Levels of skills and predictive patterns of reading comprehension in bilingual children with an early age of acquisition

Hilde Lowell Gunnerud[1] · Njål Foldnes[1] · Monica Melby-Lervåg[2]

Abstract
Bilingualism appears to be related to poor language and reading comprehension in the instructional language (L2). However, in the long term, an early age of second language acquisition (AoA) may produce higher levels of second language skills than a later onset of L2 exposure. Most studies on school-aged second language learners (L2 learners) include children with 3–7 years of second language (L2) exposure. We compared 91 early bilingual (Norwegian and a variety of different native languages) 5th graders with 196 monolingual peers on a range of linguistic skills and their relationships with reading comprehension. All bilingual learners in the sample had been exposed to the instructional language by at least the age of 2. The results, using structural equation modelling and latent variables, show that early bilingual learners' vocabulary, listening comprehension, and reading comprehension are significantly lower than those of monolingual learners. Nevertheless, they have similar levels of knowledge of conjunctions and decoding skills and corresponding predictive patterns from linguistic skills to reading comprehension. We discuss the implications of the findings.

Keywords Bilingualism · Early bilinguals · Second language learners · Oral language · Vocabulary · Reading comprehension

✉ Hilde Lowell Gunnerud
Hilde.gunnerud@uis.no

Njål Foldnes
njal.foldnes@uis.no

Monica Melby-Lervåg
monica.melby-lervag@isp.uio.no

[1] The National Centre for Reading Education and Research, University of Stavanger, Stavanger, Norway

[2] Department of Special Needs Education, University of Oslo, Oslo, Norway

 Springer

There are many advantages to mastering two languages. Being bilingual (learning two languages at home or learning the society's language through attendance of school/ECEC institutions) can support children in having close ties with their family and culture. In addition—in a time of globalization—mastering two or more languages may yield benefits such as better job opportunities. However, there may also be challenges related to learning two languages, particularly in regard to the second language (L2) or the instructional language used in schools.

Several studies suggest that L2 preadolescents (8 to 12 years) with 5–7 years of exposure to their L2 still have not caught up with their monolingual peers in terms of oral language skills, such as different aspects of vocabulary, morphology, listening comprehension (Droop & Verhoeven, 2003; Farnia & Geva, 2013) and reading comprehension skills (Herbert et al., 2020; O'Connor et al., 2019). This can seriously hamper L2 learners' ability to achieve academic success and employability (Halle et al., 2012; Han, 2012). The literature is, however, inconclusive in regard to whether early bilingual (AoA by at least 3 years of age) and simultaneous bilingual (bilingual by birth) learners also have lower levels of language and literacy than their monolingual peers.

Here, we report on a study about language and reading development in early bilingual learners that addresses factors that have not been sufficiently accounted for in prior studies, and these factors might be a reason the literature is inconsistent. There are three main factors leading to differences between findings in previous studies. (1) Studies investigate different populations. First, previous studies have mostly focused on children with a late age of acquisition (after 3 years or in samples with a mixed age of acquisition). Thus, we know little about whether early AoA could even out language-related differences between monolingual and bilingual learners. Furthermore, most previous studies have focused on early language and reading and not preadolescent children (approximately 10–11 years old). Since the language gap with monolingual children sometimes narrows over time (Farnia & Geva, 2011), it is important to examine the long-term outcomes of early bilingual learners. Additionally, most studies have investigated the language and reading levels of L2 learners with low socioeconomic status (SES). However, group differences vary across SES levels, and there are smaller language differences between bilinguals and monolinguals with high SES (Oller et al., 2011). Thus, we need more information about early bilinguals from mid- to high-SES backgrounds. (2) Studies differ in the measures they use. For instance, many studies do not measure the age of language acquisition. Studies are vary in regard to the kind of measures used for language and reading and to what extent these measures have been validated. (3) Studies use different statistical methodologies. Finally, most studies have compared bilingual and monolingual children directly, without testing whether they are comparing the same constructs. It is critical to perform valid comparisons; we test this empirically and tailor our methodology to address this issue.

To address the inconsistent findings and fill gaps in previous research, in this study, we focus on preadolescents with an early AoA from middle to high socioeconomic backgrounds using a novel methodology to handle particular measurement issues. In the introduction and discussion, we use the three factors outlined above, i.e., (1) investigation of different populations, (2) use of different measures and (3)

use of different statistical methodologies as a framework, to view and interpret the results of our study in comparison with those of previous studies.

Reading comprehension and its underlying factors

According to the simple view of reading, reading comprehension is the product of decoding skills and linguistic comprehension (Hoover & Gough, 1990). Decoding is the ability to easily and automatically derive a representation from printed input that allows access to the mental lexicon, thus enabling retrieval of semantic information at the word level, while linguistic comprehension is the ability to take semantic information at the word level and derive sentences and discourse interpretation (Hoover & Gough, 1990). Importantly, Hoover and Gouch used listening comprehension and linguistic comprehension as two overlapping constructs. However, in addition to vocabulary, this comprehension construct comprises skills such as an understanding of morphology—knowledge of the smallest meaning-bearing units of language—and conjunctions—an understanding of how an idea in one clause is related to ideas in adjacent clauses (Crosson & Lesaux, 2013; Nagy et al., 2013).

In the early years of learning to read, decoding skills are vital for reading comprehension. However, in approximately 4th grade, a shift in roles occurs because most children then master decoding and reading comprehension start to rely more on language comprehension (Catts et al., 2005). Although the simple view of reading has been validated both in bilingual and monolingual readers (e.g., Droop & Verhoeven, 2003; Hjetland et al., 2020; Lervåg et al., 2018), measuring reading comprehension is more complex. Different tests of reading comprehension are often only moderately correlated and could tap decoding and language comprehension to different extents, depending on how they are constructed (Collins et al., 2018; Keenan et al., 2008).

Levels of language and reading comprehension in simultaneous and early bilingual learners

There are few studies on the long-term outcomes of language levels on early bilingual and simultaneous bilingual children with proper monolingual comparison groups, and their findings are inconsistent. The inconsistency could, as discussed above, partly be explained by their examinations of different populations and the differences in measurement instruments or statistical analyses.

For studies of simultaneous bilinguals, a large cross-sectional study of 5- to 9-year-old identified a large effect size difference in vocabulary across all age groups in favour of monolingual learners (Bialystok & Feng, 2011). A study of 9-year-old Portuguese-English simultaneous bilingual learners also identified large effect size differences in vocabulary levels ($d = 1.20$) and moderate effect size differences in reading comprehension ($d = 0.57$) in favour of monolingual learners (Grant et al., 2011).

Not all studies find differences in reading comprehension between monolingual and simultaneous bilingual learners. Wagner (2004) found no differences in reading comprehension between a sample of diverse simultaneous bilingual learners and their monolingual 10-year-old peers (Wagner, 2004). Although the response format of the reading comprehension tests in Wagner (2004) and Grant et al (2011) both involved open-ended questions, Wagner (2004) assessed reading comprehension of texts from different genres (Progress in International Reading Literacy Study-PIRLS), while Grant assessed reading of narrative texts only [Neale Analysis of Reading Ability-Revised (NARA)]. It has been demonstrated that narrative texts are associated with larger comprehension gaps between students with reading difficulties and typical students (Collins et al., 2018). Moreover, Wagner (2004) examined simultaneous bilingual learners with one native-speaking parent in the instructional language.

