

# Bilingualism

## Advantages and Disadvantages in Cognitive Processing, Language and Reading Comprehension

by

Hilde Lowell Gunnerud

Thesis submitted in fulfilment of  
the requirements for the degree of  
PHILOSOPHIAE DOCTOR  
(PhD)



Faculty of Art and Education  
Norwegian Centre for Reading Education and Reading Research  
2021

University of Stavanger  
NO-4036 Stavanger  
NORWAY  
[www.uis.no](http://www.uis.no)

©2021 Hilde Lowell Gunnerud

ISBN:978-82-7644-989-1  
ISSN:1890-1387  
PhD: Thesis UiS No. 575

## Acknowledgements

I still find assessment of bilingual children's language and learning abilities as intriguing as I did with the first case I was assigned in my prior work as a school psychologist adviser. The process of evaluating bilingual children with potential special education needs felt as difficult as solving a puzzle while lacking key puzzle pieces. The key puzzle pieces represent a lack of knowledge of typical bilingual development for children with different exposure to learning experiences and the instructional language. This personal encounter with the consequences of the knowledge gap regarding bilingual children's development motivated me in my work with this thesis. I feel fortunate to have been able to study this important and captivating topic as well as grateful to all the wonderful people who have supported me in the process.

First, I would like to thank my main supervisor, Monica Melby-Lervåg, for first introducing me to the gratification of research. Without your encouragement, I would not have followed my dream of becoming a researcher. I am forever grateful for your excellent professional guidance, enthusiasm and expertise, which provided me with great learning opportunities and the necessary help to see them through. Your kindness and willingness to generously share your knowledge with me are highly appreciated. I am especially thankful that you introduced me to the world of meta-analysis.

Co-supervisor Elin Reikerås, thank you so much for including my PhD project within the frame of The Stavanger Project—The Learning Child. The Stavanger Project laid the foundation upon which my thesis rests. Without your enthusiasm, dedication and work, my thesis would not have been the same. Thank you for generously sharing data with me and for your help and support in developing Study 3. I am also grateful for your encouragement and repeated utterances of belief in my ability to overcome the many challenges that needed to be solved while I worked

on this thesis. Your encouragement has inspired me to work even harder when faced with challenges.

I have also been fortunate to work with outstanding co-authors, Dieuwer ten Braak, Enrica Donolato, Anne Elisabeth Dahle, Njål Foldnes, Elin Reikerås and Monica Melby Lervåg. Thank you for your hard work and expertise, all the important things you have taught me and your consistently positive attitudes in our collaborations. You are all wonderful!

To all the children, teachers, research assistants and workers in the school administration of Stavanger municipality, thank you so much for your participation in this project. Without your collaboration, this research would have been impossible. A special appreciation to Åse Kari Wagner, prior leader of the National Educational and Reading Centre, for providing extra funding for data collection in Study 3.

Finally, I would like to thank my family. To my mother and father, thank you for always encouraging me and convincing me that I could become whatever I wanted if I worked hard enough. You have always encouraged me and have also been good role models, proving that it is never too late to go back to school. Without your support, I would not have achieved this! I would also like to express my warm gratitude to my husband, Helge, for taking such good care of our two children, Iselin and Andreas, while I spent far too many hours at work. Without you, the work on this thesis would have lasted an eternity. Finally, to Iselin and Andreas, thank you for your patience, warm support and many necessary breaks from the PhD work.

Hilde Lowell Gunnerud

Stavanger, August 2020

## Summary

The overall aim of this study is, as suggested by Bialystok (2009), to investigate whether bilingual learners have an advantage in executive functions and a disadvantage in language compared to monolingual learners. In addition, the thesis examines whether the theory holds true for different groups of bilingual learners and different aspects of language and cognitive domains.

The study has a multi-method approach. It consists of a meta-analysis investigating the bilingual advantage theory in executive functions (EF) and two studies based on data from the longitudinal study The Stavanger Project—The Learning Child (The Stavanger Project). Study 2 uses data from the first wave of The Stavanger Project. The study investigates Norwegian language comprehension in a monolingual control group and three different groups of bilingual children at 2 years and 9 months. The three bilingual groups had different amounts of exposure to Norwegian. The third article is based on data from the fourth wave of The Stavanger Project and investigates different aspects of Norwegian language and reading skills across bilingual learners and a monolingual control group of 5<sup>th</sup> graders. The sample in Study 3 is a subsample of the participants in Study 2; thus, the bilingual learners had been systematically exposed to Norwegian by early childhood education and care (ECEC) attendance and schools from at least the age of 2.

The thesis contributes three main findings. The first article provides little support for a bilingual advantage in overall EF. Moderator analysis targeting sample characteristics of bilingual subgroups that are theorized to have the largest bilingual advantage in EF shows no relation to the overall outcome of the analysis of differences in executive functions between bilingual and monolingual learners. Furthermore, there is limited evidence for a bilingual advantage in any EF domain. There is an advantage in switching, but not for all populations of bilingual learners.

The second article shows that bilingual toddlers have weaker second language comprehension skills than monolingual toddlers, but the differences in second language skills between different groups of bilingual learners are not fully explained by the time on task hypothesis. Bilingual children with mostly first language (L1) input at home had poorer Norwegian language comprehension than the two other bilingual groups. Bilingual toddlers with both first and second language input at home and bilingual toddlers with mostly second language input at home had equivalent second language skills. It therefore seems likely that a threshold value exists for the amount of second language input necessary to develop good second language skills rather than a direct relationship between the amount of input and language skills.

The third article shows that even after long and massive exposure to the second language, early bilingual 5<sup>th</sup> graders have lower vocabulary depth, listening comprehension and reading comprehension in their second language than their monolingual peers. The difference cannot be explained by differences in socioeconomic status (SES). Their decoding and text cohesion vocabulary skills are equal to those of monolingual learners. In contrast to some other studies, the strength of the predictive path between different aspects of language skills and reading comprehension was found to be equal across language groups.

In total, these findings contribute to the knowledge base of what is typical development of language, reading skills and executive functions for different groups of bilingual learners. Without information of what is typical development for different bilingual groups, it is difficult to identify atypical development. Hence, the knowledge this thesis provides can support educators in identifying bilingual learners with learning disabilities earlier and with greater certainty, thereby reducing the risk of both over- and under-identifying bilingual learners in need of special needs education.

## Table of Contents

Acknowledgements.....	iii
Summary.....	v
1 Introduction.....	1
1.1 Background and aim .....	1
1.2 The hypotheses of the thesis .....	4
1.3 Outline of the thesis .....	5
1.4 Outline of the extended abstract.....	6
1.5 The individual articles in this thesis:.....	6
2 Theoretical and empirical foundation .....	9
2.1 Bilingual learners and some of the many terms for this group.....	9
2.1.1 First and second languages.....	13
2.2 Advantages and disadvantages of bilingualism .....	14
2.2.1 The bilingual advantage in executive function .....	14
2.2.2 The disadvantage in (most) language skills .....	16
2.2.3 Theoretical criticism of the bilingual advantage theory in EF.....	18
2.3 Characteristics of bilingual children’s second language skills and development.....	19
2.3.1 Bilingual children’s second language skills and development .....	19
2.4 Predictors of second language skills .....	22
2.4.1 Length of exposure to L2 .....	22
2.4.2 Amount of exposure to L2.....	25
2.4.3 Age of second language acquisition (AoA).....	27
2.4.4 The quality of L2 input.....	29
2.4.5 Bilingual first children’s language acquisition .....	30
2.4.6 One native majority speaking parent .....	33
2.4.7 Socio- economic status (SES) .....	35
2.4.8 Research gaps.....	38
2.5 The connection between language and reading comprehension for the bilingual reader .....	39
2.5.1 Bilingual children’s mastery of specific linguistic aspects in L2 .....	40

2.5.2	Differences and similarities in predictive patterns from linguistic skills (and decoding) to reading comprehension in monolingual and bilingual readers .....	49
2.5.3	Predictive patterns of specific L2 skills on reading comprehension in L2 .....	53
2.5.4	Research gaps.....	63
2.6	Executive functions and the alleged bilingual advantage.....	64
2.6.1	Different cognitive domains in executive functions .....	65
2.6.2	Measures of different cognitive domains .....	71
2.6.3	The alleged bilingual advantage in EF .....	73
2.6.4	Research gaps.....	75
2.7	Research questions addressed in the different articles in this thesis .....	76
3	Methodological perspectives and considerations .....	79
3.1	Study 1 .....	80
3.1.1	Preregistration/data collection .....	80
3.1.2	Analysis.....	81
3.2	Study 2 .....	83
3.2.1	The Stavanger Project and participants for article 2 .....	83
3.2.2	Selection of measures.....	85
3.2.3	TRAS as a tool for research.....	85
3.3	Study 3 .....	89
3.3.1	Participants and content forms .....	89
3.3.2	Selection of measures .....	90
3.3.3	Data collection .....	93
3.3.4	Missing data .....	94
3.3.5	Analysis.....	95
3.3.6	Statistical methods – choice of model .....	96
3.4	Construct validity .....	98
3.5	Ethical considerations .....	100
3.5.1	Research on vulnerable groups.....	100
4	Results.....	101
4.1	Research question 1: Do bilingual learners have superior EF skills and poorer language skills?.....	102
4.2	Does the claim of superior EF levels and poorer language levels hold across different groups of bilingual learners? .....	103



4.3	Does the claim of superior EF levels and poorer language levels hold across different cognitive domains and linguistic aspects?.....	105
4.4	Is the predictive pattern of language and decoding skills to reading comprehension different for bilingual and monolingual learners?.....	106
5	Discussion .....	107
5.1	Bilingual learners have superior EF levels and poorer language levels than monolingual learners.....	107
5.2	The assumption of superior EF and poorer language levels holds across different groups of bilingual learners.....	112
5.3	Bilingual learners have superior EF levels and poorer language levels across different EF domains and linguistic aspects.....	119
5.4	The predictive patterns of language and decoding skills are different for (early) bilingual than for monolingual learners.....	128
5.5	Ethical reflections .....	131
6	Practical implications.....	135
7	Limitations .....	141
8	Recommendations for future research.....	147
	References.....	151
	Article I.....	185
	Article II.....	211
	Article III .....	235
	Appendices .....	285
	Appendix 1. Study characteristics of primary studies in reviews where the authors examine bilingual learners' language levels after 5-7 years of L2 exposure .....	285
	Appendix 2. The primary studies included in De Houwer's (2009a, 2009b, 2012) reviews.....	286
	Appendix 3. Summary of study characteristics of prediction studies of reading comprehension .....	290
	Appendix 4. Additional analysis of language skills between different subgroups of bilingual toddlers and their monolingual peers.....	291

Appendix 5. Information about Study 3 sent to parents accompanied by a consent form for their children’s participation in the study. ....	293
Information letter sent to the 2006 birth class .....	293
Information letter sent to the 2007 birth class. ....	295
Appendix 6. Parent questionnaire sent to participants in Study 3 .....	297

## Table of Figures

Figure 1. Visual illustration of which hypotheses are addressed in the different articles. ....	5
Figure 2. Graphic representation of the three measurement models as they could be applied at the item level.....	98

## List of Tables

Table 1. Study characteristics of the targeted bilingual subgroup, information regarding what their language levels are compared to, and the bilingual learners’ L2 outcomes in the primary studies included in the reviews examining bilingual learners’ proficiency levels after 5-7 years.....	285
Table 2. Overview of the study characteristics describing participant age, sample size, longitudinal status and type of comparison study for the sources referenced in De Houwer’s reviews (2009a, 2009b, 2012) .....	286
Table 3. Study characteristics of studies examining the predictive pattern of multiple linguistic skills for reading comprehension for bilingual children with a minimum of 4 years of exposure to their second language.....	290

## Appendices

Appendix 1. Study characteristics of primary studies in reviews where the authors examine bilingual learners' language levels after 5-7 years of L2 exposure .....	285
Appendix 2. The primary studies included in De Houwer's (2009a, 2009b, 2012) reviews .....	286
Appendix 3. Summary of study characteristics of prediction studies of reading comprehension .....	290
Appendix 4. Additional analysis of language skills between different subgroups of bilingual toddlers and their monolingual peers.....	291
Appendix 5. Information about Study 3 sent to parents accompanied by a consent form for their children's participation in the study.	293
Information letter sent to the 2006 birth class .....	293
Information letter sent to the 2007 birth class. ....	295
Appendix 6. Parent questionnaire sent to participants in Study 3 .....	297



# **1 Introduction**

## **1.1 Background and aim**

Large comparative studies investigating bilingual and monolingual children's academic achievement have found that bilingual learners score lower than their monolingual peers (Halle, Hair, Wandner, McNamara, & Chien, 2012; Han, 2012), even when they have the same socio-economic backgrounds (Kieffer, 2008). In fact, bilingual learners have over decades been over-represented as children in need of special education (Artiles & Ortiz, 2002; Jim Cummins, 1984; Mercer, 1973; Nordahl & Overland, 1998; Pihl, 2010) and have higher school drop-out rates than monolingual learners (Midtbøen, 2019; U.S. Department of Education, 2012).

A key for bilingual learners to succeed in school is being fluent enough in their second language (L2) to benefit from and understand the information directed at them in school, both verbally and through reading. A meta-analysis showed that bilingual children have both lower oral language skills and lower reading comprehension skills in the instructional language than their monolingual peers (Melby-Lervåg & Lervåg, 2014). In fact, bilingual children's L2 skills at kindergarten age predict their learning outcome in 8<sup>th</sup> grade (Halle et al., 2012; Han, 2012). This is perhaps not surprising considering the central role that language comprehension plays in reading comprehension (Hoover & Gough, 1990). In fact, new empirical studies find that language and decoding explain 96-99.7% of variations in reading comprehension (Hjetland et al., 2018; Lervåg, Hulme, & Melby-Lervåg, 2018). However, some researchers argue that the definition of "the simple view of reading", where reading comprehension is explained as a product of decoding and language comprehension skills (Hoover & Gough, 1990), should also include executive function (EF) skills (Liu et al., 2018; Reynolds, 2000).

