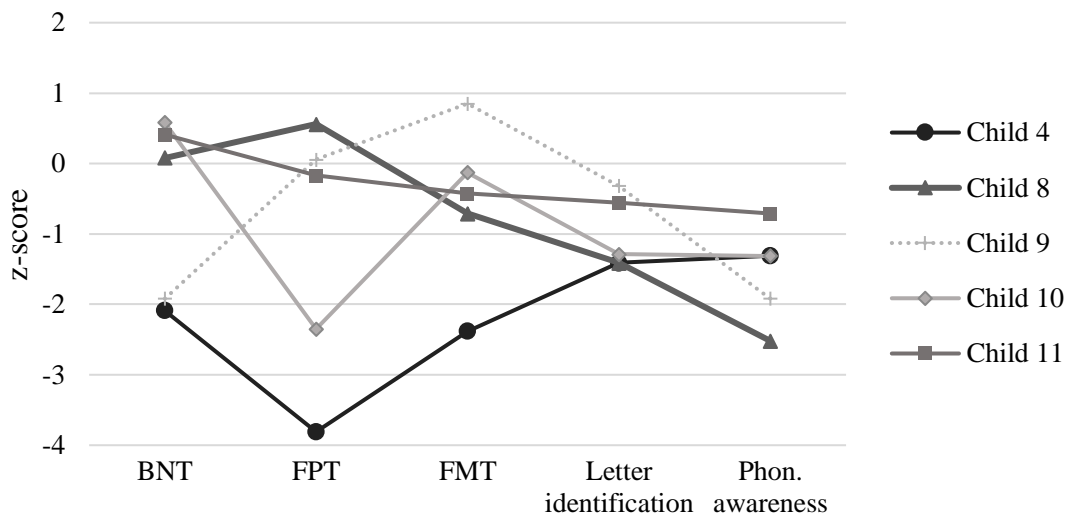


Figure 3. Z-scored language profiles at 5;0 years of children with weak expressive lexical skills (<10th percentile value in Finnish short-form version of the MacArthur–Bates Communicative Development Inventories, Toddler version) at 2;0 years. BNT = Boston naming test, FPT = Finnish Phonology test, FMT = Finnish Morphology test, Phon. = Phonological.