For studies comparing levels in early bilingual learners (and not simultaneous bilinguals as discussed above), the results are also mixed. Whereas some studies have found superior vocabulary levels in early bilingual learners or no difference in language and reading levels between language groups (i.e., Hsu et al., 2019; Hwang et al., 2017), others have found full or partial support for large differences in favour of monolingual learners (e.g., Bonifacci & Tobia, 2016; Vernice & Pagliarini, 2018; Kovelman et al., 2008). Again, the inconsistency appears to be caused mainly by the use of samples from different populations, the use of different measurement instruments or differences in statistical analysis.

The studies that show no differences or differences in favour of bilingual learners used rather atypical samples. In Hwang et al. (2017), many bilingual children were enrolled in programmes for gifted students, and in the study by Hsu et al. (2019), bilingual learners were selected based on vocabulary levels above − 1 SD compared to monolingual norms. The results could be different in studies that recruit early bilingual learners based on their AoA without restrictions related to a threshold of already acquired language levels in L2.

In Bonifacci and Tobia (2016), early bilingual learners had poorer reading comprehension but similar levels of listening comprehension. Even though early bilingual learners with diverse L1 statuses were less exposed to L2 than simultaneous bilingual learners in the study by Grant et al. (2011) (AoA by age 4 with assessment of 6- to 12-year-olds; mean age: 8:72), the effect size difference in reading comprehension was comparable ($d = 0.69$ to $d = 0.57$). For listening comprehension, given the poorer levels of reading comprehension, it is surprising that listening comprehension is similar. Nevertheless, the results by Bonifacci and Tobia (2016) might have been influenced by the choice of population and the statistical approach. The authors highlighted the lack of SES information as a limitation to the generalizability of the performance of early bilingual learners. Furthermore, levels were compared across sum scores.

Hence, our literature review indicates that when language and literacy levels are compared with sensitive tests and across equal groups, bilingual learners with low AoA appeared to have lower levels of vocabulary, reading, and listening comprehension. The level of knowledge of conjunctions has, however, not yet been examined in bilingual children with early AoA. Studies of Dutch–Turkish and Dutch–Moroccan

L2 8- to 9-year-old show that they lag behind their monolingual peers in their use of conjunctions when they have to read a text. The group difference in favour of monolingual learners is medium to large (Droop & Verhoeven, 2003).

Predictive patterns for reading comprehension and its underlying factors in early bilingual children

Few studies have examined predictive patterns from linguistic skills to reading comprehension in early or simultaneous bilingual children. Only Grant et al. (2011) examined predictive patterns from linguistic skills to reading comprehension in pre-adolescent children. They studied 3rd-graders' vocabulary and decoding skills in relation to reading comprehension and found that the pattern differed across language groups. Both decoding and vocabulary predicted reading comprehension in simultaneous bilingual children, yet only decoding significantly predicted reading comprehension in English-speaking monolingual children.

The results from previous studies are inconsistent in regard to whether some components of language comprehension are more strongly related to reading comprehension than others in bilingual children. A meta-analysis synthesized correlational studies between different L2 aspects and L2 reading comprehension in bilingual children (Jeon & Yamashita, 2014). The results showed correlations similar in size for reading comprehension with vocabulary ($r=0.79$), listening comprehension ($r=0.77$), and morphological skills ($r=0.61$). For decoding, the relationship with reading comprehension appears somewhat weaker than for the other aspects of language ($r=0.56$) (Jeon & Yamashita, 2014). However, again, this could be because the studies investigated children from different populations (e.g., different age groups).

Regarding primary studies, some have shown that the predictive strength of L2 skills is similar across language groups (Babayiğit, 2015; Verhoeven et al., 2019); for instance, Babayiğit (2015) indicated that the direct path coefficients from oral language (the latent variables of sentence repetition, verbal working memory, and vocabulary) and decoding are comparable across English-diverse L1 learners and their monolingual peers. On the other hand, Proctor and Louick (2018) concluded in their review, based on preliminary support, that language skills are a stronger predictor of reading comprehension for bilingual (versus monolingual) children. Thus, if true, targeting language skills in instructional language could be critical to improving reading comprehension in bilingual children.

Regarding the individual contribution of different language comprehension predictors to reading comprehension, most studies on L2 3rd to 7th graders have found L2 vocabulary to uniquely predict L2 reading comprehension over and above other linguistic skills (Burgoyne et al., 2011; Hutchinson et al., 2003; Kieffer, 2012; Proctor et al., 2012). There is also support for listening comprehension uniquely predicting L2 reading comprehension over other language comprehension constructs in bilingual children in this age group (Burgoyne et al., 2011; Hutchinson et al., 2003; Kieffer, 2012; Proctor et al., 2012). However, the contribution from listening comprehension sometimes overlaps with vocabulary skills or is not present at all when

controlling for other L2 linguistic skills (Burgoyne et al., 2011; Hutchinson et al., 2003; Kieffer, 2012).

Some studies have shown that knowledge of conjunctions can predict L2 learners' reading comprehension beyond vocabulary breadth (Crosson & Lesaux, 2013; Fraser et al., 2021; Rydland et al., 2012). After controlling for vocabulary, knowledge of conjunctions predicted concurrent reading comprehension in English-diverse L1 4th graders and mediated reading comprehension through vocabulary (Fraser et al., 2021). Additionally, studies have shown the relationship between knowledge of conjunctions and reading comprehension to be weaker for English–Spanish learners than for monolinguals, which has led researchers to suggest that readers with high L2 proficiency are the only ones who fully benefit from their conjunction skills (Crosson & Lesaux, 2013). However, the individual contribution of knowledge of conjunctions to reading comprehension seems to also be dependent on the measure used to assess L2 reading. Knowledge of conjunctions sometimes explains the largest proportion of Norwegian-diverse L1 learners' L2 reading comprehension of some texts, but when reading comprehension is assessed by other texts, it is often found to be redundant after controlling for vocabulary (Rydland et al., 2012).

The present study

Thus, there are few studies on the long-term language and literacy outcomes of simultaneous or early bilingual children. Furthermore, most involve children with lower SES.

Thus, the research questions are as follows:

(1) To what extent do bilingual 5th graders with an AoA in the instructional language from birth to 2 years old have levels of language and reading comprehension skills similar to those of their monolingual peers across different aspects of language and reading?
(2) Are the predictive patterns for language comprehension, decoding skills, and SES in relation to reading comprehension the same for bilingual and monolingual children?

Method

Participants

We recruited 196 monolingual (mean age 18.52 months, N girls = 116) and 91 bilingual children (mean age 18.54 months, N girls = 42) to participate in this study. We obtained ethical approval from the Norwegian Social Science Data Service and collected informed parental consent. The majority of the sample had middle to high SES (above 3 years of college); the SES was at the same level in both groups. Sixty of the bilingual children were simultaneous bilingual children with one native Norwegian-speaking parent, while 31 of the children had two bilingual parents with an

AoA of at least age two. The largest language groups were English ($N=22$), German ($N=14$), French (N=6), Kurdish (N=5), Dutch (N=4), Turkish (N=4), Arabic (N=3) and Polish (N=3), with 31 languages represented. The participants were a subsample of children enrolled in the longitudinal Project the longitudinal, interdisiplinary Stavanger Project, The Learning Child. At the onset of the project, most parents of bilingual children provided information on which languages they used in interactions at home during toddlerhood (N=76). Most parents of L2 learners only interacted in L1 at home, while the majority of the simultaneous bilingual learners had some level of Norwegian exposure at home. See Table S4 in the supplementary materials for more information.