EF can be understood as the ability to work strategically towards an aim, to inhibit irrelevant information, to flexibly switch between tasks and to adjust information in working memory (Diamond, 2013). A recent meta-analysis found a correlation of medium effect size between EF and reading comprehension, with positive correlations between specific EF domains and reading comprehension (Follmer, 2018). The specific role that EF plays in reading is, however, discussed. Some researchers argue that since language and decoding skills explain most of the variation in reading comprehension, EF could affect reading comprehension only indirectly through decoding or oral comprehension (Haft et al., 2019; Kieffer, Vukovic, & Berry, 2013). Other studies, however, find a direct impact of EF on reading comprehension beyond decoding and oral language skills (Guajardo & Cartwright, 2016; Liu et al., 2018). Furthermore, EF is known to influence and predict children's academic performance beyond reading comprehension (Best & Miller, 2010; Raghubar, Barnes, & Hecht, 2010). Good EF skills could therefore promote academic success and potentially act as a protective factor among children with low language comprehension skills.

The empirical findings from group comparison studies of monolingual and bilingual learners suggest that learning two or more languages affects bilingual learners' language and executive function skills but affects them differently. The findings suggest that bilingual learners develop superior skills in executive functions but poorer language skills in each of their languages than monolingual children (Bialystok, 2009, 2017, 2018; Bialystok & Feng, 2011; Friesen & Bialystok, 2012). In fact, lower levels than those of their monolingual peers are found for all aspects of language, with the possible exception of meta-linguistic abilities. The lower language level may also cause the predictive paths from linguistic skills to reading comprehension to be stronger for bilingual than for monolingual readers. Additionally, the relative strength of bilingual learners in metalinguistic awareness is theorized to be enhanced by bilingual children's early understanding of their

language as one particular system among many (Jim Cummins, 1979; Vygotsky, 1964).

When comparing group differences between monolingual and bilingual learners, it is important to understand the background that bilingual learners as a group are highly heterogeneous, perhaps more so than monolingual learners. For instance, even though the number of bilingual children who attend universities is increasing in Norway, the individual differences in academic achievement within the group of bilingual children are greater than the differences within the group of monolingual children (Barne- likestillings- og inkluderingsdepartementet, 2012). This heterogeneity could be related, for instance, to cultural differences, bilingual children's fluency level in the language in which they are taught, or their socio-economic status.

To investigate why some bilingual students succeed in school while others struggle, more studies of different populations of bilingual learners investigating different aspects that influence academic learning are of great importance. Nevertheless, it is important to bear in mind that even if bilingual children on average have poorer performance in school, and in reading comprehension in particular, a large number of bilingual children succeed in school. In Norway, for instance, second-generation immigrants are over-represented in higher education in general and in highly prestigious fields such as medicine in particular (Midtbøen, 2019) and are more prone to successful socio-economic class transitions than monolingual learners (Hermansen, 2016). However, for those who struggle, it is important to examine what explains individual differences among bilingual children and between monolingual and bilingual learners to find plausible causes so that we can develop efficient interventions.

## **1.2 The hypotheses of the thesis**

The overarching objective of this thesis is to investigate cognitive profiles related to language and EF in bilingual children compared to monolingual children and whether there are differences in predictive patterns for reading comprehension between the groups. Thus, the overall hypotheses examined are as follows:

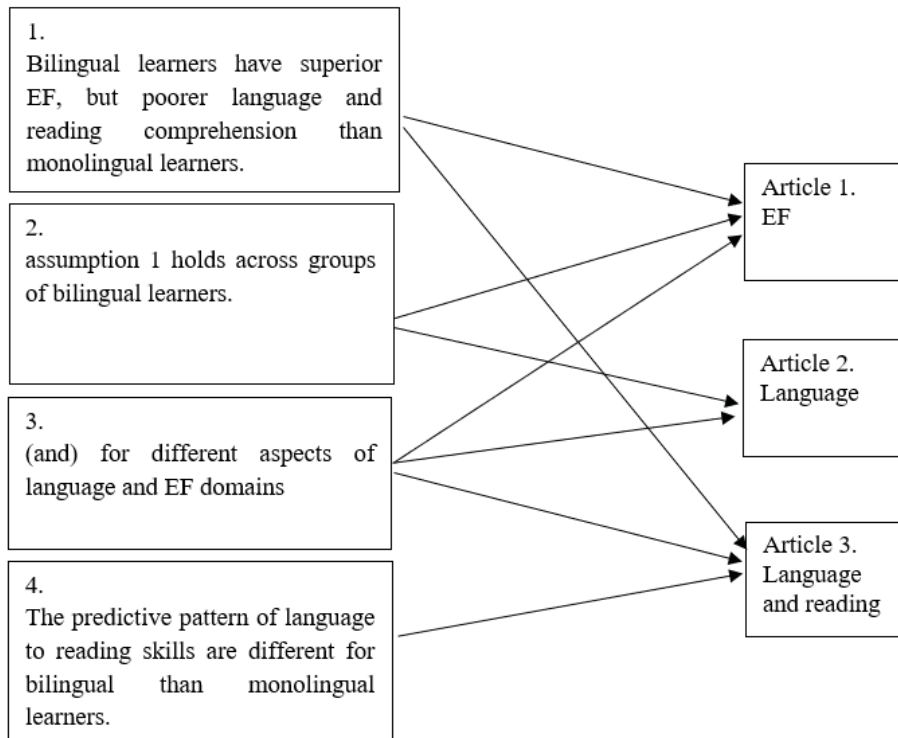
Bilingual children have superior executive functions and poorer language skills. This is in line with what has been suggested by Bialystok (Bialystok, 2009, 2017, 2018) and evidence from the large “Miami project” (Oller, 2005; Oller, Pearson, & Cobo-Lewis, 2007).

Poorer language and superior EF skills hold across different groups of bilingual children and across different cognitive EF domains and aspects of language. The lower level of language skills among bilingual children leads to different patterns between monolingual and bilingual learners in the prediction of reading comprehension. The different articles in this dissertation provide different angles for investigating these hypotheses that are summarized in Figure 1 below.

Visual illustration of which hypotheses are addressed in the different articles.



Figure 1. Visual illustration of which hypotheses are addressed in the different articles.



This thesis will examine whether these hypotheses are supported.

### 1.3 Outline of the thesis

The thesis consists of two main parts: a) the extended abstract (narrative) and b) three papers. The three papers were written in cooperation with different co-authors and represent individual studies. Study 1 is a meta-analysis investigating the theory of bilingual learners' superior abilities in executive functions (EF), both in overall EF and on the cognitive domain level. Differences in executive functions for different groups of bilingual learners are explored by moderator analysis. Studies 2 and 3 are part of the longitudinal study The Stavanger Project and investigate

different groups of bilingual children's second language (L2) skills compared to those of monolingual children. The bilingual groups in the articles differ in age and in exposure to Norwegian. Furthermore, the studies also examine possible differences in bilingual and monolingual learners' language skills in different aspects of the instructional language and differences in predictive patterns between language and reading comprehension.

#### **1.4 Outline of the extended abstract**

No single theory covers the development of all EF domains and different aspects of language skills and reading comprehension across different groups of children. Therefore, there is a need for a combination of different theoretical frameworks to examine the possible differences between bilingual and monolingual learners explored in this thesis.

Chapter 2 defines the central terms used in this thesis and outline the theoretical and empirical findings relevant to bilingual children's language, reading comprehension and EF skills. Chapter 3 is devoted to methodological perspectives and considerations related to the three studies. Chapter 4 provides a summary of the thesis results. In Chapter 5, the findings of the three studies are discussed in relation to the theory of lower linguistic levels but superior EF levels for bilingual learners. This discussion is embedded in the theory and empirical findings of studies of language, reading comprehension and EF. Chapter 6 discusses the practical implications of the thesis findings, and the thesis limitations are addressed in Chapter 7. Finally, Chapter 8 outlines recommendations for future research.

#### **1.5 The individual articles in this thesis:**

Article 1      Gunnerud, H., L., ten Braak, D., Reikerås, E., Donolato, E., & Melby-Lervåg, M. (accepted). Is Bilingualism Related to a Cognitive Advantage in Children? A

*Introduction*

---

Systematic Review and Meta-Analysis. *Psychological Bulletin*. Advance online publication. <https://psycnet.apa.org/fulltext/2020-67422-001.html>

Article 2 Gunnerud, H. L., Reikerås, E., & Dahle, A. E. (2018). The influence of home language on dual language toddlers' comprehension in Norwegian. *European Early Childhood Education Research Journal*, 26(6), 833-854.

Article 3 Gunnerud, H., L., Foldnes, N., Melby-Lervåg (in process). Level of skills and predictive patterns of reading comprehension in bilingual children with an early age of acquisition.

*Introduction*

---

## **2 Theoretical and empirical foundation**

### **2.1 *Bilingual learners and some of the many terms for this group.***

Bilingual learners are a very heterogeneous group, yet they have in common the mastery of two or more languages. The level at which they master their languages can, however, vary greatly. This variation has resulted in the development of several different definitions and terms for bilingualism. Some of the terms are intended to be strictly differentiated from other terms, while others have some overlap.

One way to differentiate between different groups of bilingual learners is the age of acquisition (AoA) of their second language. Examples of applicable terms are *simultaneous (2L1 learners/bilingual first children)* and *sequential bilingual learners, (early bilingual learners and child L2 learners)* (Chondrogianni, 2018; Genesee, 2010; Kovelman, Baker, & Petitto, 2008). According to this categorization, *simultaneous bilingual learners*, also often referred to as *bilingual first children*, have two first languages and are regularly exposed from birth to two or more languages (De Houwer, 2009a). In contrast, *sequential bilingual learners* are bilingual learners who speak one language at home and are introduced to L2 after they are one to two years old through attendance at early childhood education and care (ECEC) institutions (Genesee, 2010). Early *sequential bilingual learners* can again be divided into two subgroups, *early bilingual learners* or *child L2 learners*. *Early bilingual learners* are introduced to L2 at a timepoint from birth up to the age of 3 (Kovelman et al., 2008). (Note that this definition implies that some *early bilingual learners* could also be *bilingual first children*.) In contrast, a *child L2 learner* is introduced to L2 after 3-4 years of age and before the age of seven. Thus, these learners have a well-developed L1 (first language) before being introduced to L2.

Another way of differentiating between groups of bilingual learners is the learners' proficiency level in the two languages. Terms in use here are, for instance, *balanced bilingual learners* versus *second language learners/minority language learners*. *Balanced bilingualism* refers to bilingual learners who master both (all) of their languages equally well (Kohnert & Bates, 2002). In contrast, *minority language learners* or *second language learners* refers to children from homes in which a language other than that of the society is the primary spoken language (August & Shanahan, 2006). Note that even though research on *minority language learners* has often shown that these students lag behind their monolingual peers in the instructional language (Kieffer, 2008; Mancilla-Martinez & Lesaux, 2011a; Oller et al., 2007), these definitions simply state that another primary language is spoken at home but do not state how well the children have mastered their second language.

In recent years, the categorization of bilingual learners into different subgroups of learners has been criticized (e.g., *sequentially bilingual learners, child L2 learners*). One reason for criticizing the categorization of different types of bilingual learners is that such groupings are complex due mainly to the difficulties of creating valid and non-overlapping groups of bilingual children (Genesee, 2010; Luk & Bialystok, 2013). It is also questionable whether the term bilingual can be used as a categorical variable in terms of whether someone is or is not bilingual. The findings from a confirmatory factor analysis found that the term bilingual is a multi-dimensional construct (Luk & Bialystok, 2013). The construct consisted of how fluent bilingual learners are in their languages, the bilingual learners' usage of the languages on a daily basis, and the age of onset of active bilingualism. Only by exploring all these dimensions is it possible to fully capture the history, insensitivity and performance of bilingual experience. Most researchers do not explore these dimensions when they define the bilingual participants in their studies as bilingual learners or before they label them by a specific term for a subgroup of bilingual learners. Additionally, from a methodological

view, it is unfortunate to categorize variables that are likely to be continuous (MacCallum, Zhang, Preacher, & Rucker, 2002). Thus, it is recommended that bilingualism should be treated as a continuous variable where the degree of mastering different languages could vary from little to full mastery (Bialystok, 2017). Article 1 in this dissertation explains this understanding of the term bilingual learners in more detail. This thesis acknowledges the contribution made by this definition to the research field concerning bilingualism and recommends that current and forthcoming research report studies of bilingual learner samples by measuring bilingualism as a continuous construct. However, most of the prior research summarized in this thesis has treated bilingualism as a categorical variable, not a latent continuous variable consisting of different dimensions of bilingualism. The term bilingual learners is therefore used throughout the extended abstract and in the articles not as a continuous multiple construct but as an overall term for bilingual learners in general without including any combined measures of the degree of bilingualism.

Note, however, that the term bilingual is not used consistently throughout the three articles and the extended abstract. The reason is partly that prior research was conducted on specific populations of bilingual learners. When referencing these studies, the most accurate term for bilingual learners is used to facilitate readers' understanding of when and to which population of bilingual learners these results can be generalized. Furthermore, article 1 examines a different population of bilingual learners than articles 2 and 3. The populations in articles 2 and 3 partly overlap, and these articles investigate bilingual learners' language skills at different timepoints in a longitudinal study.

In article 1, the meta-analysis includes the full range of bilingual learners from children who have just started to master their second language to *balanced bilingual learners*. When searching for articles to include in the Meta-Analysis several terms for bilingual learners were used, yet all the articles that were coded in the meta-analysis referred to bilingual

learners by the term bilingual. The reason could be the name of the theory that most of the identified articles addressed, the bilingual advantage theory. To address this specific research question, the term bilingual learners was most accurate. This article is also the one that comes closest to treating bilingualism as a continuous construct. In this article, the impacts of different bilingual learners' language experiences are used as separate moderator variables as an alternative to treating bilingualism as a continuous multi-construct variable (De Cat, 2020; Kaushanskaya & Prior, 2015; Luk & Bialystok, 2013; Unsworth, 2013). Different bilingual language experiences can be regarded as different dimensions of bilingualism. The moderator variables were AoA, the degree of balanced bilingualism (difference in level of L1 and L2 proficiency), and L2 proficiency (difference in L2 level between bilingual and monolingual learners).

The other two articles are based on data from participants in The Stavanger Project. The participants in these studies differ from those investigated in the meta-analysis. In these studies, only data on the children's level in L2 were available, preventing the possibility of calculating the degree of balanced bilingualism; bilingualism is therefore labelled a categorical variable targeting the children's low AoA.

In Study 2, the only available information on the children's bilingual status was that their parents regarded them as bilingual learners and the language(s) spoken at home. We lacked information on the parents' nationality and could not determine whether parents speaking Norwegian were native Norwegian speakers. We did, however, know that all the bilingual children had been introduced to L2 by at least two years of age. In this article, *early bilingual learners* are referred to as *dual language learners*. The term *dual language learners* refers to children who have acquired two or more languages prior to the age of five (Genesee, 2010). In article 2, *dual language learners* are often referred to as dual language toddlers, which explicitly means that this



subpopulation of *dual language learners* was introduced to L2 at an earlier timepoint than 5 years of age.

More sample characteristic information of the bilingual children was available in Study 3, when the children turned 10, due to yet another round of data collection. After the data collection, it was clear that the majority of the bilingual children in Study 3 were bilingual learners from birth, with one minority language-speaking parent and one native Norwegian-speaking parent. The subselection of the sample without a native Norwegian-speaking parent all came from minority language-speaking households. These terms are used in the section of article 3 describing the sample characteristics. However, when the bilingual learners in this article are referred to as a combined group, they are labelled *early bilingual learners*, which is a more precise description of this subgroup than *dual language learners*. The switch in terminology from *dual language learners* to *early bilingual learners* is also grounded in an adaptation to the terminology used in the papers addressing the theme of article 3.

Finally, to explore whether Bialystok's claims of poorer language levels but superior EF levels hold across different groups of bilingual learners (Bialystok, 2009, 2017), multiple terms are used for bilingual learners in the discussion of the thesis results. To avoid confusion, however, the term *dual language learner* was replaced with *early bilingual learners* in the presentation and discussion of the results in Study 2. This enables one consistent term for the bilingual participants from The Stavanger Project to be used throughout this extended abstract.

### **2.1.1 *First and second languages***

All the articles in this thesis include a combination of bilingual children with one native Norwegian-speaking parent and one minority language-speaking parent and bilingual children with two minority language-speaking parents. This means that for some of the investigated bilingual

learners, the majority language of the society could be the child's first language (L1); for others, it could be their second language (L2). For *simultaneously bilingual learners*, both the majority language and the minority language are the child's first language. To address the bilingual learners' skills in the different languages at the group level, the children's skills in the minority language are always referred to as the children's L1 and the children's skills in the majority language as their L2. This categorization is carried through the whole thesis regardless of which language the bilingual children actually were first introduced to or were superior in.

## **2.2 Advantages and disadvantages of bilingualism**

There are numerous benefits of being bilingual. Studies show that being bilingual supports children in maintaining strong ties with their family, culture and ethnic identity (Tse, 2000; Zelasko & Antunez, 2000). Bilingualism can also lead to advantages in the job market related to both salary and position in the employment hierarchy (Rumbaut, 2014). Additionally, mastering multiple languages provides the opportunity to be immersed in different cultures and to gain unfiltered access to their history, art and literature (Gabszewicz, Ginsburgh, & Weber, 2011). It has, however, been suggested that bilingual learners have poorer language skills in both their languages, and there is ongoing debate regarding a possible bilingual advantage in executive function.

### **2.2.1 The bilingual advantage in executive function**

According to the theory of bilingual advantage in executive function (EF), cognitive processing in bilingual learners, particularly those who frequently shift between languages, differs from that in monolingual learners. The putative reason is that the additional requirement of bilingual learners to function in two or more languages creates experiences that monolingual learners do not have to undergo. These

experiences influence the development of bilingual learners' cognitive performance. Some researchers argue that the underlying mechanism is that both (all) of the different languages that bilingual learners master is active simultaneously in the same conversation. This activation allegedly happens regardless of whether the bilingual speaker communicates in only one of his/her languages. According to Green's (1998) inhibitory control model, the activation of all languages creates a need to select the linguistic criteria of the form and meaning of the corresponding language rather than those of the activated and competing other language. According to this theory, this competition requires a mechanism for controlling attention to one language by inhibiting the other. The competing language systems make bilingual learners' speech production differ from that of monolingual learners and cause cognitive and linguistic consequences of bilingualism. For an in-depth discussion of this theory, see the introduction to article 1.

The theory also suggests that competing language activation creates a need for attention control that is unique for bilingual learners and causes a greater need for attention control for bilingual than for monolingual learners (Bialystok, 2009, 2017). This attention control has been theorized to be responsible for both the linguistic and cognitive consequences of bilingualism. It has been suggested that handling conflicts related to which activated word to select in a given conversation, among other language-related outcomes, leads to word-finding difficulties. On the more optimistic side, this mechanism supposedly works as a brain exercise and thereby positively affects EF skills.

Bialystok explains this attention control as a complex ability involving underlying skills such as attention, switching, working memory and inhibition (Bialystok, 2017). Furthermore, she suggests that this attention control draws upon general EF components. Thus, the same attention control ability is activated in the regulation of both linguistic and non-linguistic stimuli.

Nevertheless, in this theory, not all bilingual learners have the same advantage in EF. It has been suggested that early age of acquisition of the second language, equal degrees of fluency in both languages, and frequent changes between the two languages are factors that positively affect bilingual learners' level of attention control. For a more detailed explanation, see article 1.

This theory has gained widespread support on different levels. For neuroscience, there is empirical support from neuro-imaging showing greater structural density in bilingual learners' brains, which resemble those obtained from older children or adults (e.g., for review, see Bialystok, 2017). This could be seen as evidence of better brain development in bilingual learners (Bialystok, 2017). In addition, behavioural data support the claim of bilingual advantage in EF from early childhood (Bialystok, Barac, Blaye, & Poulin-Dubois, 2010; Woods, 2013; Yoshida, Tran, Benitez, & Kuwabara, 2011) throughout adulthood (Lee Salvatierra & Rosselli, 2011). The advantages hold across various language pairs and are distinct from the effects of culture, immigration history, and language of instruction (Barac, Bialystok, Castro, & Sanchez, 2014). There is even support for bilingualism postponing the onset of dementia (Bialystok, Craik, Binns, Osher, & Freedman, 2014; Craik, Bialystok, & Freedman, 2010).

### *2.2.2 The disadvantage in (most) language skills*

In addition to creating an advantage in EF, bilingual learners' need for attention control of the selected language of conversation also has linguistic consequences (Bialystok, 2009). Here, Bialystok refers to empirical findings of weaknesses in lexical retrieval functions such as picture naming, verbal fluency tasks and tip of the tongue experience, all aspects of bilingual learners' speech production. Bialystok draws support from Green's speech production model (Green, 1998) and claims that at least part of the problems with speech production are

caused by interference from the non-selected language (Bialystok, 2009). The cause of the disadvantage of receptive vocabulary is, however, explained differently. Bilingual learners are building two lexical systems to integrate into a semantic network. They have less exposure to each of their languages than monolingual learners do; thus, their learning opportunities are divided between their languages. It is therefore not surprising that they know fewer words in each language (Bialystok & Feng, 2011; Friesen & Bialystok, 2012).

However, not all linguistic-dominated abilities are considered bilingual disadvantages. In a critical review of the cognitive development of *dual language learners*, Bialystok and colleagues argue that metalinguistic awareness is a strength in bilingual learners and define metalinguistic awareness as morphological, syntactic and phonological awareness (Barac et al., 2014). The reason Bialystok and colleagues suggest that this linguistic ability is a strength and not a weakness is that metalinguistic awareness draws on both linguistic and cognitive abilities (Barac et al., 2014; Bialystok, 2001a, 2001b; Bialystok, Peets, & Moreno, 2014), with different metalinguistic tasks requiring different levels of executive control and linguistic knowledge (Bialystok, 2018). More specifically, Bialystok suggests that bilingual learners are superior to monolingual learners in metalinguistic tasks that additionally require controlled and selective attention to avoid interference from conflict. Such tasks enable bilingual learners to use their EF skills, and these superior skills offset the costs of weaker L2 knowledge (Friesen & Bialystok, 2012). Friesen & Bialystok (2012) acknowledge that linguistic abilities are the core of language proficiency, and such measures generally favour monolingual learners. Even though metalinguistic abilities in bilingual learners are enabled through enhanced control, they suggest that limited formal linguistic knowledge in L2 would work as a restriction of the ability (Bialystok, Peets, et al., 2014), especially when tasks depend heavily on linguistic knowledge of L2 (Friesen & Bialystok, 2012).

The theory of a bilingual advantage in EF is the rationale for bilingual learners' superior abilities in some metalinguistic awareness tasks that are also linked to bilingual learners' experience with two simultaneously active languages. The attention control created by handling two active languages directly supports bilingual learners in those metalinguistic awareness tasks that require a high level of (attention) control (Bialystok, 2018).

### *2.2.3 Theoretical criticism of the bilingual advantage theory in EF*

The theory that bilingual learners' constant need for attention control creates non-linguistic superiority has been criticized. Paap and Greenberg (2013) outlined three underlying assumptions on which the theory regarding superior EF abilities for bilingual learners is based. The three assumptions function as a step theory, which means that they build on one another, and all of them need to be true for Bialystok's theory to be supported.

First, all conversations, in monolingual as well as bilingual learners, create the need to monitor signals of turn switching, topic changes, sarcasm, misunderstanding, etc. Furthermore, multiple semantic and syntactic candidates in the semantic associative network are activated simultaneously in every conversation, creating a need to select one word and suppress others. There is also a need to suppress the irrelevant meaning of homographs. Thus, the theory of a bilingual advantage is based on the assumption that for bilingual learners, there is an additional and unique requirement to recruit sufficiently more executive control during language processing and that this creates group differences in EF between bilingual and monolingual learners. Second, the theory is also based on the assumption that bilingual learners' attention control during language production and conversations is controlled by an independent executive function domain that processes both linguistic and non-linguistic stimuli. Alternatively, handling inhibition, monitoring, and

switching between languages, etc. could be specialized within the language module (Fodor, 1983; Frazier, 1987); thus, these functions are specific to tasks that depend on linguistic representation. The third assumption is that a threshold value for the extent of EF activities that learners must engage in to develop good EF skills does not exist. If it does, then everyday-life activities such as inhibition of distractors and inappropriate responses, planning activities, pursuing goals, switching activities, etc. could perhaps be enough to optimize the capacity for control.

There is also reason to question the theoretical rationale for the alleged linguistic profiles of bilingual learners. Since the metalinguistic awareness advantage in tasks, which relies heavily on (attention) control, is theoretically directly caused by superior EF skills, Paap and Greenberg's (2013) criticism of the three underlying assumptions of EF advantage theory also applies here.

## **2.3 *Characteristics of bilingual children's second language skills and development***

### **2.3.1 *Bilingual children's second language skills and development***

Children's language development is complex. This is, however, easy to forget, as the manifestation of linguistic development in most children appears to be uncomplicated and effortless. In reality, language development is a complicated and lifelong process involving several aspects of language, such as morphology, lexicon, syntax and phonological development (Nippold, 2016).

Regarding the development of a lexicon, children gradually learn new words by beginning with a restricted, contextual meaning of a word to partly understand it before they fully master the differentiated meaning

of the word in different contexts (Nagy & Herman, 1987). To complicate this even further, lexicon development increases rapidly, from a mean vocabulary of 5,200 root words in 2<sup>nd</sup> grade to approximately 8,400 root words by 5<sup>th</sup> grade to 26,000 words at undergraduate age (Biemiller & Slonim, 2001). Bilingual children need to learn these words not in one language but in two or more and therefore have a much larger learning task than monolingual learners. Learning two languages seems to be challenging.

The results of a meta-analysis show that bilingual children on average have lower second language skills than their monolingual peers (Melby-Lervåg & Lervåg, 2014). In fact, in only 4 of 124 studies did the bilingual samples show language skills better than or equal to those of their monolingual peers. The mean effect size was -1,12 in favour of the monolingual learners, whereas samples with low socio-economic background or those in which only the first language was used at home showed the largest group differences. Furthermore, bilingual children, on average, have weaker vocabulary skills for both receptive and expressive vocabulary than monolingual learners (Bialystok & Feng, 2011; Bialystok, Luk, Peets, & Yang, 2010). The difference is found for both vocabulary breadth and depth and even for high-frequency words (Schwartz & Katzir, 2012; Verhallen & Schoonen, 1993). Bilingual learners can also show a delay in morphology and lexical retrieval tasks (Bialystok & Feng, 2011; Droop & Verhoeven, 2003). Hence, most studies of bilingual and monolingual learners find differences in language levels, while a few studies conclude that comparable levels exist (e.g., Jim Cummins, 1984, 2017; Hakuta, Butler, & Witt, 2000; Paradis & Ruiting, 2017).