The Norwegian language and context

Norwegian is a Germanic language; the lexicon is predominantly Germanic but contains many words from other languages, both Germanic and non-Germanic. The Norwegian morphology is more complex than that of English. In addition to containing both regular and irregular verb classes, verbs are inflected by mood and tense in Norwegian (Simonsen & Bjerkan, 1998). Depending on the dialect, nouns have two or three classes and are inflected for definiteness, while adjectives are inflected for number, gender, and definiteness (Simonsen et al., 2013). For reading comprehension, Norwegian children learn to decode fluently earlier than English-speaking children; hence, linguistic variables play the dominant role in reading comprehension at an earlier timepoint (Lervåg & Aukrust, 2010).

The Kindergarten Act does not regulate any special rights for L2 learners to become fluent in Norwegian (Ministry of Education and Research, 2005). According to the Norwegian framework plan for content and tasks, early childhood education centre (ECEC) staff are responsible for providing a varied and good language environment and for working systematically to promote every child's communication and language skills (Kunnskapsdepartementet, 2017). The ECEC institutions in Norway follow a social pedagogic tradition regulated by the Norwegian framework plan for content and tasks (Kunnskapsdepartementet, 2017). The framework focuses on the importance of introducing playful activities that encourage learning through formal and informal learning situations rather than a curriculum with set learning goals for different age groups. Teachers are responsible for the early identification of children in need of extra support, the initiation of activities that would promote the progression of language skills for these children and the constant evaluation of the effect of the activities on the children's language development (Kunnskapsdepartementet, 2017). Children start school at the age of 6.

Measures

We tested the children using a wide range of language tasks, as well as reading comprehension and decoding skills.

We examined *reading comprehension* using a Norwegian adaptation of the NARA (Neale, 1997). The NARA was translated and adapted to Norwegian by

a team of linguists, and it has frequently been used to gauge reading comprehension in large-scale longitudinal studies of Norwegian children (e.g., Lervåg et al., 2018). The test contains six texts of increasing length and complexity. If the children erroneously decoded a word, the correct word was presented to them. The children were asked questions about the texts immediately after reading them. If a child could not correctly answer four consecutive questions related to one of the texts, the test was stopped, and all subsequent items were scored as 0. The alpha reliability was 0.82.

We tested *listening comprehension* with passages taken from the Norwegian adaptation of the NARA (Neale, 1997). Six texts were read aloud to the child, and each text was followed by questions related to it. If a child could not correctly answer four consecutive questions related to one of the texts, the test was stopped, and all subsequent items were scored as 0. The alpha reliability was 0.94.

We tested knowledge of *conjunctions* with a Norwegian translation of the test developed by Droop and Verhoeven (2003). The test involves listening to two cloze texts and filling in the conjunctions (*in spite of* and *in contrast to*) using a multiple-choice format. One such example is that the children were asked to identify which of the 4 alternatives (because, although, before, during) would fit with the following sentence: Put on your socks……. you put on your shoes. To choose the correct alternative, the child had to understand both the meanings of the parts of the sentences and the relationship between them. The difficulty level of the conjunctions varies in both the Norwegian and English versions of the test, but how it varies differs. Direct matching of items to word familiarity levels was impossible across languages. Word familiarity levels were therefore matched on the category level (e.g., contrastive conjunctions) instead of the item level. The test was group-based, administered in a pen-and-paper format, and accompanied by a verbal presentation of the sentences and the response alternatives to reduce the likelihood that the children's decoding and reading comprehension would impact the results. The alpha reliability was 0.63.

We measured *vocabulary* with the vocabulary subtest of the WISC-4 battery (Wechsler, 2003). On this subtest, children are asked to provide an explanation of a verbally presented word. The scoring was performed in line with the manual, scoring the child's description with 0, 1, or 2 points depending on the quality and accuracy of the description. The alpha reliability was 0.73.

Morphological knowledge was assessed with a version of the test used in the study by Brinchmann et al. (2016), which was supplemented with additional items of increasing difficulty to provide normally distributed data for 5th graders. This was also a group-based cloze test in pen-and-paper format, with additional verbal support to prevent the influence of the child's decoding skills on the test results. The children were orally presented with a sentence that included a nonword and were asked to describe the meaning of this nonword within a multiple-choice format. The nonword comprised two meaningful morphological items (a derivational morpheme in combination with a prefix or suffix) and was interpretable if the core meanings of the morphological items were combined and understood. For example, the children were asked to identify the alternative (scared, truly tough, tired or fearless) that explained the morphological nonword in the sentence "On his way home, Andy felt

unbrave". The alpha reliability calculated across 17 items was 0.66 but dropped to 0.52 after the removal of variant items.

We tested *decoding* with a word chain test (Høien & Tønnesen, 2008), which is a Norwegian instrument resembling the Test of Silent Word Reading Fluency (TOSWRF) (Jacobson, 1993). The test was administered in pen-and-paper format on a group level with a time limit (4 min). The test contains 60 chains of high-frequency words, where four words are presented together in a continuous string of letters. The children were asked to mark where one word ended and the next began. Each word chain in which all marks were correctly placed was awarded 1 point. The test–retest reliability was 0.84 (Høien & Tønnesen, 2008).

The parental questionnaire provided information on the nationality of the bilingual children's parents, which language they considered to be their native language, and at what age their child first attended ECEC institutions.

We used *parents' education level* as an indicator of SES and obtained it through a question with 4 response options (high school, vocational education, 3 years of college, and more than 3 years of college). Because few parents were in the categories of high school and vocational education, we collapsed the two lower-SES categories. Thus, we used three categories of SES for further analysis.

Procedure

Research assistants or the first author tested the children on vocabulary and the two NARA tests separately in a quiet room at their school. The test order was fixed, and the test took 1 h on average. The group-administered tests were conducted by the children's own teachers after they had attended a course on how to administer the tests. The teachers were instructed to spend time on the first items on the test to provide support and ensure that all children understood the task. No formal testing began until the teacher was certain that all children understood the test format. The teacher presented one question, and all pupils were asked to raise their hand to signal when they were ready for the next question.

Data analysis

Measurement invariance analysis

We tested invariance to ensure that differences across groups represented true differences across language groups rather than a comparison of skills across different constructs. Given the categorical-ordinal nature of the items (i.e., correct/incorrect), we used multigroup CFA based on polychoric correlations to evaluate measurement invariance. We performed model estimation using the WLSMV estimator in Mplus version 8.4 (Muthén & Muthén, 1998). In total, we explored five linguistic factors to ensure that all items loaded on the invariant latent variables.

We then used the invariant latent variables, along with a manifest variable of decoding, to examine the predictive patterns of linguistic variables and decoding in relation to reading comprehension. In line with Brown (2015), we conducted

preliminary tests to identify items not recommended for inclusion in the latent variables before running formal invariance tests. First, we removed items that did not have significant factor loadings on the overall latent variable. Additionally, we removed items that were not significantly related to the latent variables across the different language groups and items with negative factor loadings. Before conducting formal invariance tests, we investigated the overall model fit and the model fit for the two language groups separately to ensure an acceptable model fit. In some cases, we had to adjust the model to obtain an appropriate model fit for both groups. When the sample size is moderate (as in this case), items with little variance contain little information. This can produce a disproportionate amount of zero cell frequencies in the observed contingency table, which in turn could lead to bias in the polychoric correlations and pose a threat to inference in the CFAs (Brown & Benedetti, 1977; Olsson, 1979). Therefore, we removed items with limited variance.

Performing a multigroup analysis in Mplus with categorical data requires an equal number of categories across groups. We deleted problematic items in cases in which the item contained an unequal number of categories for the two groups. We then applied model fit statistics to ensure that the model being tested for measurement invariance had an acceptable fit. Acceptable or good model fit is typically defined as RMSEA values below 0.8 or 0.6, respectively, and CFI and TLI above 0.90 and 0.95, respectively (Hu & Bentler, 1999). We thereafter investigated formal measurement invariance.