Hoff et al. (2012) point out that studies that conclude with comparable language levels between bilingual and monolingual children are often misleading. According to Hoff, the reason is that instead of comparing the bilingual learners to a monolingual control group, they compare bilingual learners to test norms of monolingual learners. The researchers



then often conclude that language skills are equal if the bilingual children's performance is within +1 to -1 SD of the test norms in use. However, in reality, the scores of bilingual children are in the lower range of this scale compared to the average language levels of monolingual children.

In addition to vocabulary skills, bilingual children need to acquire the grammar/sentence structure and phonology of both languages. Since every language has its own grammatical rules for how words are inflected, as well as how sentences are constructed, grammatical and syntactical differences can provide an additional challenge when learning a new language (Bjerkkan, Monsrud, & Thurmann-Moe, 2013). However, whereas the total number of words bilingual learners need to learn in L2 is extensive (Biemiller & Slonim, 2001; Nagy & Anderson, 1984), one should assume that the average L2 learner needs less time to master a limited set of grammatical rules than vocabulary skills. This assumption is supported by Gisela, Eva-Kristina, and Ulrika's (2003) study of Arabic preschool *L2 learners* learning Swedish. In their study, the *L2 learners* mastered Swedish grammar (e.g., adding morphemes to a stem, phrasal morphology and inter-phrasal morphology) within a mean length of 1;6 years. *Sequential bilingual children* seem to master the phonology after relatively brief systematic L2 exposure (Chondrogianni, 2018). This does not imply that they have a fully developed grammar by two years of age, yet they master it well enough to take part in and understand everyday conversations (Jim Cummins, 2017).

Fluent pronunciation and mastery of basic grammar are both an advantage and a disadvantage for bilingual children. Correct pronunciation and grammar definitely make it easier to communicate successfully with others, which again provides access to more meaningful situations in which bilingual child can improve their second language skills. However, based on interaction in everyday conversations, it is surprisingly difficult to detect whether a child has

language problems (Im-Bolter & Cohen, 2007) and determine if the child understands what is said based on semantics alone or only in combination with situational knowledge and the interlocutor's non-verbal communication signals (Espenakk et al., 2011). It is much easier to detect language challenges based on pronunciation errors or grammatical errors. In this way, native-like grammar and pronunciation can create the image of a bilingual child who fully understands his or her second language (Espenakk et al., 2011). This often false image can lead to late rather than early identification of a possible need for more targeted language intervention (Espenakk et al., 2011). However, as discussed earlier, there is a large difference in how well bilingual children master their first and second languages.

## **2.4 Predictors of second language skills**

There is an ongoing discussion of to what extent bilingual children develop language levels equal to those of their monolingual peers. Some researchers claim that bilingual children do develop comparable levels (Collier, 1989; Jim Cummins, 2017), and others suggest that bilingual learners close the gap between their language skills and those of their monolingual peers (Han, 2012; Kieffer, 2008) yet perhaps never truly become equally proficient (Abrahamsson & Hyltenstam, 2009). The section below presents some of these studies, displaying important predictors of good second language skills and evidence from studies comparing different subgroups of bilingual to monolingual learners. The latter is important since most researchers agree that there are large group differences among bilingual learners and predict different developmental trajectories for different subgroups of bilingual learners.

### **2.4.1 Length of exposure to L2**

Perhaps the most intuitive factor that influences *minority language learners'* L2 development is the length of exposure to L2. In line with the time on task hypothesis, one can expect that the longer the length of

L2 exposure is, the better the learning outcome in L2 (Carroll, 1963). However, how long does it take for *minority language learners* to develop L2 proficiency? L2 proficiency is often referred to as cognitive academic language proficiency (CALP) (Jim Cummins, 1984, 2017) and is operationalized as the academic language required to succeed in school. Several narrative reviews of the field suggest that it takes 5-7 years to develop CALP (Collier, 1989; Jim Cummins, 1984, 2017; Hakuta et al., 2000; Saunders & O'Brien, 2006). Some reviews take the impact of 5-7 years of L2 exposure further and claim that this timespan is the number of years it takes for bilingual learners to develop levels of language equal to those of their monolingual peers (Collier, 1989). This synthesis of research mostly involves studies on children introduced to L2 early in life (often just before or after entry into school). (see study details of these summaries presented in Table 1, appendix 1). Narrative reviews are, however, often vulnerable to bias since the criteria for inclusion are often unclear, and the researcher can cherry-pick studies to report (Borenstein, Hedges, Higgins, & Rothstein, 2011).

Studies published after these narrative reviews were conducted question the likelihood of developing equal language levels in the instructional language within 5-7 years. Several large-scale studies, most of them comparing several hundred *minority language learners'* L2 levels with those of monolingual control groups, show that when sufficiently sensitive tests are used, the results indicate that bilingual children do not catch up with monolingual children within this timeframe (Bialystok, Luk, et al., 2010; Droop & Verhoeven, 2003; Farnia & Geva, 2013; Oller et al., 2007). Moreover, when the effect sizes for the three bilingual groups in the ECLS-K study by Halle et al. (2012) of approximately 17,190 monolingual and 2,700 bilingual children are combined, the bilingual children's skills in the instructional language still lagged behind those of their monolingual peers, even after 9 years of L2 exposure.

The primary studies in the much-cited reviews by Jim Cummins (1984), Hakuta et al. (2000) and Saunders and O'Brien (2006) most often compare *sequential bilingual learners'* L2 skills to norms of monolingual children (see Table 1 in appendix 1). The same is true of some later published studies, such as Paradis and Ruiting (2017). This research approach is questionable for several reasons. First, even though most bilingual children were approaching the normed mean, any particular cohort could differ from the normed mean due to an unexpected difference between that particular cohort and the norming sample. The same is true for the mean of the selected monolingual sample. For instance, the bilingual children in the study of Paradis and Ruiting (2017) had a non-verbal IQ of 120, a factor known to influence language skills (Teepen, 2004). It is therefore possible that uneven distribution of non-verbal IQ across the bilingual sample and the norming sample was the reason that the bilingual children had levels of language skills equal to those of the monolingual sample. However, it is impossible to control for differences across possible third factors between the compared samples, such as IQ, when using normed means as comparisons.

Furthermore, at first glance, it might seem as if Cummins (1984;2017), Hakuta et al. (2000), Collier (1987) and Saunders and O'Brien (2006) performed the same investigations, and reached identical conclusions, yet this is not the case. In fact, some of the studies base their conclusions on the bilingual sample reaching within or close to -1 SD of the norms of monolingual children; others refer to the 32<sup>nd</sup>-50<sup>th</sup> percentile of norms, yet only a minority of the studies compare bilingual children's L2 skills to an actual monolingual control group. In addition, many of the primary studies cited in these reviews examine how many years of L2 exposure it takes for a bilingual learner to be reclassified as proficient in L2 (see Table 1 in appendix 1). Once proficient in L2, bilingual learners are not considered in need of L2 support. A variety of tests is used to identify these students; however, most of the tests do not rely on comparisons to

the norms of native language speakers (Hakuta et al., 2000; Thompson, 2017). The studies that do compare bilingual learners' L2 levels to the norms of native language speakers find that bilingual learners score below the mean of monolingual learners yet within or close to -1 SD (Jim Cummins, 2017; Hakuta et al., 2000). This is clearly not the same as having equal levels, although it could indeed mean that bilingual learners' skills are good enough for them to benefit from ordinary classroom lessons without supplementary support. It is also important to note that there are large individual differences. Notably, even though Cummins (2017) claims that some *sequential bilingual learners* reach the level of their monolingual peers within 5-7 years, he also acknowledges that some bilingual samples will never obtain language levels equal to those of their monolingual peers regardless of length of exposure. Cummins (2017) then questions whether sample characteristic factors other than length of exposure might cause these differences.

#### **2.4.2 Amount of exposure to L2**

Hammer et al. (2014) finds in a critical review of the knowledge base of bilingual learners' language and literacy skills that the amount of exposure to and young bilingual learners' usage of two languages appears to play a key role in their language development. In early childhood, a bilingual child's L2 skills are a reflection of the amount of L2 exposure the child has received at home. This is evident for the child's vocabulary size and grammatical and narrative abilities (Bridges & Hoff, 2014; Hipfner-Boucher et al., 2015; Hoff, Rumiche, Burrridge, Ribot, & Welsh, 2014; Marchman, Martínez, Hurtado, Grüter, & Fernald, 2017; Place & Hoff, 2011; Scheele, Leseman, & Mayo, 2010). The more L2 exposure children have at home, the narrower the gap between bilingual and monolingual peers in language skills in the instructional language (Vagh, Pan, & Mancilla-Martinez, 2009). Although the amount of parental exposure explains variations in young bilingual children's L2 skills, some of the few longitudinal studies of this topic cast doubts on

the long-term effect of parental L2 exposure (Hammer, Davison, Lawrence, & Miccio, 2009; Mancilla-Martinez & Lesaux, 2011b; Oller et al., 2007). One possible reason is that once a child attends kindergarten or school, the amount of exposure to L2 received in kindergarten/school also affects the child's L2 skills (Bohman, Bedore, Peña, Mendez-Perez, & Gillam, 2010; Gathercole & Thomas, 2009; Paradis & Kirova, 2014). It might be that over time, the amount of L2 exposure provided at school/kindergarten is enough for the bilingual learner to fully master L2 regardless of which language(s) the family speaks at home. This does not imply that bilingual learners will eventually develop language levels equal to those of their monolingual peers in the instructional language (Mancilla-Martinez & Lesaux, 2011b; Mancilla-Martinez & Lesaux, 2011a; Oller et al., 2007). Although some studies of *minority language learners* find that differences in language levels caused by different distributions of parental L1 and L2 at home even out over time, not all researchers agree. Hoff, Giguere, Quinn, and Lauro (2018) investigate growth in Spanish/English *early bilingual learners'* language skills from the age of two years 6 months to five years. They find that the relative amount of exposure significantly predicts L1 and L2 vocabulary scores. The growth in L2 scores was quadratic, showing an increase in the gap in English skills between children with little and much English exposure at home while controlling for the effect of education. Furthermore, none of the bilingual groups caught up with their monolingual peers.

Amount of exposure also affects older children's language skills beyond the L2 exposure provided in school. Living in neighbourhoods with high co-ethnic concentrations leads to higher exposure to L1 and less developed L2 skills (Rydland, Aukrust, & Fulland, 2013). Perhaps more interesting is the effect of the cumulative amount of exposure to L2 in a child's lifetime (from parents, school and other interlocutors) on children's language skills. This measure could be regarded as a purer measure of time invested in acquiring a language (Unsworth, 2013).

De Cat (2020) examined the language skills of 5- to 7-year-old bilingual learners to identify the threshold value for exposure necessary to develop L2 skills at a level of -1.25 compared to their monolingual peers. The threshold value, controlled for SES, varied as a function of different language constructs. The threshold value was 32-33 months for lexical semantics and sentence repetition and 44 months for discourse semantics. Note, however, that it is difficult to determine how this transfers to other age groups. The older the monolingual learners are, the better language skills they have. As a consequence, the gap that a bilingual child must then close is larger. From that perspective, one can expect the effect of the amount of exposure to be intermingled with the age of acquisition. The time a child needs to master L2 fluently is essential. If the child does not master L2 fluently by the end of 1<sup>st</sup> grade, he or she will most likely still lag behind his or her monolingual peers by the end of 5<sup>th</sup> grade (Halle et al., 2012). The age at which a child starts kindergarten/ECEC education therefore matters.

### *2.4.3 Age of second language acquisition (AoA)*

AoA could also affect L2 development. AoA is often understood as the age at which a child is immersed in the L2 context (Birdsong, 2006). Studies of the impact of AoA have often targeted bilingual adults, for whom late AoA and the level of later second language skills are negatively correlated (Birdsong, 2005; DeKeyser & Larson-Hall, 2005). A critical onset of AoA has been suggested as the reason for why adults struggle to obtain language levels similar to those of their majority native language-speaking peers (DeKeyser, 2013; DeKeyser & Larson-Hall, 2005). However, the cut-off age for this alleged critical period varies: some researchers suggest that the cut-off age is adolescence, middle school age, preschool age and even infancy (Nicoladis, 2018). Nevertheless, even if the critical period is not agreed upon, it seems clear that it could be more difficult for bilingual learners to acquire native-like L2 skills by adulthood than earlier studies assumed. One study shows

that after undergoing sensitive language testing without ceiling effects, only 3 of 193 adults who self-reported having language levels in the instructional language equal to those of monolingual learners actually had scores equal to those of their native language-speaking monolingual peers (Abrahamsson & Hyltenstam, 2009). The participants' AoA ranged from 1 to 47 years of age, and the three who had language levels equal to those of monolingual learners all had low AoA (Abrahamsson & Hyltenstam, 2009).

Studies of the impact of AoA on children's L2 skills are limited, especially for AoA between 0 and 5 years of age (Nicoladis, 2018). Theoretically, a distinction is often made between those who learn L2 before and those who learn L2 after 3-4 years of age. Bilingual children who are introduced to L2 after 3-4 years of age have already developed a broad foundation of L1 skills with most properties (e.g., word order, verbal inflation) already in place. Children learning L2 between 3-4 years and 7-8 years of age are often referred to as *child L2 learners* (Chondrogianni, 2018). Given their older AoA, by the time they are introduced to L2, they have developed more cognitive, linguistic, social, and literacy-related resources than bilingual learners introduced to L2 before the age of 3. These skills may provide the *child L2 learner* with an advantage in developing L2 skills. There is some evidence that *child L2 learners* have faster vocabulary growth than their peers with lower AoA (Blom & Bosma, 2016), yet this does not necessarily imply that they will eventually develop better L2 levels than *early bilingual learners*. Several studies find that bilingual learners with AoA before the age of 3-4 years out-perform *child L2 learners* (Kovelman et al., 2008; Meisel, 2018; Sebastián-Gallés, Echeverría, & Bosch, 2005; Unsworth et al., 2014). Notably, Unsworth et al. (2014) identified a complex interplay between input quantity and AoA, suggesting that both early AoA and amount of exposure positively influence L2 development. In a critical review of young bilingual learners' language and literacy skills, Hammer et al. (2014) find preliminary support for the hypothesis that the



age of second language acquisition affects the development of L2. They do, however, specify that the evidence for this stems mostly from studies of children's vocabulary development.

#### **2.4.4 *The quality of L2 input***

The input quality of L2 used at home and in school also seems to affect L2 development. In fact, minority language-speaking mothers' self-evaluated L2 proficiency scores predict which bilingual 7- to 8-year-olds would approach their monolingual peers after 4 years of exposure and which would not (Chondrogianni & Marinis, 2011).

Additionally, the quality of the language input that monolingual learners received from teachers is found to predict growth in language skills (Huttenlocher, Vasilyeva, Cymerman, & Levine, 2002). However, longitudinal studies investigating this relationship among bilingual learners are limited (Unsworth, 2016). While one study finds that teacher-led talk in ECEC institutions predicts differences at age 5 that remain stable throughout 5<sup>th</sup> grade (Rydland, Grøver, & Lawrence, 2014), the results from the study of Bowers and Vasilyeva (2011) differ to some extent. They find that the growth of vocabulary in bilingual learners is positively related to the total number of words produced by the teacher and negatively related to the number of words per utterance. The researchers suggest that the reason is that the bilingual children were at a stage in their L2 development where their L2 skills were relatively low. Furthermore, the teachers' use of complex language was far too advanced hence unlikely to facilitate L2 vocabulary growth. However, other studies show that attending ECEC institutions without a curriculum does not in itself promote young bilingual learners' L2 skills, even though L2 is the primary language spoken there (Hoff et al., 2018). Thus, taken together, there is support for high-quality L2 input, customized to the *second language learners'* L2 level, positively influencing their L2 development. This assumption is strengthened by evidence from experimental studies that show that interventions at an early point in life

can limit the gap between bilingual and monolingual learners' language levels (Heller, Lervåg, & Grøver, 2019; Rogde, Melby-Lervåg, & Lervåg, 2016). The quality of input therefore seems to be an important factor influencing L2 development as long as the intervention is fitted to the bilingual children's L2 level and consists of a fair number of lessons over time. Nevertheless, it should be noted that the effects of these interventions tend to fade over time and that they must be refreshed regularly.

#### **2.4.5 Bilingual first children's language acquisition**

*Bilingual first children*, also often referred to as *2L1 children* or *simultaneous bilingual learners*, are a very special group of bilingual learners. They are introduced to two or more languages from birth and have the advantage of a very low AoA and long exposure to L2 (the society's language), two factors that positively impact their language skills. The comparison of this group of bilingual learners' long-term L2 development to that of monolingual learners is of special interest since such group comparisons provide a good design for disentangling possible "deficits" in bilingual learners' language learning from other mechanisms influencing language development, such as AoA (De Houwer, 2009a). However, not even for this group of children do researchers agree upon whether they have language levels comparable to those of their monolingual peers.

De Houwer claims that *2L1 children* develop language skills equal to those of monolingual learners and that they reach language milestones at the same pace as their monolingual peers (De Houwer, 2009a, 2009b, 2012). She goes so far as to say that if they do not reach these milestones at the same ages as their monolingual peers, there is a need for language evaluation to ensure that they do not have a language disorder (De Houwer, 2018). The studies she builds her case on, claiming *bilingual first children* exhibit language levels equal to those of monolingual learners, could, however, be more solid. The book "An introduction to

bilingual development” (De Houwer, 2009b) is criticized for not including any longitudinal studies (Ringblom, 2010). A search through the rest of her references in the articles/books cited above shows that many of the studies with which De Houwer supports her claim are of *bilingual first learners* only, combined with comparisons of literature describing the ages of language milestones for monolingual learners. The characteristics of the studies referenced that do compare a *2L1* sample to an actual monolingual sample are presented in Table 2 in appendix 2. The majority of these studies focus on toddler age. With the exception of case studies, longitudinal studies are scarce (3 studies of toddlers), as are studies with participants aged 5 years or older (4 studies where the oldest sample is 8-year-olds). The studies of children 5 years or older either compare the *2L1 children’s* L2 levels to the norms of monolingual learners or have small sample sizes ( $N = 30$ ,  $N = 36$ ). Small sample size and/or comparisons with norms are general problems with most of the cited studies. The larger studies that she cites, such as Umbel, Pearson, Fernández, and Oller (1992) and Barrena, Ezeizabarrena, and Garcia (2008), compare the *2L1 children’s* language skills to the norms of monolingual learners. The study of Barrena et al. (2008) is, however, interesting. They find that *bilingual first learners* 16-30 months old exposed to L2 more than 60% of the time obtain language levels comparable to monolingual norms. Less exposure to L2 was associated with lower language levels.

Although De Houwer argues for equal language levels between *2L1 children* and monolingual learners, other researchers disagree (Bialystok & Feng, 2011; Grant, Gottardo, & Geva, 2011; Hoff et al., 2018). Hoff et al. (2018) compare the language levels of what seems to be a combined cohort of 139 *early bilingual learners/bilingual first learners* with 39 monolingual children. Hoff et al. (2018) find that bilingual children at 5 years of age were lagging behind their monolingual peers by 6 months to a year in English, depending on the amount of parental exposure to English. Note, however, that even

though the researchers report that the bilingual learners were exposed to both languages by infancy (30 months), they also state that some of the children lived in homes where Spanish was spoken exclusively. It is therefore difficult to determine what percentage of the bilingual learners were truly *bilingual first learners*. The lack of comparable language levels in this study could therefore be caused by the inclusion of bilingual children with an onset of L2 later than birth. Bialystok and Feng (2011) study a group that contained purely *bilingual first learners*. Bialystok and Feng (2011), however, do not divide *bilingual first children* into groups based on their amount of L1/L2 exposure. The bilingual children are all treated as one sample. The researchers investigate the effect size difference in language levels across all children in the age span of 5-9 years of age. Their large cross-sectional study consists of 963 participants in the age group of 5- to 9-year-olds, of whom half were monolingual learners and the rest were bilingual learners with an AoA from birth. The effect size difference in language skills in the majority language was large across all age groups and favoured the monolingual learners. Furthermore, a reanalysis of data from Pearson, Fernandez, and Oller (1993) performed by Bialystok and Feng (2011) shows the same pattern. (This study is one of the references cited by De Houwer in her 2009 books claiming equal levels in the instructional language.) Bialystok and Feng (2011) therefore claim that the patterns of lower vocabulary skills hold true for the productive vocabulary of children in the first two years of life as well as receptive vocabulary in preschool and early school years (Bialystok & Feng, 2011). Since large-scale studies have the advantage of increased statistical power (Ingre, 2013), the large-scale study of Bialystok and Feng (2011) is likely a more solid proof than De Houwer (De Houwer, 2009a, 2009b, 2012, 2018). It is, however, worth mentioning that none of these studies base their conclusions on large scale longitudinal data.

#### **2.4.6 One native majority speaking parent**

A native majority language-speaking parent is exposed to the majority language only throughout his or her childhood, with the exception of possibly attending foreign learning classes, is. This group of parents' language skills in the majority language differs from the language skills of bilingual learners who have been introduced to their L2 after 7 years of age, even in cases where the bilingual learners themselves self-report full proficiency in L2. (Abrahamsson & Hyltenstam, 2009). The L2 proficiency level of parents affects children's L2 development. Notably, having one native majority language-speaking parent could potentially enable different language trajectories than those for *2L1 children* with two non-native majority language-speaking parents. However, identifying studies of *2L1 children* with one native majority language-speaking parent is difficult.

De Houwer (2009b) claims that it is most common for *2L1 children* to have one majority language-speaking monolingual parent and one bilingual parent or for two bilingual parents to foster a *2L1 child* in a society with two languages considered of equal usage/importance in society (i.e., Canada, Wales, Ireland or the Basque Country—a region in Spain). This might be true. However, the presence of studies that show that having an older sibling increases native minority language-speaking parents' usage of L2 interaction with their younger children (Bridges & Hoff, 2014) makes this assumption an over-generalization. Being exposed to two languages at home from birth therefore does not necessarily imply that the L2 input is from at least one native majority language-speaking parent. There is therefore a need for clarification in all studies of *2L1 children's* language development. This clarification is lacking in the studies of Bialystok and Feng (2011), Pearson et al. (1993), De Houwer, Bornstein, and Putnick (2014) and most of the comparative studies of *2L1* and monolingual children cited in De Houwer (2009a, 2009b). One reason could be the shifting patterns of L1 and L2 language usage and abilities of parents providing bilingual children with L1 input

(Biedroń & Birdsong, 2019; Keijzer & De Bot, 2019; Kopke, 2019). In addition, large variations in language levels among native majority language speakers (Treffers- Daller, 2019) make it difficult to categorize some minority language-speaking parents' L2 input as non-native-like.

Studies that do provide information on parents' status as native majority language-speaking or not often consist of a combination of children with one native majority language-speaking parent and children with parents who themselves have grown up learning L1 (e.g., Place & Hoff et al. 2011, 2016, Hoff et al 2018) yet now use a combination of L1 and L2 when communicating with their children at home. What we know so far is that, as for other bilingual learners, *2L1 children's* language skills are affected by their amount of exposure to the different languages (Hoff et al., 2018; Sebastián-Gallés et al., 2005). Furthermore, they most often develop a dominant language. When dual language education is provided, the dominant language in early childhood could prevail as dominant, even for university *2L1* students highly proficient in both languages (Sebastián-Gallés et al., 2005). However, for *sequential bilingual learners*, the amount of native majority input explains variations in their majority language even after controlling for the amount of exposure to the majority language (Hoff et al., 2018; Place & Hoff, 2011, 2016).

Due to limitations in the studies, it is more uncertain whether *2L1 children* with one native majority language-speaking parent develop language levels in the majority language comparable to those of monolingual learners. Four-year-old *2L1* bilingual learners with one native language-speaking parent are found to have vocabulary levels in the majority language similar to those of monolingual learners (Hoff et al., 2014). However, some of the studies that have found non-significant group differences, including that of Hoff et al. (2014), must be interpreted with caution. The sample of *2L1 children* consisted of only 15 children; hence, the power to detect group differences was low. To my knowledge, there is no large-scale study investigating the language

skills of pre-adolescent *2L1 children* with one parent who is a native society language speaker. Analysis of data from the Norwegian PIRLS 2001 shows that monolingual learners and *2L1* 5<sup>th</sup> graders have equal reading comprehension skills (Wagner, 2004). Even though reading comprehension at this age depends heavily on oral language skills, it is not the same measure as that of language skills. It could, however, be seen as an indicator that this student group's language and literacy skills may differ from other bilingual learners' language and literacy skills.

#### **2.4.7 Socio- economic status (SES)**

SES describes an individual or family ranking on a hierarchy according to access to resources such as wealth, power or social status, education or occupational prestige (Bornstein & Bradley, 2014) and is a component often examined when predictors of academic achievement are studied.

SES in this thesis is limited to parental education level, and the words SES and parental education are used interchangeably. A meta-analysis of over 100,000 students found a relationship between parental education and academic achievement of a medium effect size ( $r = .27$ , 95% CI [.23, .39]) (Sirin, 2005). The effect size was moderated by academic outcome, with a larger correlation between parental education and verbal abilities than general academic outcome. Additionally, in a synthesis of a large body of research, Hoff (2013) concluded that children from low-SES homes have lower language skills than children from middle-SES homes. Children from low-SES homes lagged behind their middle-SES peers in a range of language abilities involving vocabulary size, grammatical development, narrative skills, phonological awareness and speed of language processing. Thus, the effects of SES on children's early language skills are large, pervasive and robust (Hoff, 2013).

However, SES is argued to be a proxy for something, representing a spectrum of many factors that may most likely have causal effects on children's learning outcomes (Kirby & Hogan, 2008). It is often

claimed that the SES effect is caused partly by parents directly providing resources at home for their children and indirectly by providing the social capital necessary to succeed in school. Common explanation factors are that qualities within mother and child dialog and the amount of storybook reading differ from one SES class to another (Hart & Risley, 2003; Hoff, 2006; Huttenlocher, Waterfall, Vasilyeva, Vevea, & Hedges, 2010; Rowe, 2012). Furthermore, parents with low SES living in poverty have less access to nutrition and health care (Black et al., 2017). Another explanation is that parents' beliefs and behaviours differ for different SES classes (Davis-Kean, 2005). All these factors can cause different developmental patterns in children from different SES classes.

Nonetheless, the importance of quality differences in parent-child interaction and basic living standards over the "poverty line" as the main explanation for why SES differences impact language outcomes has recently been challenged by, for instance, the study of Puglisi, Hulme, Hamilton, and Snowling (2017). Based on their findings, they argue that the relationship between home literacy activities and children's language and reading skills is largely accounted for by maternal skills and may reflect genetic influences. In fact, after controlling for variations in maternal language and phonological skills, Puglisi et al. (2017) find that storybook exposure is not a significant predictor of children's outcomes. Other studies find that environmental influences do matter, yet not as much as parents' genes. For instance, a review by Olson, Keenan, Byrne, and Samuelsson (2014) finds that although environmental influences are generally statistically significant for reading disabilities, the average influence of genes is twice as strong as that of (shared) environmental influences. The meta-analysis of de Zeeuw, de Geus, and Boomsma (2015) also points to the importance of parents' genes and concludes that genetic variation is an important contributor to individual differences in educational achievement. They find the heritability of reading to be 73% for reading, 66% for language and 49% for reading comprehension. Notably, environmental influences accounted for only 10% of variance



in reading, 13% of reading comprehension and 15% of language. Thus, the mechanism that underlies SES as a moderator of academic outcomes seems to be complex.