Measurement invariance involves testing a sequence of nested models and assessing in each step whether the imposed constraints are in line with the data. We tested four models—referred to as the configural, metric, scalar, and strict models—using theta parameterisation (Millsap, 2012). We used modification indices to identify variant factor loadings, thresholds, and residual variances. We deleted variant items to create comparable constructs. We followed the recommendations of Sass et al. (2014) and used chi-square difference tests to formally detect invariance across nested models when analysed with the estimator WLSM, and we reported the RMSEA and CFI for transparency.

Latent means and regression patterns across groups

We studied differences in latent variable means by comparing the means across groups in the strict model. In this model, we constrained the factor loadings and thresholds to be equal between the two groups and the fixed residuals to 1 in the bilingual group while freely estimating them in the monolingual group. The invariance testing made such a comparison meaningful since differences in group levels could then be ascribed to true differences in performance, not to comparisons of skills across different constructs.

Given the complexity of our measurement models, combined with the categorical-ordinal nature of the items and relatively small sample sizes, model estimation could be complicated because of nonconvergence and unstable estimates. We therefore decided to parcel items to obtain a more robust and simpler model estimation. A parcel is an aggregated level indicator comprising the sum scores of

two or more manifest variables. Parcels can be used in SEM when the underlying nature and dimensions of such items are known (Little et al., 2002). Parcelling significantly reduced the number of parameters to be estimated while maintaining the content validity of our latent variables.

We performed parcelling systematically by replacing at least five items with a single parcel item containing their mean score. We included only items that were invariant in the parcel items, which ensured that the parcel items would also be invariant. The resulting parcels had at least six categories each. Since the parcels had symmetrical distributions (see the output of the parcel distributions on the project page of the Open Science Framework: https://osf.io/d8myc/), we treated them as continuous variables. This approach facilitated model estimation since we then did not need polychoric correlations but rather used the observed Pearson correlations. We treated only the three-level indicators of the SES of fathers and mothers as truly categorical by employing polyserial and polychoric correlations.

To examine regression patterns across groups, we tested the model in Fig. 1 but without the morphological knowledge variable. The model examining the prediction of reading comprehension is more complex than the individual measurement models used for latent mean and variance testing. To address missing data, we employed multiple imputation. We combined the results of model estimation in each imputed dataset using the R package semTools (Team, 2019). This change in software was necessary to analyse multiple datasets using WLSM as an estimator. For transparency, we have reported our R code with accompanying output via the Open Science Framework at https://osf.io/d8myc/. Notably, we also conducted the above analysis using a different statistical approach, whereby we combined the item factor score regressions with multiple imputation. This procedure led to the same conclusions as those obtained with the SEM approach described above.

Fig. 1 Theoretical model to be tested via multigroup SEM in monolingual and bilingual children. Parcel items used as indicators for each latent variable were not included for the simplicity of presentation. Only invariant variables will be included in the final model
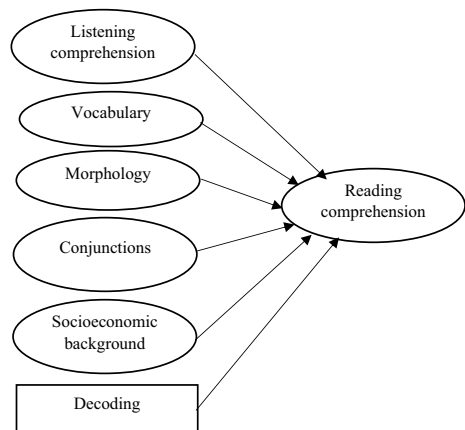
Table 1 Min, max, mean, SD, skewness, and kurtosis for the variables SES, decoding, reading comprehension (RC), listening comprehension (LC), vocabulary, knowledge of conjunctions, and morphology for the monolingual and bilingual groups

Variable	Monolingual learners			Bilingual learners		
	Min/Max	Mean (SD)	Skewness/Kurtosis	Min/Max	Mean (SD)	Skewness/Kurtosis
SES mother	1/3	2.52 (.70)	− 1.13/− .85	1/3	2.37 (0.80)	− .76/− .97
SES father	1/3	2.31 (0.83)	− .64/− 1.25	1/3	2.37 (.082)	− .80/− 1.04
Decoding	5/59	31.47 (9.91)	.12/− .04	9/49	30.12 (8.95)	.10/− .46
Reading comprehension	1/28	12.69 (6.09)	.71/− .11	0/25	9.51 (5.08)	.92/.93
Listening comprehension	2/33	16.23 (8.31)	.04/− 1.23	0/33	12.72 (8.66)	.25/− .78
Vocabulary	0/21	10.31 (4.15)	.42/− .16	0/20	8.26 (3.8)	.61/.78
Conjunctions	1/11	7.24 (1.97)	− .06/− .24	1/11	6.91 (2.29)	− .32/.51
Morphology	0/5	4.16 (1.04)	− 1.21/1.08	0/5	3.87 (1.27)	− 1.04/.30

The different latent variables consist of sum scores based exclusively on the invariant items of these variables

Table 2 Correlations between decoding, reading comprehension (RC), listening comprehension (LC), vocabulary, knowledge of conjunctions, and morphology

Variable	1	2	3	4	5
1 Decoding	1				
2 RC	.21**	1			
3 LC	.08	.69**	1		
4 Vocabulary	.24**	.48**	.49**	1	
5 Conjunctions	.25**	.46**	.31**	.49**	1
6 Morphology	.10	.14*	.16**	.20**	.25**

Correlations between sum scores of different variables created by merging invariant items only

*Indicates $p < .05$. **Indicates $p < .01$

Results

Descriptive statistics

We calculated descriptive statistics using SPSS, version 25. Table 1 displays the mean, SD, minimum, maximum, skewness, and kurtosis values for all manifest variables. Table 2 depicts the correlations between the variables. Table 1 shows that all variables (with the exception of morphology) were normally distributed. As shown in Table 2, all language variables and reading comprehension were correlated. Decoding skills were only correlated with knowledge of conjunctions, vocabulary, and reading comprehension; decoding skills were not correlated with listening comprehension or morphology.

Confirmatory factor analysis: overall and for each group

We first fitted reading comprehension as a one-factor model, with correlated items on the text level. We then removed nonsignificant items from the model. We did not assume any correlation between items in the other CFA models. Table 3 presents model fit statistics for both groups for the final CFA models.

As outlined in Table 3, the fit indices for the CFAs of reading comprehension and listening comprehension indicated that the model had good fit in both language groups. The model fit indices for the CFA of conjunctions differed slightly between language groups, with a somewhat better fit for the bilingual group. This finding suggests that there may be some differences in CFI structure between the two groups. However, since the model fit indices were within acceptable ranges for both groups, we proceeded with invariance testing.

The primary analysis of the vocabulary variable was more challenging. After we removed all items that could reduce model fit, the overall model fit was good (χ^2 (104) = 124.93, p = 0.079, RMSEA = 0.026, CFI = 0.977), accompanied by an acceptable model fit of the monolingual group [χ^2 (104) = 107.31, p = 0.392, RMSEA = 0.013, CFI = 0.995]. The model fit for the bilingual group did, however, indicate a mismatch between the data and the model [χ^2 (104) = 143.43, p = 0.006, RMSEA = 0.065, CFI = 0.853]. The modification indices for the configural model suggested a correlation between two items for the bilingual group but not for the monolingual group. Since the correlations between these items differing across groups were a hindrance for configural invariance, we excluded the most difficult item. In line with suggestions in the modification indices, we excluded some additional items to identify an invariant model for the two language groups. The model fit for each of the two language groups showed a difference in fit between the two groups, where the CFI model fit index for the bilingual group was on the borderline for what is considered to be acceptable (RMSEA: 0.052, CFI: 0.895). In contrast, the model for the monolingual group had an excellent fit (RMSEA: 0.017, CFI: 0.993). However, the χ^2-difference test that constrained factor loadings, thresholds, and residuals to be equal between the groups was not significant; therefore, we concluded that the model fit was adequate for both groups.