International and national studies find that the population of *minority language-speaking* children is over-represented in the national statistics of low-SES families (Barne- likestillings- og inkluderingsdepartementet, 2012; Capps et al., 2005). Fifty-one percent of the children living in poverty in Norway during 2011-2013 were children of *minority language-speaking* immigrants (Justis- og beredskapsdepartementet, 2015). It has been argued that the low SES of *minority language-speaking* students could be causing the achievement gap (Bakken, 2003). It is therefore common to statistically control for SES when comparing monolingual and bilingual learners' language and reading abilities (e.g., Bakken & Elstad, 2012; Kieffer, 2012a; Strand, Wagner, & Foldnes, 2017).

A recent meta-analysis finds that the differences in oral language skills between monolingual and bilingual children in the instructional language were larger for children from low-SES backgrounds than for children from high-SES backgrounds (Melby-Lervåg & Lervåg, 2014). However, importantly, when SES was controlled for, the group differences were moderate to large. This is in line with the large-scale longitudinal ECLS-K study results of Halle et al. (2012), who find low-SES bilingual children less likely than their bilingual peers of higher SES to be proficient in English before entering kindergarten.

To date, bilingual children from low- and middle-class SES backgrounds have been the main focus of research. A critical review of knowledge of bilingual learners' literacy and language skills identifies a lack of large-scale studies of high-SES bilingual children's language and reading development (Hammer et al., 2014). The relationship between high-SES bilingual children's SES background and language skills is therefore uncertain. However, because there is a relationship between bilingual

learners' SES and academic outcome (Sirin, 2005), although not as strong as for monolingual learners (Hermansen, 2009, 2013; Sirin, 2005; Steffensen & Ziade, 2009), one should expect the gap in language skills between monolingual and bilingual learners' language skills in the instructional language to be smaller for high-SES children. This is supported by findings in the "Miami" study, one of the few studies with a large sample of high-SES bilingual learners (Oller, Jarmulowicz, Pearson, & Cobo-Lewis, 2011). Unpublished findings from the National Household Education Survey of 2001, although measuring different skills, present an even more optimistic view (Woodard & Rodman, 2007). The researchers suggest that SES for bilingual learners works as a mediating factor of cognitive development. In a conference paper, they indicate that a high SES level could work as a threshold value and enable the high-SES bilingual child to outperform the high-SES matched monolingual child in basic math and reading skills.

#### **2.4.8 Research gaps**

Descriptions of sample characteristics are typically lacking or limited in primary studies of bilingual learners; thus, it is difficult to draw conclusions, to generalize or to make comparisons between studies (Hammer et al., 2014). Based on their review, Hammer et al. (2014) request that more studies provide information on variables such as bilingual learners' exposure to and usage of language, timing of exposure (AoA) and demographic characteristics. There is a specific lack of studies of high-SES children's language and literacy skills and a critical need for large-scale studies of bilingual groups other than Spanish-/English-speaking ones (Hammer et al., 2014).

## **2.5 The connection between language and reading comprehension for the bilingual reader**

According to the simple view of reading (Hoover & Gough, 1990), reading comprehension is a product of decoding skills and linguistic comprehension, and different aspects of skills play different roles at different timepoints during the development of reading skills. During the first years of development, decoding skills play a central role. Hoover and Gough (1990) defined decoding as efficient word recognition. A skilled decoder can read isolated words quickly and accurately. However, in order to understand the content of a text, one must combine meaning on the word level and derive sentence and discourse interpretations (Hoover & Gough, 1990). As the complexity of texts increases, so do the demands of linguistic comprehension (Storch & Whitehurst, 2002). As soon as decoding skills are functioning at an automatic level, linguistic skills, as in the Hoover and Goughs model, where they are also referred to as listening comprehension, play the most dominant role (Hoover & Gough, 1990). This pattern is the same across transparent (Hjetland et al., 2018; Lervåg, Hulme, & Melby-Lervåg, 2018) and non-transparent languages (García & Cain, 2014), as well as across the monolingual and bilingual reader (Geva & Farnia, 2012; Verhoeven & van Leeuwe, 2012).

Although both decoding and listening comprehension are important for both the monolingual and bilingual reader, these skills do not seem to be equally distributed across language groups. As mentioned in section 2.4.1 in this thesis, a meta-analysis investigating this matter found a monolingual advantage of 1.12 *d* in language skills (Melby-Lervåg & Lervåg, 2014). The same analysis found a monolingual advantage in reading comprehension in .62 *d* yet equal decoding skills across language groups. The gap in reading skills is perhaps not surprising given previously presented studies of bilingual learners' language skills, where bilingual learners as a group have lower second language skill than their

monolingual peers. It is then interesting to investigate differences in linguistic skills across monolingual and bilingual learners as well as possible differences in predictive patterns of reading comprehension.

### ***2.5.1 Bilingual children's mastery of specific linguistic aspects in L2***

To fully map bilingual children's linguistic abilities, there is a need to examine bilingual children's competence in both their languages. The reason is that a bilingual child's language competence is distributed across two languages (Monsrud, Rydland, Geva, Thurmann-Moe, & Lyster, 2019; Oller, 2005). They know some words in their L1 and somewhat different words in their L2. Therefore, when this thesis solely investigates bilingual children's L2 skills, only a part of their linguistic competence is examined. Such examination is, however, important since bilingual children's L2 skills strongly predict their academic outcomes (Halle et al., 2012; Han, 2012; Kieffer, 2008).

In this thesis, bilingual learners mastery of different linguistic aspects in L2 will be explored. This exploration is mostly limited to the linguistic skills of pre-adolescent (8-12-year-old) bilingual learners. Studies of emergent readers (1<sup>st</sup>-2<sup>nd</sup> graders) and adolescents (13-15-year-old) are sometimes included, often to exemplify the development of the targeted L2 skills. Due to the lack of studies of pre-adolescent *early bilingual learners* examining this topic, most of the presented studies are of *minority language learners*. For these examples to deviate the least from the expected skills of *early bilingual learners*, almost all the studies are of *minority language learners* exposed to at least 4-9 years of L2 (some of the studies of emergent readers are less exposed to L2). This timeframe of L2 exposure somewhat overlaps with the timeframe claimed to be sufficient for *minority language learners* to obtain a level in the instructional language within the normal variation of their monolingual peers (Jim Cummins, 1984, 2017; Hakuta et al., 2000; Saunders & O'Brien, 2006). Note, however, that how well the results of

these findings transfer to studies of *early bilingual learners*, or whether the linguistic abilities of *early bilingual learners* more closely resemble those of monolingual learners, is uncertain.

There are several relevant research questions targeting this topic. Whether some specific linguistic aspects in L2 are more delayed than others compared to monolingual peers is, for instance, uncertain. If so, this could provide grounds for arguing that it is more useful to target some linguistic aspects for intervention than others. Of the linguistic aspects under investigation in this section of the thesis are L2 vocabulary breadth and depth, morphology, listening comprehension and L2 text cohesion vocabulary.

*Vocabulary breadth* is the number of known words, and *vocabulary depth* is the richness of knowledge about the known words (Anderson & Freebody, 1981; Stahl & Nagy, 2006). The difference between these two aspects can be validated through observations of early childhood language development (Ouellette, 2006). The knowledge of newly learned words is restricted, without a full understanding of the meaning of the words (Nagy & Herman, 1987). However, as the child grows, word meanings are refined, adding to the child's depth of vocabulary knowledge. The extension of knowledge spans from some familiarity with a word to a full understanding of various meanings of the word, the syntactic and morphological knowledge involved in using the word in different contexts, and knowledge of the word's superordinates and semantically related words (Nagy & Scott, 2000). However, whereas some researchers claim that vocabulary breadth and vocabulary depth are different dimensions of vocabulary, others find the relationship between these aspects of vocabulary so strongly correlated that they may be interchangeable variables (Harkio & Pietilä, 2016; Kieffer & Lesaux, 2012; Li & Kirby, 2014). L2 vocabulary depth and breadth have been extensively investigated. A number of studies of *minority language learners* have detected low vocabulary depth and breadth compared to monolingual peers across different age spans and SESs (Kieffer &

Lesaux, 2012; Leider, Proctor, Silverman, & Harring, 2013; O'Connor, Geva, & Koh, 2019; Oller, 2005; Silverman et al., 2015; Verhoeven, 2000; Verhoeven, Voeten, & Vermeer, 2019), even for high-frequency words (Schwartz & Katzir, 2012; Verhallen & Schoonen, 1993).

Whether the vocabulary gap between bilingual and monolingual learners is also evident for older samples of *bilingual first children* or *early bilingual learners* has been less studied. The studies that have targeted vocabulary in this group have focused mainly on vocabulary breadth. The study of Bialystok and Feng (2011) examined *bilingual first children's* vocabulary skills on an aggregated dataset consisting of 16 studies, a sample of 963 children across the age span of 5-9 years. They found a large vocabulary gap across all age groups in favour of monolingual learners. This is in line with the results of Grant et al.'s (2011) study of *bilingual first 9-year-olds*. Studies of *early bilingual learners* show mixed results, ranging from no differences between monolingual and bilingual 1<sup>st</sup>-6<sup>th</sup> graders introduced to L2 by at least the age of 2 (Hsu, Ip, Arredondo, Tardif, & Kovelman, 2019) to a large gap in disfavour of bilingual 1<sup>st</sup>-4<sup>th</sup> graders systematically exposed to L2 from at least the age of 3 (Vernice & Pagliarini, 2018). The results of the latter study and the one by Grant et al. (2011) could, however, be influenced by the danger of small sample sizes and hence low statistical power (Ingre, 2013).

*Text cohesion vocabulary* is words/expressions that refer to inter-clausal relationships and work as guiding cues to assist listeners in understanding how an idea in one clause is related to those in adjacent clauses (Crosson, Lesauc, & Martiniello, 2008). "In spite of", "in contrast to", "since" and "therefore" are examples of text cohesion vocabulary. If the meaning of text cohesion vocabulary is confused, the meaning of a whole sentence/section of narrative might change. One example of such a mix-up would be if the word "because" were understood as "in spite of" in the following sentence: "Alan is happy because he has attended a football match". Thus, such linguistic devices

carry a high level of meaning (Crosson et al., 2008). Furthermore, if the meaning of text cohesion vocabulary is unknown, knowing all other words in the sentence will not help the listener's interpretations of the section of the narrative that contains the unfamiliar text cohesion (Crosson et al., 2008). Thus, unlike other dimensions of vocabulary, learning the meaning of new text cohesion vocabulary through interpretations of a narrative is difficult.

L2 text cohesion vocabulary is investigated less than other L2 aspects. One of the few studies comparing monolingual and bilingual pre-adolescents' text cohesion vocabulary skills, in this case on a sample of *minority language learners*, found a large gap in text cohesion vocabulary (-1.04 *d*) between *minority language learners* and their monolingual 4<sup>th</sup> grade peers (Droop & Verhoeven, 2003) but less of a gap than for morphological knowledge and vocabulary skills. Empirical studies of *minority language learners* show that L2 listening comprehension and L2 vocabulary both explain variations in *minority language learners'* L2 text cohesion vocabulary; thus, text cohesion vocabulary is suggested to play a role in the underlying abilities of oral language that *minority language learners* need in order to be able to grasp complicated narratives (Crosson et al., 2008). Although other linguistic variables can explain some of the variation in text cohesion vocabulary, CFA modelling provides grounds for considering this aspect of language as a separate construct (Droop & Verhoeven, 2003). There is a lack of studies investigating the level of text cohesion vocabulary across monolingual and *early bilingual learners*. How *early bilingual learners* master this linguistic aspect is therefore uncertain.

There is more disagreement on the nature of L2 morphological knowledge/metalinguistic abilities. It has been theorized that since bilingual children know two or more languages, their awareness of linguistic operations and analytic orientations to linguistic input might be a strength in comparison to monolingual learners (James Cummins, 1978; Jim Cummins, 1987). The long-standing assumption is that

bilingualism leads to an understanding of their language as one particular system among many, which again enhances metalinguistic awareness (Vygotsky, 1964). The meta-analysis of Adesope, Lavin, Thompson, and Ungerleider (2010) could be seen as support for the claims of bilingualism leading to superior metalinguistic abilities. One variable of metalinguistic awareness is morphological awareness (Berthiaume, Daigle, & Desrochers, 2018). Some researchers have separated morphological awareness from morphological knowledge (Bialystok, 2001a), whereas others do so only partly (Kuo & Anderson, 2006) or not at all (Carlisle & Feldman, 1995). Morphological knowledge can be defined as knowledge of the smallest meaningful units of language and how to use word-building rules to construct and understand morphologically complex words (Kuo & Anderson, 2006).

Friesen and Bialystok (2012) argue that bilingual children's performance on morphological awareness differs in tasks that draw highly on executive control and more purely linguistic knowledge tasks. In their article, they synthesize evidence of bilingual learners' performance on the Wug test (Berko, 1958). The Wug test targets children's awareness of inflections. An example of a test item is "Here is a Wug, here are two ...". The child's task is to fill in the last word in the sentence, in this case adding the inflection of plurality to the nonsense word Wug. According to the authors, the executive function demands for this task are low, yet the linguistic demands are fairly high. The reason is that the children must supply the correct morphological form. Since monolingual learners often show superior performance in linguistic tasks (Bialystok & Feng, 2011; Bialystok, Luk, et al., 2010), Friesen and Bialystok (2012) explain that one should expect monolingual learners to outperform bilingual learners on this task. However, in a study by Barac and Bialystok (2012), bilingual learners outperformed monolingual learners. Friesen and Bialystok (2012) argue that the reason was that the two groups had equal levels of the instructional language, allowing the bilingual advantage to emerge even in a task with high language demands. A study by



Bialystok, Peets, et al. (2014) seconds these findings. After two years of attendance in immersion school, bilingual children outperformed monolingual children on the Wug test (measured in L1), and the gap in favour of the bilingual learners increased after 5 years of immersion.