Factor analysis of morphological knowledge indicated a multifactorial structure, which was supported by the identification of a three-factor model in an exploratory factor analysis (EFA) [χ^2 (187) = 202.626, p = 0.206, RMSEA = 0.018, CFI = 0.979].

Table 3 Model fit statistics across language groups for the final models for each construct

Model fit statistics for	Monolingual group				Bilingual group			
	χ^2	p	RMSEA	CFI	χ^2	p	RMSEA	CFI
Listening comprehension	646.86	0.00	0.034	0.983	540.42	0.33	0.017	0.996
Reading comprehension	254.78	0.10	0.025	0.986	269.70	0.03	0.045	0.933
Vocabulary	57.22	0.36	0.017	0.993	67.20	0.11	0.052	0.895
Conjunctions	66.82	0.13	0.036	0.905	59.55	0.28	0.034	0.950

Invariance testing of the factor containing most of the items for the morphology variable denoted that few factor loadings were significant in both groups, resulting in a reduction of 20 of 25 test items. Since 73.5% of all participants scored within the two highest performance levels, we considered the remaining test items to be unsuitable for identifying possible differences across groups. We thus excluded morphology from further analysis.

Invariance analysis

Based on the previous examination, we tested invariance for the variables of reading comprehension, listening comprehension, vocabulary, and knowledge of conjunctions. Table 4 presents the results.

As shown in Table 4, apart from vocabulary, all chi-square tests for the configural model were significant. However, chi-square tests are sensitive to sample size, which has led to a recommendation to rely on other fit statistics that are often acknowledged to be more reliable (Hooper et al., 2008). Other fit indices for the configural model of reading comprehension (RMSEA=0.033, CFI=0.969), listening comprehension (RMSEA=0.027, CFI=0.989), and knowledge of conjunctions (RMSEA=0.035, CFI=0.926) were all acceptable; hence, we concluded with configural invariance for all variables. Chi-square tests between the nested models showed no significant differences for the configural, metric, scalar, or strict models for any of the variables. We hence concluded with strict invariance across all tested variables.

Differences in SES, decoding skills, and linguistic constructs across groups

First, the robust Mann–Whitney U test did not lend support for any differences in SES between monolingual and bilingual children ($p=0.26$ for the SES of mothers and $p=0.51$ for the SES of fathers). We investigated group differences in the manifest variable of decoding via t test; there were no significant differences between the decoding skills of the monolingual ($M=31.47$, $SD=9.91$) and bilingual learners ($M=30.12$, $SD=8.95$, effect size $d=-0.14$) [$t(283)=-1.10$, $p=0.52$].

Comparisons of latent means between the monolingual and bilingual groups require scalar invariance. Here, we had full scalar invariance for the vocabulary, knowledge of conjunctions, listening comprehension, and reading comprehension variables. There were significant differences in the means for reading comprehension, listening comprehension, and vocabulary, while the means for knowledge of conjunctions were equal across groups. Table 5 depicts the differences in factor means. The differences in means are standardized and can be interpreted as group differences measured by Cohen's *d*. Table S1 (online supplemental material) shows the differences in variance.

Comparison of predictive patterns for reading comprehension between the two groups

We first tested the regression pattern for listening comprehension, vocabulary, knowledge of conjunctions, decoding skills, and SES in relation to reading in the

Table 4 Test of measurement invariance across language groups for reading comprehension, listening comprehension, vocabulary, and knowledge of conjunctions

	χ^2	df	diff χ^2	diff df	RMSEA	CFI
Reading comprehension						
Overall sample	273.01**	227			0.027	0.984
MI across groups						
Configural	526.94**	454			0.033	0.969
Metric	544.34**	476	24.89	22	0.032	0.971
Scalar	571.49**	503	27.706	27	0.031	0.971
Strict	556.88**	480	21.979	23	0.033	0.967
Listening comprehension						
Overall sample	795.10***	561			0.038	0.979
MI across groups						
Configural	1167.67**	1055			0.027	0.989
Metric	1175.28**	1088	32.72	33	0.024	0.991
Scalar	1210.64**	1121	41.68	33	0.024	0.991
Strict	1192.90**	1092	31.73	29	0.025	0.990
Vocabulary						
Overall sample	63.155	54			0.024	0.985
MI across groups						
Configural	98.175	88			0.028	0.979
Metric	99.416	98	6.158	10	0.010	0.997
Scalar	129.668	119	32.543	0.052	0.025	0.978
Strict	120.679	108	10.004	11	0.029	0.974
Conjunctions						
Overall sample	52.211	54			0.000	1.000
MI across groups						
Configural	126.07	109			0.035	0.926
Metric	129.20	119	6.791	11	0.025	0.958
Scalar	143.97	130	16.811	11	0.028	0.942
Strict	135.81	118	9.052	12	0.033	0.926

MI measurement invariance

**$p < .05$, ***$p < .001$

Table 5 Standardized differences in factor means across language groups for reading comprehension (RC), listening comprehension (LC), knowledge of conjunctions, vocabulary, and morphology

Construct of comparison	Standardized difference in factor means	SE	*P*
Reading comprehension	0.78	0.145	0.009
Listening comprehension	0.60	0.191	0.002
Conjunctions	0.34	0.206	0.102
Vocabulary	0.74	0.210	<0.001

The factor mean of the bilingual learners was set to zero and used as a reference group. A positive difference in means in the standardized model indicates an advantage for monolingual learners

overall sample (Fig. 2). This approach provided an excellent model fit ($N = 287$, χ^2 [80.0] = 53.996, $p = 0.989$, RMSEA = 0.000, CFI = 1.000, TLI = 0.937). Listening comprehension had the strongest relationship with reading comprehension ($\beta = 0.52$, $SE = 0.07$, $p = 0.000$), followed by knowledge of conjunctions ($\beta = 0.428$, $SE = 0.19$, $p = 0.03$). When controlling for knowledge of conjunctions and listening comprehension, vocabulary and SES were not significant. Decoding was only marginally related to children's reading comprehension skills ($\beta = 0.003$, $SE = 0.001$, $p = 0.02$). Notably, the intercorrelations among parcels of latent variables were, unsurprisingly, large, while the correlations between parcels across the examined constructs were moderate (see Table S3, online supplemental material).

In the final step, we constrained all regressions to be equal between the two groups and compared them to a model in which the regressions were freely estimated. The χ^2-difference test showed no significant differences between models [F (5.0) = 0.555, $p = 0.734$]. Thus, the two groups had equal strength in predictions from decoding and linguistic skills in relation to reading comprehension.

Discussion

This study reveals interesting findings about reading and language in preadolescent bilingual children with early AoA. The first research question investigated to what extent early bilingual 5th graders have similar levels of language and reading comprehension skills to their monolingual peers. The results revealed that early bilingual learners with (primarily) middle to high SES have levels of decoding skills and knowledge of conjunctions similar to those of their monolingual peers. However, there are moderate to large differences in favour of monolingual learners in listening comprehension, reading comprehension, and vocabulary. The second research question asked whether the predictive patterns for language comprehension and decoding skills in relation to reading comprehension were the same for bilingual and monolingual children. The size of the predictive paths of linguistic skills and decoding to reading comprehension were equal across groups. This finding implies that linguistic skills do not play a more critical role in early bilingual children than in monolingual readers. Listening comprehension, knowledge of conjunctions, and

Fig. 2 Regression model predicting reading comprehension in bilingual and monolingual children. Model fit (N = 287). χ^2 [80.0] = 53.996, p = 0.989, RMSEA = 0.000, CFI = 1.000, TLI = 0.937, **Indicates $p < .01$. *Indicates $p < .05$. **Parcels** were not included for simplicity
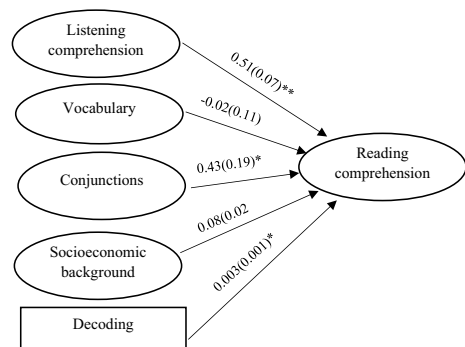
decoding were the only constructs related to reading comprehension, with listening comprehension explaining the largest proportion of reading comprehension.

Levels of language and reading comprehension skills in early bilingual children compared to their monolingual peers

The results support prior findings that bilingual children have poorer language comprehension than monolingual children (Bialystok, 2009; Melby-Lervåg & Lervåg, 2014). Since the participants in this study were exposed to Norwegian from at least the age of two and had middle- to high-SES backgrounds, one should perhaps expect potential group differences to be evened out. However, this was not the case; group differences in favour of monolingual children were moderate to large (ds = 0.60–0.78) and were comparable in size to those in some studies of early bilinguals, albeit somewhat smaller than in studies on L2 learners with 5–7 years of L2 exposure (with a Cohen's d of approximately 1) (Cummins, 1984; Hakuta et al., 2000; Melby-Lervåg & Lervåg, 2014).

Our results concur with the findings of most prior studies on early or simultaneous bilingual learners since these studies also show lower levels of reading comprehension in bilinguals than in their monolingual peers (Bonifacci & Tobia, 2016; Grant et al., 2011; Kovelman et al., 2008). However, the current study is inconsistent with Wagner's (2004) study on simultaneous bilinguals. One reason could be that the above studies used different measures of reading comprehension. Both the present study and Grant et al. (2011) assessed reading compression with the NARA, while the study of Wagner (2004) used the PIRLS. Assessment of reading comprehension of expository genre (NARA) produces larger gaps in reading comprehension for struggling readers than some of the texts in the PIRLS. Additionally, in contrast to the PIRLS, in the NARA, the test administrator informs the child about the correct word when the child fails to decode it correctly (Martin et al., 2021; Neale, 1997). Hence, reading comprehension assessed by the NARA most likely relies more heavily on the child's linguistic skills than does the PIRLS test used in Wagner's (2004) study.

For listening comprehension and vocabulary, our findings are inconsistent with the work of Bonifacci and Tobia (2016), where the two groups had similar listening comprehension levels, and Hwang et al. (2017), where early bilingual learners outperformed their monolingual peers in terms of vocabulary. However, the early bilingual learners in Bonifacci and Tobia (2016) had a higher AoA than the sample in the present study and were tested on listening comprehension at a younger age. Hence, factors other than AoA might cause nonsignificant differences in listening comprehension.

Methodological differences could be another plausible reason for the differences in the study results. Bonifacci and Tobia (2016) compared levels of listening comprehension using sum scores rather than testing for invariance between groups. This might lead to bias because there could be items favouring one group rather than the other. In the present study, even after the removal of items associated with test bias in favour of one of the language groups, a comparison across sum scores and

the latent means produced different results (0.42 d and 0.60 d). Additionally, samples from different populations might also play a role: while we compared language levels across groups from similar SES backgrounds, Bonifacci and Tobia (2016) recruited participants from the same neighbourhoods but did not examine potential SES differences between early bilingual and monolingual participants. Furthermore, samples from different populations are also a plausible explanation for the differences in results between the present study and Hwang et al. (2017) since the latter examined a bilingual sample with a large percentage of gifted participants.

For knowledge of conjunctions, the effect size difference between monolingual and early bilingual learners in our study was smaller than that for the other measures and was not significant ($d = 0.34$, $p = 0.12$). This could be because conjunctions, in contrast to vocabulary, comprise a limited number of words. Nevertheless, prior studies of children with later AoA have revealed that bilingual children have challenges in regard to knowledge of conjunctions (Droop & Verhoeven, 2003). These conflicting findings suggest, not surprisingly, that bilingual children and early bilingual learners have different profiles and different needs for intervention.

Regarding morphology, the Cronbach's alpha was low, and a large number of test items were invariant; hence, the comparison of morphological levels on latent means across language groups was invalid. With the exception of Droop and Verhoeven's (2003) study of L2 learners, no other studies have used morphology measures that are invariant across groups to examine differences in levels or predictive patterns; rather, they have used sum scores where the invariance assumption is assumed but not tested. Although using sum scores per se does not imply poor quality of the study, using sum scores can imply that items that are rather different are being summed. Thus, the results from studies using sum scores and finding that morphology is a relative strength in bilingual children should be interpreted with some caution (e.g., Barac & Bialystok, 2012; Hsu et al., 2019).

The predictive patterns for aspects of language comprehension and decoding skills in relation to reading comprehension

In this study, the predictive pattern from language skills to reading comprehension was similar for monolingual and early bilingual learners. In contrast, in Grant et al. (2011), vocabulary predicted simultaneous bilingual 3rd graders' reading skills; however, only decoding predicted their monolingual peers' reading comprehension. The bilingual learners in both the present study and Grant et al. (2011) had lower levels of linguistic skills than their monolingual peers. Thus, the difference in results between the two studies is unlikely to arise because the language levels between the bilingual samples differed. Since decoding is the most crucial component in reading comprehension in the first years of gaining reading skills (Hoover & Gough, 1990), the difference across studies is likely due to the examination of different populations. The participants in Grant et al. (2011) were 3rd graders learning to decode in a nontransparent language (English), while the participants in the present study were 5th graders decoding in a semitransparent language.

Regarding the magnitude of the effect of knowledge of conjunctions on reading comprehension, in contrast to Crosson and Lesaux (2013), we found a similar relationship for bilinguals and monolinguals. Crosson and Lesaux (2013) hypothesized that the proficiency level in L2 of bilingual learners was insufficient; hence, their reading comprehension could not fully benefit from their knowledge of conjunctions because the conjunctions were embedded in passages with a high proportion of unknown words. The group difference in vocabulary was, however, much larger in Crosson and Lesaux's (2013) study than in ours (1.64 d versus 74 d). This suggests that early bilingual learners with a large amount of exposure to their L2 over a minimum of 8 years might develop sufficient proficiency in their L2 to fully benefit from their conjunctional skills.

The predictive strength of decoding to reading comprehension in the present study was weak. However, Norwegian has a semitransparent orthography; furthermore, the participants were 5th graders. By this timepoint, typically developed children tend to master the alphabetic principle, which most likely explains why there was only a weak relationship between decoding and reading comprehension in the present study.

Listening comprehension explained 26.01% of the variation in reading comprehension in our study, while vocabulary had no explanatory value after controlling for listening comprehension. This finding is inconsistent with previous research on bilingual language learners. Comparisons between studies are, however, complicated by differences in methodology. Often, studies do not include both vocabulary and listening comprehension as measures (e.g. Grant et al., 2011; Hutchinson et al., 2003) or use sum scores in the analysis (e.g. Burgoyne et al., 2011). Note, however, that the measures in the present study could have inflated the strength of the relationship between listening comprehension and reading comprehension since the NARA listening and reading comprehension tests are based on the same format and are highly similar (except the content of the stories).