Except for the studies of Bialystok, no other studies differentiate morphological tasks based on the degree of misleading information they contain or how balanced bilingual children might be in their L1/L2. Furthermore, studies of *early bilingual learners* targeting morphological skills, most likely to be more comparable with monolingual learners in terms of linguistic skills in the instructional language, are limited. Vernice and Pagliarini (2018) examine bilingual learners with an AoA before the age of 3. In a sample of 2<sup>d</sup>-4<sup>th</sup> graders, they find a large effect size in favour of monolingual learners. However, the gap was smaller for morphological skills than for vocabulary, indicating a relative advantage in morphological abilities. This is, however, a small study, and clear conclusions therefore cannot be drawn. The study of Hsu et al. (2019) has a larger sample size and is thus more trustworthy. Hsu et al. (2019) find no significant differences between monolingual and *early bilingual learners* for either vocabulary or morphology. If the assumption of Friesen and Bialystok (2012) is right, given equal vocabulary skills, one should expect the presumed bilingual advantage to occur; thus, the bilingual learners should have outperformed their monolingual peers, yet they did not. Thus, the results of *early bilingual learners'* morphological skills are mixed. It is therefore not clear whether the morphological abilities of *early bilingual learners* are in line with or even better than monolingual language learners' skills, comparable with *minority language learners'* skills or inhabiting a skill level somewhere in between these two groups.

Studies of *minority language learners* targeting morphological skills show that the gap between monolingual and *minority language learners* (2<sup>nd</sup> graders, 4<sup>th</sup> graders and children in kindergarten [4-5-year-olds]) is large and favours monolingual learners (Droop & Verhoeven, 2003;

Verhoeven et al., 2019). Furthermore, in most of these studies, the gap is roughly comparable to the gap in vocabulary skills. In contrast to these studies, Lipka and Siegel (2012) find the morphological skills of monolingual and *minority language* 7<sup>th</sup> graders to be equal. This study did not provide measures of vocabulary. It is therefore difficult to compare gaps between morphological skills and vocabulary. However, measures of syntactic awareness and sentence repetition are known to highly correlate with other language skills (Klem et al., 2015). When comparing bilingual and monolingual learners on these skills, the gap in language skills is small (.26 *d* and .05 *d*, respectively). This might suggest that given enough exposure to L2 and small differences in overall language skills in L2, *minority language learners* can catch up with their monolingual peers on morphological skills.

*Listening comprehension* is the ability to understand language and can be assessed by presentations of stories followed by questions related to the contexts of the narrative (Hoover & Gough, 1990). This requires interpreting semantic information at the word level (vocabulary) and deriving sentence and discourse meaning (Hoover & Gough, 1990). Listening comprehension can by this definition be regarded as a multiple construct consisting of all aspects of linguistic subskills, such as vocabulary, morphology, syntax, and inferences.

Most studies comparing the listening comprehension of pre-adolescent bilingual learners to that of monolingual learners are of *minority language learners*. Such studies vary in their results, ranging from a large gap in listening comprehension in favour of monolingual 3<sup>rd</sup>-5<sup>th</sup> grade peers (Droop & Verhoeven, 2003; Hutchinson, Whiteley, Smith, & Connors, 2003; Verhoeven & van Leeuwe, 2012) to a difference of a medium effect size (Burgoyne, Whiteley, & Hutchinson, 2011; Farnia & Geva, 2013; Geva & Farnia, 2012) to no significant differences in spite of good sample sizes (O'Connor et al., 2019). Due to a lack of studies, it is unknown whether the same mixed results will be found when examining *early bilingual learners*' listening comprehension.

The study of Bonifacci and Tobia (2016) is one of the few to investigate *early bilingual learners'* listening comprehension. The study has a robust sample size, yet its sample of bilingual learners is introduced to L2 from birth until the age of 4. They compare monolingual learners (mean age 8.69 years) and bilingual children (mean age 8.72 years) in the age span of 1<sup>st</sup>-5<sup>th</sup> grade on listening comprehension skills and found no significant differences across language groups.

One of the very few studies that target larger overarching oral language structures in pre-adolescent bilingual learners with lower AoA than that of the sample in Bonifacci and Tobia (2016) is the study of Kovelman et al. (2008). Kovelman et al. (2008) examine *early bilingual learners* with an AoA from birth until the age of 3, yet instead of targeting listening comprehension, they measure *early bilingual learners'* expressive language competence. The *early bilingual learners* were asked to re-tell the content of a 1½-minute carton video, and the participants' grammatically correct utterances and the amount of story events they produced were scored. They find that the *early bilingual learners* exhibited the same quality of narrative skills as their monolingual 3<sup>rd</sup> grade peers. Based on this outcome, Kovelman et al. (2008) argue that AoA is a better predictor of L2 success than years of exposure, claiming that *early bilingual learners* have the best possibilities for a good outcome.

However, even though the studies of Kovelman et al. (2008) and Bonifacci and Tobia (2016) target overarching linguistic structures, the constructs they examine are different; therefore, comparisons should be made with caution. The task of Kovelman et al. (2008) does not provide language scores of the complexity of utterances, just how many events were reported and whether the grammatical structures of these utterances were correct. Listening comprehension involves understanding complex text in terms of the meaning of both specific words and complex syntactic structures. To be able to answer questions on such a text, just from hearing it without the acumination of visual support (as a cartoon

video would provide), the difficulty level could very well be considered higher than that of re-telling the content of a video. In the re-telling task, the *early bilingual learners* could choose the complexity level of the vocabulary and syntactic structures they used without their choices hampering the test results. However, both results point in the same direction, and there seems to be no difference in listening comprehension/larger overarching language structures across monolingual and *early bilingual learners*. However, this is an understudied topic.

*Reading comprehension of early bilingual learners* is more studied than listening comprehension. Kovelman et al. (2008) compare monolingual and *early bilingual* 2<sup>nd</sup> and 3<sup>rd</sup> graders' reading comprehension. The bilingual learners had an AoA from birth until 3 years of age. They found a large difference in reading comprehension (1.47 *d*) in favour of the monolingual children. The results of this study are, however, vulnerable to possible overestimation of effect size, and even the direction of the effect size, due to the sample size (N=25) of the bilingual group (Ingre, 2013). A study by Bonifacci and Tobia (2016) also examines *early bilingual learners'* reading comprehension (AoA birth-4 years old) yet finds only a medium gap (.69 *d*) in reading comprehension. One should perhaps expect the reading comprehension skills of *bilingual first children* to resemble those of monolingual learners even more than the studies of *early bilinguals*. Studies of bilingual learners at birth, however, show conflicting results, from no significant differences (Wagner, 2004) to a medium gap (.69 *d*) in reading comprehension (Grant et al., 2011). Different tests of reading comprehension are often only moderately correlated (Leider et al., 2013) and could tap different aspects of reading (Keenan, Betjemann, & Olson, 2008). Wagner measures reading comprehension using data from the Norwegian PISA investigation in 2003, whereas Grant et al. (2011) use the Nara-a test, which relies heavily on readers' linguistic skills (Nation, 2006). Another difference is that the participants in Wagner's study were 5<sup>th</sup> graders,

whereas Grant et al studied 3<sup>rd</sup> graders. It is therefore also possible that *bilingual first learners* have a different growth trajectory than monolingual learners, and that a potential gap in reading comprehension is evened out by 5<sup>th</sup> grade.

### ***2.5.2 Differences and similarities in predictive patterns from linguistic skills (and decoding) to reading comprehension in monolingual and bilingual readers***

Differences and similarities in predictive patterns could imply two different questions. First, do the same variables of oral language predict reading comprehension at the same timepoints across language groups? Next, are the strengths of these predictions the same across language groups? Both these aspects of differences across predictive patterns will be explored below. The presentation of study outcomes is, however, limited to the predictive patterns of the linguistic variables presented in section 2.5.1 in this thesis. Since most of these studies target *minority language learners*, such studies will be presented first.

Based on the findings of equal validity for the simple view of reading across monolingual and bilingual readers (Farnia & Geva, 2013; Lesaux, Rupp, & Siegel, 2007; Verhoeven & van Leeuwe, 2012; Verhoeven et al., 2019), one should expect reading comprehension to be built on the same building blocks. Research shows mixed results regarding equal predictive patterns across language groups despite such evidence. Some studies of *minority language learners* find that the same variables predict reading comprehension across groups, even with equal strength (Babayigit, 2015; Verhoeven et al., 2019), yet other studies find that different variables predict reading comprehension across language groups (Bellocchi, Tobia, & Bonifacci, 2017; Burgoyne et al., 2011; Geva & Farnia, 2012; Hutchinson et al., 2003; Limbird, Maluch, Rjosk, Stanat, & Merkens, 2014).

Limbird et al. (2014) use data from a 3-year longitudinal study of 100 bilingual and 69 monolingual learners. They find that phonological awareness and decoding skills in 2<sup>nd</sup> graders predicted monolingual learners' reading comprehension in 3<sup>rd</sup> grade, yet vocabulary was an additional predictor of bilingual learners' reading comprehension skills. The study lacks usage of linguistic structures other than vocabulary as possible predictors. The studies of Geva and Farnia (2012) and Hutchinson et al. (2003) include a broader range of linguistic predictors. Hutchinson et al. (2003) trace 43 monolingual and 43 *minority language learners* from 2<sup>nd</sup> through 4<sup>th</sup> grade and find that 2<sup>nd</sup> grade reading comprehension predicted 4<sup>th</sup> grade reading comprehension across language groups. However, only morphosyntax (TROG) contributed uniquely to the monolingual 4<sup>th</sup> graders' reading comprehension, while vocabulary did so for the bilingual learners. Due to sample sizes, the results from this study must, however, be interpreted with caution. The study of Geva and Farnia (2012) has a much more robust sample size, with 390 *minority language learners* and 149 monolingual learners. The study has a longitudinal design and follows the children from 2<sup>nd</sup> to 5<sup>th</sup> grade. It finds that 2<sup>nd</sup> grade vocabulary skills predicted both monolingual and *minority language learners'* reading comprehension in 5<sup>th</sup> grade. Of concurrent predictors provided syntactic and listening comprehension skills, a unique contribution to 5<sup>th</sup> grade reading comprehension for the *minority language readers*, yet only 5<sup>th</sup> grade vocabulary did so for the monolingual reader. Burgoyne et al. (2011) study the impact of some of the same linguistic aspects as those selected by Geva and Farnia (2012) on reading comprehension yet include a more limited number of predictors. In contrast to the study of Geva and Farina (2012), Burgoyne et al. (2011) find that listening comprehension predicted monolingual learners' reading comprehension and vocabulary predicted bilingual learners' reading comprehension. There might be several reasons why these two outcomes differ. First, Burgoyne et al. (2011) examine this relationship across 2<sup>nd</sup> and 3<sup>rd</sup> graders, not 5<sup>th</sup> graders. It is possible that vocabulary skills play a more dominant role in

*minority language learners'* reading comprehension at younger ages, yet as they grow older and develop more complex and competent L2 skills, L2 listening comprehension plays the most dominant role. It is also possible that the results of Burgoyne et al. (2011) are hampered by the small sample size; they investigate the reading skills of only 39 monolingual and 39 bilingual children. Unfortunately, many of the studies targeting this theme are small, as is the study of Bellocchi et al. (2017) investigating the reading comprehension of somewhat younger children than those who are the main focus of Study 3 in this thesis. They follow 30 *minority language* and 56 monolingual emergent readers from 1<sup>st</sup> through 2<sup>nd</sup> grade yet include different variables, and several more, than the study of Burgoyne et al. (2011). They find that both morphosyntax and vocabulary predicted the monolingual children's reading comprehension, yet only morphosyntax predicted the bilingual children's reading comprehension. This challenges the results of the study of Hutchinson et al. (2003), in which vocabulary predicted bilingual learners' reading comprehension, yet morphosyntax predicted monolingual learners' reading comprehension. There is therefore some evidence for different linguistic constructs predicting reading comprehension across language groups and some support for *minority language learners'* reading comprehension drawing on more linguistic constructs than that of monolingual learners. Due to a lack of studies, little is known about this relationship for *early bilingual learners*.

One of the few studies examining the predictive pattern of linguistic skills in relation to reading comprehension in pre-adolescent bilingual learners is the one by Grant et al. (2011). They trace the language and reading skills of 29 monolingual and 32 *bilingual first learners* from 2<sup>nd</sup> through 3<sup>rd</sup> grade. As with most of the studies of *minority language learners*, this study also finds a different predictive pattern from language skills to reading comprehension across language groups. They find that both decoding skills and vocabulary predicted *bilingual first learners'* reading comprehension, yet only decoding predicted

monolingual learners' reading comprehension. Note that this is a small study and might have enough statistical power to detect true predictors, yet the results are the same as those of the much larger study of *minority language learners* by Limbird et al. (2014). The two studies have in common that the bilingual learners had much lower language skills than their monolingual peers. This might indicate that vocabulary skills play a much more dominant role at a younger age for both *bilingual first learners* and *minority language learners* than for monolingual children, at least when the two groups do not have equal language skills. This does not, however, imply that the relationship is stable across the years of pre-adolescents, where one might expect the gap in the instructional language skills to decrease.