Furthermore, even though vocabulary here did not explain variation in reading comprehension, knowledge of conjunctions did (18.49%). This finding is in line with Rydland et al. (2012), who examined the impact of conjunctions on L2 learners' reading comprehension. Thus, the impact of conjunctions on preadolescent children's reading comprehension appears to hold across L2 learners and early bilingual and monolingual readers. Thus, listening comprehension and knowledge of conjunctions have important impacts on reading comprehension in early bilingual preadolescents (8- to 12-year-old bilingual learners with an AoA before 3). Nevertheless, the independent contribution of knowledge of conjunctions to reading comprehension seems to depend on the instrument used to assess L2 reading. Knowledge of conjunctions sometimes explains the largest proportion of L2 reading comprehension but is often redundant when controlling for other vocabulary skills (Rydland et al., 2012).

Practical implications and limitations

Our main finding that early bilinguals have not caught up by 5th grade implies that medium to high SES background, early AoA, and a large amount of exposure to the instructional language across a minimum of 8 years are not sufficient for early bilingual learners to develop language levels comparable to those of their monolingual peers in

all aspects of the instructional language. More longitudinal studies of early bilingual children are needed to examine how this pattern evolves over time. Furthermore, there is a need to create interventions to ensure improvement in language and literacy trajectories for the early AoA subgroup of bilingual learners.

Furthermore, we found nonsignificant group differences for levels of knowledge of conjunctions in early bilingual and monolingual children. Notably, our sample size was moderate; a larger sample would have had more power to detect differences. Regardless, the difference across monolingual and bilingual learners was larger for other examined language aspects, suggesting that those language aspects should be targeted for interventions rather than knowledge of conjunctions.

Regarding the prediction of specific L2 skills in reading comprehension, few studies (including the present study) have examined dimensionality in linguistic constructs. This is an important step in the prediction of reading comprehension to draw solid conclusions regarding the unique contributions of specific L2 skills to reading comprehension. Otherwise, one might rely too heavily on theoretical differences across constructs and falsely assume that there are empirical differences between language aspects when this might not be the case. Regardless of how different L2 aspects are related to reading comprehension, interventions to improve early bilingual children's language comprehension could be implemented before formal reading instruction begins. Intervention studies of young bilingual children of ECEC age show promising results, which suggests that such interventions could change young bilingual children's learning trajectories (Rogde et al., 2016).

Data availability No data are available.

Code availability R codes for the SEM analysis is reported at the Open Science Framework at osf.io/d8myc/.

Declarations

Conflict of interest The authors have no relevant financial or nonfinancial relationships to disclose.

Ethical approval Ethical approvement was obtained from the Norwegian Social Science Data Service.

Consent to participate Parents signed parental consent papers.

Consent for publication Yes.

References

Babayiğit, S. (2015). The relations between word reading, oral language, and reading comprehension in children who speak English as a first (L1) and second language (L2): A multigroup structural analysis. *Reading and Writing, 28*(4), 527–544. https://doi.org/10.1007/s11145-014-9536-x

Barac, R., & Bialystok, E. (2012). Bilingual effects on cognitive and linguistic development: Role of language, cultural background, and education. *Child Development, 83*(2), 413–422. https://doi.org/10.1111/j.1467-8624.2011.01707.x

Bialystok, E. (2009). Bilingualism: The good, the bad, and the indifferent. *Bilingualism: Language and Cognition, 12*(1), 3–11.

Bialystok, E., & Feng, X. (2011). Languageproficiency and its implications for monolingual and bilingual children. In A. Y. Durgunoglu & C. Goldenberg (Eds.), *Language and literacy development in bilingual settings* (pp. 121–138). The Guildford Press.

Bonifacci, P., & Tobia, V. (2016). Crossing barriers: Profiles of reading and comprehension skills in early and late bilinguals, poor comprehenders, reading impaired, and typically developing children. *Learning and Individual Differences, 47*, 17–26.

Brinchmann, E. I., Hjetland, H. N., & Lyster, S. A. H. (2016). Lexical quality matters: Effects of word knowledge instruction on the language and literacy skills of third- and fourth-grade poor readers. *Reading Research Quarterly, 51*(2), 165–180. https://doi.org/10.1002/rrq.128

Brown, M. B., & Benedetti, J. K. (1977). On the mean and variance of the tetrachoric correlation coefficient. *Psychometrika, 42*(3), 347–355.

Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. Guilford Publications.

Burgoyne, K., Whiteley, H., & Hutchinson, J. M. (2011). The development of comprehension and reading-related skills in children learning English as an additional language and their monolingual, English-speaking peers. *British Journal of Educational Psychology, 81*(2), 344–354.

Catts, H. W., Hogan, T. P., & Adlof, S. M. (2005). Developmental changes in reading and reading disabilities. In W. K. InH & A. G. Catts (Eds.), *The connections between language and reading disabilities* (pp. 25–40). LawrenceErlbaum Associates Publishers.

Collins, A. A., Lindström, E. R., & Compton, D. L. (2018). Comparing students with and without reading difficulties on reading comprehension assessments: A meta-analysis. *Journal of Learning Disabilities, 51*(2), 108–123.

Crosson, A. C., & Lesaux, N. K. (2013). Does knowledge of connectives play a unique role in the reading comprehension of English learners and English-only students? *Journal of Research in Reading, 36*(3), 241–260.

Cummins, J. (1984). *Bilingualism and special education: Issues in assessment and pedagogy*. Multilingual Matters.

Droop, M., & Verhoeven, L. (2003). Language proficiency and reading ability in first- and second-language learners. *Reading Research Quarterly, 38*(1), 78–103. https://doi.org/10.1598/RRQ.38.1.4

Farnia, F., & Geva, E. (2011). Cognitive correlates of vocabulary growth in English language learners. *Applied Psycholinguistics, 32*(4), 711–738.

Farnia, F., & Geva, E. (2013). Growth and predictors of change in English language learners' reading comprehension. *Journal of Research in Reading, 36*(4), 389–421. https://doi.org/10.1111/jrir.12003

Fraser, C., Pasquarella, A., Geva, E., Gottardo, A., & Biemiller, A. (2021). English language learners' comprehension of logical relationships in expository texts: Evidence for the confluence of general vocabulary and text-connecting functions. *Language Learning, 71*, 872.

Grant, A., Gottardo, A., & Geva, E. (2011). Reading in English as a first or second language: The case of grade 3 Spanish, Portuguese, and English speakers. *Learning Disabilities Research & Practice (Wiley-Blackwell), 26*(2), 67–83. https://doi.org/10.1111/j.1540-5826.2011.00327.x

Hakuta, K., Butler, Y. G., & Witt, D. (2000). *How long does it take english learners to attain proficiency?* Santa Barbara: The Linguistic Minority Research Institute.

Halle, T., Hair, E., Wandner, L., McNamara, M., & Chien, N. (2012). Predictors and outcomes of early versus later English language proficiency among English language learners. *Early Childhood Research Quarterly, 27*(1), 1–20. https://doi.org/10.1016/j.ecresq.2011.07.004

Han, W.-J. (2012). Bilingualism and academic achievement. *Child Development, 83*(1), 300–321.

Herbert, K. E., Massey-Garrison, A., & Geva, E. (2020). A Developmental examination of narrative writing in EL and EL1 school children who are typical readers, poor decoders, or poor comprehenders. *Journal of Learning Disabilities, 53*(1), 36–47.

Hjetland, H. N., Brinchmann, E. I., Scherer, R., Hulme, C., & Melby-Lervåg, M. (2020). Preschool pathways to reading comprehension: A systematic meta-analytic review. *Educational Research Review, 30*, 100323.

Høien, T., & Tønnesen, G. (2008). *Introduksjonshefte til Ordkjedeprøven*. Logometrica AS.

Hooper, D., Coughlan, J., & Mullen, M. R. (2008). *Structural equation modelling: Guidelines for determining model fit.Electronic journal of business research methods, 6*(1), 53–60.

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing, 2*(2), 127–160. https://doi.org/10.1007/BF00401799

Hsu, L.S.-J., Ip, K. I., Arredondo, M. M., Tardif, T., & Kovelman, I. (2019). Simultaneous acquisition of English and Chinese impacts children's reliance on vocabulary, morphological and phonological awareness for reading in English. *International Journal of Bilingual Education and Bilingualism, 22*(2), 207–223. https://doi.org/10.1080/13670050.2016.1246515

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Hutchinson, J. M., Whiteley, H. E., Smith, C. D., & Connors, L. (2003). The developmental progression of comprehension–related skills in children learning EAL. *Journal of Research in Reading, 26*(1), 19–32.

Hwang, J. K., Lawrence, J. F., & Snow, C. E. (2017). Defying expectations: Vocabulary growth trajectories of high performing language minority students. *Reading and Writing, 30*(4), 829–856. https://doi.org/10.1007/s11145-016-9703-3

Jacobson, C. (1993). *Manual till Ordkedjetestet [The wordchains test; manual]*. Psykologiförlaget.

Jeon, E. H., & Yamashita, J. (2014). L2 reading comprehension and its correlates: A meta-analysis. *Language Learning, 64*(1), 160–212.

Keenan, J. M., Betjemann, R. S., & Olson, R. K. (2008). Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. *Scientific Studies of Reading, 12*(3), 281–300.

Kieffer, M. J. (2012). Early oral language and later reading development in Spanish-speaking English language learners: Evidence from a nine-year longitudinal study. *Journal of Applied Developmental Psychology, 33*(3), 146–157. https://doi.org/10.1016/j.appdev.2012.02.003

Kovelman, I., Baker, S. A., & Petitto, L.-A. (2008). Age of first bilingual language exposure as a new window into bilingual reading development. *Bilingualism: Language and Cognition, 11*(2), 203–223.

Kunnskapsdepartementet. (2017). *Rammeplan for barnehagers innhold og oppgaver*. Kunnskapsdepartementet.

Lervåg, A., & Aukrust, V. G. (2010). Vocabulary knowledge is a critical determinant of the difference in reading comprehension growth between first and second language learners. *Journal of Child Psychology and Psychiatry, 51*(5), 612–620.

Lervåg, A., Hulme, C., & Melby-Lervåg, M. (2018). Unpicking the developmental relationship between oral language skills and reading comprehension: It's simple, but complex. *Child Development, 89*(5), 1821–1838.

Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling: A Multidisciplinary Journal, 9*(2), 151–173. https://doi.org/10.1207/S15328007SEM0902_1

Martin, M. O., von Davier, M., Foy, P., & Mullis, I. V. S. (2021). *PIRLS 2021 assessment design*. TIMSS & PIRLS International Study Centre, Lynch School of Education, Bosten College.

Melby-Lervåg, M., & Lervåg, A. (2014). Reading comprehension and its underlying components in second-language learners: A meta-analysis of studies comparing first-and second-language learners. *Psychological Bulletin, 140*(2), 409.

Millsap, R. E. (2012). *Statistical approaches to measurement invariance*. Routledge.

Ministry of Education and Research. (2005). *The Kindergarten Act (LOV-2005-06-17-64)*.

Muthén, L., & Muthén, B. (1998). *Mplus (Version 8.4)[computer software]*. Author: statmodel.com.

Nagy, W. E., Carlisle, J. F., & Goodwin, A. P. (2013). Morphological knowledge and literacy acquisition. *Journal of Learning Disabilities, 47*(1), 3–12. https://doi.org/10.1177/0022219413509967

Neale, M. D. (1997). *Neale analysis of reading ability—revisited. Second revised British Edition*. GL Assessment.

O'Connor, M., Geva, E., & Koh, P. W. (2019). Examining reading comprehension profiles of grade 5 monolinguals and English language learners through the lexical quality hypothesis lens. *Journal of Learning Disabilities, 52*(3), 232–246.

Oller, D. K., Jarmulowicz, L., Pearson, B. Z., & Cobo-Lewis, A. B. (2011). Rapid spoken language shift in early second- language learning. The role of peers and effects on the first language. In A. Y. Durgunoglu & C. Goldberg (Eds.), *Language and literacy development in bilingual settings.* The Guilford Press.

Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika, 44*(4), 443–460.

Proctor, C. P., & Louick, R. (2018). Development of vocabulary knowledge and its relationship with reading comprehension among emergent bilingual children: An overview. In A. Bar-On, D. Ravid, & E. Dattner (Eds.), *Handbook of communications disorders: Theoretical, empirical, and applied linguistic perspectives.* De Gruyter Inc.

Proctor, C. P., Montecillo, C., Silverman, R. D., & Harring, J. R. (2012). The role of vocabulary depth in predicting reading comprehension among English monolingual and Spanish-English bilingual children in elementary school. *Reading and Writing, 25*(7), 1635–1664. https://doi.org/10.1007/s11145-011-9336-5

RC Team. (2019). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Rogde, K., Melby-Lervåg, M., & Lervåg, A. (2016). Improving the general language skills of second-language learners in kindergarten: A randomized controlled trial. *Journal of Research on Educational Effectiveness, 9*(sup1), 150–170.

Rydland, V., Aukrust, V. G., & Fulland, H. (2012). How word decoding, vocabulary and prior topic knowledge predict reading comprehension. A study of language-minority students in Norwegian fifth grade classrooms. *Reading and Writing, 25*(2), 465–482. https://doi.org/10.1007/s11145-010-9279-2

Sass, D. A., Schmitt, T. A., & Marsh, H. W. (2014). Evaluating model fit with ordered categorical data within a measurement invariance framework: A comparison of estimators. *Structural Equation Modelling: A Multidisciplinary Journal, 21*(2), 167–180.

Simonsen, H. G., & Bjerkan, K. M. (1998). Testing past tense inflection in Norwegian: A diagnostic tool for identifying SLI children? *International Journal of Applied Linguistics, 8*(2), 251–270.

Simonsen, H. G., Kristoffersen, K. E., Bleses, D., Wehberg, S., & Jørgensen, R. N. (2013). The Norwegian Communicative Development Inventories: Reliability, main developmental trends and gender differences. *First Language, 34*(1), 3–23. https://doi.org/10.1177/0142723713510997

Verhoeven, L., Voeten, M., & Vermeer, A. (2019). Beyond the simple view of early first and second language reading: The impact of lexical quality. *Journal of Neurolinguistics, 50*, 28–36.

Vernice, M., & Pagliarini, E. (2018). Is morphological awareness a relevant predictor of reading fluency and comprehension? New evidence from Italian monolingual and Arabic-Italian bilingual children. *Frontiers in Communication, 3*, 11.

Wagner, Å. K. H. (2004). *Hvordan leser minoritetsspråklige elever i Norge? En studie av minoritetsspråklige og majoritetsspråklige 10 åringers leseresultater og bakgrunnsfaktorer i den norske delen av PIRLS 2001*. Høgskolen i Stavanger.

Wechsler, D. (2003). *WISC- IV: Wechsler intelligence scale for children*. Pearson Inc.