However, most of the studies cited until now examining the unique impact of different linguistic constructs on reading comprehension have used manifest variables. Thus, measurement error might have hampered their results. From that perspective, the results of studies examining differences in predictive strength across language groups might be more interesting. Most of the studies cited below have used SEM analysis to examine this relationship using latent variables. Unfortunately, all these studies are of *minority language learners*.

Studies examining differences in the strength of predictors of reading comprehension across language groups often find oral language skills to be more crucial for bilingual learners' reading comprehension than for that of monolingual readers. This is supported by studies of the reading comprehension of emergent readers, somewhat younger *minority language learners* than those who are the main focus of this thesis (Lervåg & Aukrust, 2010; Verhoeven, 2000), as well as studies of pre-adolescent *minority language learners* (Proctor, Montecillo, Silverman, & Haring, 2012). Although the data are analysed differently than those in the prior cited studies, the findings of Kieffer and Vukovic (2013) also confirm a stronger relationship between vocabulary and reading comprehension for *minority language learners* than for monolingual



learners, at least during the years of first reading instruction. In their longitudinal study, correlations between vocabulary and reading comprehension increased for monolingual 1<sup>st</sup>-4<sup>th</sup> graders (.12, .16, .31, .43) but were consistently strong for bilingual learners (.49, .50, .40., .41). This pattern is supported by the study of Droop and Verhoeven (2003), which examines the predictive pattern of a broad range of linguistic structures to reading comprehension. Their longitudinal study examines the reading comprehension of children in 2<sup>nd</sup>-4<sup>th</sup> grades and finds that both listening comprehension and vocabulary are stronger for bilingual than for monolingual learners. Although some studies using latent variables in SEM analysis imply that L2 is a stronger predictor for *minority language learners*, there are also studies that find the strength of prediction to be equal across groups (Babayiğit, 2015).

### *2.5.3 Predictive patterns of specific L2 skills on reading comprehension in L2*

Two recent longitudinal studies of reading comprehension use latent variables and find that almost all variation in listening comprehension (95% and 97%) can be explained by one factor of verbal language skills, which consist of vocabulary, morphosyntax, listening comprehension, verbal working memory and inference skills (Hjetland et al., 2018; Lervåg, Hulme, & Melby-Lervåg, 2018). This implies that there is one underlying oral language skill that has an impact on reading rather than a range of linguistic subskills. Another SEM analysis finds more support for different linguistic dimensions affecting reading comprehension rather than one underlying ability. One such study uses a bifactor model to assess the relationship between vocabulary, syntactic awareness and morphological awareness in the reading comprehension of 311 3<sup>rd</sup>-5<sup>th</sup> graders (Kieffer, Petscher, Proctor, & Silverman, 2016). As a first step, the measurement model is tested. Different manifest variables of syntactic awareness, morphological awareness and vocabulary are used to create latent

variables of these constructs. This measurement model is tested across 3 versions of the model: a one-factor model, a correlated model and a bifactor model. The bifactor model allows all latent variables to remain separate latent constructs in addition to being merged together to create a new latent variable. All latent variables can be used in regressions in SEM models, thereby separating the contributions of the individual variables as well as the contribution of the common variable. The bifactor model fit the data well and was significantly better than the other models. Kieffer et al.'s (2016) study indicates that there is something in common for all the language variables but also something specific to each of these linguistic constructs. The common variable explained most of the variance in reading comprehension (.77), and the only other linguistic construct that significantly predicted reading comprehension was morphology (.16). However, none of the studies examining the dimensionality of oral language skills have investigated whether these results hold across the monolingual and bilingual reader.

There is some support for different linguistic constructs functioning as separate constructs across the monolingual and bilingual learner rather than just a common factor (Droop & Verhoeven, 2003; Kieffer & Lesaux, 2012). Kieffer and Lesaux (2012) use multi-group confirmatory factor analysis to test the model structure of a broad range of linguistic constructs on a sample of 583 6<sup>th</sup> graders. They find that linguistic skills can be identified as three highly related yet distinct dimensions: vocabulary, morphological awareness and contextual sensitivity. This model provided a better fit to the data than a two-factor and one-factor model. While testing other linguistic constructs in a longitudinal design, Droop and Verhoeven (2003) use the same approach as Kieffer and Lesaux (2012). First the validity of each construct is tested across time, and thereafter, it is separately fitted to the two language groups. Again, the confirmatory factor analysis identified distinct differences across the linguistic constructs. Morphosyntax, vocabulary, listening comprehension and reading

comprehension were all identified as different constructs. The model provided acceptable fit for both language groups, with a significantly better fit for the 4-factor solution than for 2- and 3-factor solutions. Based on these studies, one can find some support for different linguistic constructs being distinctly different from one another across the monolingual and bilingual reader.

Regarding the connection of different linguistic constructs in reading comprehension, the meta-analysis of Jeon and Yamashita (2014) examines which linguistic aspects explain most of the variation in L2 reading comprehension. They find that vocabulary skills and listening comprehension are the linguistic constructs that correlate most highly with L2 reading comprehension (.79, 95% CI [.69- .86] and (.77, 95% CI [.58- .88], respectively). Moderator analysis of vocabulary showed that age as a moderator approached significance, with a stronger correlation between L2 vocabulary and L2 reading comprehension for children than for adolescents/adults. Due to a lack of studies, text cohesion vocabulary is not included as a separate construct in this analysis, yet morphological knowledge is. The correlation between L2 morphological skills and L2 reading comprehension was also high, yet somewhat lower than for the other constructs ( $r = .61$ ). Based on information from this meta-analysis, there is some support for L2 vocabulary and listening comprehension being the strongest predictors of L2 reading comprehension; however, please note the methodological differences between assessing correlational data in a meta-analysis and regression analysis. While the meta-analysis of Jeon and Yamashita (2014) presents pure correlations among different outcomes, it does not provide information on the unique contribution each linguistic construct has to reading L2 comprehension. It is therefore possible that other linguistic constructs could work behind the scenes as third variables and in fact actually drive the high correlation identified between vocabulary and reading comprehension by Jeon and Yamashita (2014).

Several studies have examined the unique impact of specific linguistic aspects on L2 reading comprehension. The results from studies of pre-adolescent *minority language learners* with 4-7 years of L2 exposure do, however, show conflicting results. Some argue that vocabulary plays a critical role (Kieffer, 2012b; Silverman et al., 2015), while others support morphology (Kieffer & Lesaux, 2008), listening comprehension (Geva & Farnia, 2012) or text cohesion vocabulary (Rydland, Aukrust, & Fulland, 2012). (Other linguistic aspects are also argued to play a central role in L2 reading comprehension, but such results will not be reported here since that subject is beyond the focus of this dissertation.)

Unfortunately, many of the studies that have argued for the unique contribution of specific L2 aspects included a limited number of linguistic constructs as predictors of L2 reading. Moreover, most of them used manifest variables and hence did not control for measurement error. This random noise can threaten the validity or generalizability of a measure or construct (Little, 2013). The inclusion of random noise might also result in an overestimate of the impact of specific linguistic aspects on bilingual children's reading comprehension. Additionally, the majority of the studies relied on assumptions of linguistic constructs based on face validity rather than empirically investigating the underlying constructs. As Kieffer et al. (2016) put it when arguing for the importance of the empirical settlement of their bifactor model before approaching regression analysis, "If this assumption (of face validity) does not hold, then inferences about the underlying constructs' relations to outcomes are called into question". Hence, these methodological weaknesses of most studies on this topic could cause the differences in study outcomes. To avoid the presentation of lengthy clarifications of methodological weaknesses with single studies, the characteristics of studies included

in the upcoming review of research are clarified in Table 3 in Appendix 3.

*Vocabulary* is considered a fundamental component of the linguistic proficiency necessary for facile reading comprehension across the monolingual and bilingual reader (Geva, 2006). This is perhaps not surprising given that vocabulary is the building block of sentences. Without an understanding of the words of a sentence, the sentence cannot be understood. Several studies have also identified vocabulary skills as the most dominant variable in the prediction of L2 reading comprehension (Droop & Verhoeven, 2003; Hutchinson et al., 2003; Kieffer, 2012b; Leider et al., 2013; Proctor et al., 2012; Silverman et al., 2015).

Hutchinson et al. (2003) investigate the impact of listening comprehension, vocabulary and morphosyntax in 2<sup>nd</sup> grade on L2 reading comprehension in 4<sup>th</sup> grade. After 2<sup>nd</sup> grade reading comprehension was controlled, only vocabulary skills predicted L2 reading comprehension in 4<sup>th</sup> grade (Hutchinson et al., 2003). Leider et al. (2013) examine vocabulary and morphologic skills across two different reading comprehension tests with 123 participants in 3<sup>rd</sup>-5<sup>th</sup> grade. Vocabulary was the variable across both tests that best predicted reading comprehension. Kieffer, Biancarosa, and Mancilla-Martinez (2013) study the relationship between reading vocabulary, morphology and listening comprehension in the reading comprehension of *minority language learners* in 6<sup>th</sup>-8<sup>th</sup> grades. Again, vocabulary was identified as the main predictor.

Studies of growth in reading comprehension confirm the role of vocabulary in reading comprehension. Proctor et al. (2012) study the growth of L2 reading comprehension from 2<sup>nd</sup>-4<sup>th</sup> grade using vocabulary and morphology as predictors. Only vocabulary predicted initial reading comprehension, and it was the only variable to predict growth in reading comprehension for *L2 learners*. Silverman et al.

(2015) study 173 *minority language learners* and 213 monolingual learners in a cohort-sequential design with three cohorts providing data of children in the age span of 2<sup>nd</sup>-4<sup>th</sup> graders. All children were tested at four timepoints across the next two years. Vocabulary skills predicted the initial status of reading comprehension as well as growth, while morphological skills were unrelated to reading comprehension. The results were the same across language groups. Kieffer (2012b) examines the prediction of listening comprehension and vocabulary skills on the growth of L2 reading comprehension by following 295 *minority language learners* in kindergarten through 8<sup>th</sup> grade. Of the *minority language learners*' linguistic skills in kindergarten, vocabulary alone predicted 3<sup>rd</sup> grade reading comprehension (intercept), yet vocabulary did not predict further growth in reading comprehension. He concludes that once vocabulary is considered, the inclusion of other linguistic constructs to predict reading comprehension is redundant. Silverman et al. (2015) CFA-tests every linguistic construct before using the sum scores as predictors in the growth model. The morphology construct used in the analysis had a good fit statistic, yet vocabulary depth did not. Droop and Verhoeven (2003) have the strongest design of the presented examples, yet they use a rather small sample given the complexity of the model (Wolf, Harrington, Clark, & Miller, 2013). When studying the relationship between and development of L2 reading comprehension and vocabulary, listening comprehension and morphology from the beginning of 3<sup>rd</sup> grade through the end of 4<sup>th</sup>, they apply separate SEM models for monolingual and *minority language learners*. The vocabulary skills of the L2 children at the end of 3<sup>rd</sup> grade were the strongest predictor of 4<sup>th</sup> grade reading comprehension. Based on these studies, there is evidence that initial vocabulary skills predict pre-adolescents' reading comprehension skills as well as growth in reading comprehension across different age spans and reading comprehension tests.

*Morphology.* Morphological awareness has an impact on reading in several aspects. Morphological awareness enhances decoding and fluency in reading and vocabulary and is argued to have both an indirect and direct effect on reading comprehension (Carlisle, 2000; Chen, Ramirez, Luo, Geva, & Ku, 2012; Kieffer, 2014; Levesque, Kieffer, & Deacon, 2017; Lyster, Lervåg, & Hulme, 2016). Having knowledge of the smallest meaning-bearing language units is a clear advantage in efficient word recognition. Morphological knowledge helps the reader to recognize small familiar units in new words (Jeon & Yamashita, 2014) and could therefore, with support from the contexts of the text, enable understanding of the newly introduced and until now unfamiliar word. The presentation of studies that have investigated the impact of morphological skills on reading comprehension is limited to the main focus of this dissertation – the direct effect of different linguistic constructs on reading comprehension. Indirect effects of morphological skills on *minority language* pre-adolescents' reading comprehension through decoding or vocabulary skills are therefore not included.

As several of the examples listed under the explanation of the impact of vocabulary skills on reading comprehension show, the impact on morphology on reading comprehension is often redundant when the studies have controlled for a set of other linguistic variables. Even studies that identify a direct impact of morphology on pre-adolescents and adolescent *minority language learners'* reading comprehension often find that vocabulary skills explain a slightly larger part of reading comprehension than morphology does (Droop & Verhoeven, 2003; Kieffer, Biancarosa, et al., 2013; Kieffer & Lesaux, 2008; Leider et al., 2013).

Several studies have tested the impact of morphology on the same sample across multiple reading comprehension tests. Leider et al. (2013) test 3<sup>rd</sup>-5<sup>th</sup> graders with two tests. They find that morphology explained a good proportion of reading comprehension when assessed with the Woodcock Passage Comprehension subtest (Woodcock, Muñoz-

Sandoval, Ruef, Alvarado, & Wendling, 2005); however, only vocabulary skills predicted reading comprehension measured by the Gates-MacGinities Reading Comprehension Test (MacGinitie, MacGinitie, Maria, & Dreyer, 2002). This confirms the findings of Keenan et al. (2008) that different reading comprehension tests measure different constructs. It is also interesting that although vocabulary explained more than morphology, the prediction strength of the linguistic constructs was quite even. Different test types therefore seem to matter, although this does not seem to explain the full picture. Another factor that might influence the relationship between L2 morphology and reading comprehension is participant age. Kieffer and Lesaux (2008) investigate the prediction of vocabulary skills and morphology across the same reading comprehension tests as Leider et al. (2013) at two timepoints in a longitudinal study. They find that 4<sup>th</sup> grade reading comprehension was not predicted by 4<sup>th</sup> grade morphological knowledge, whereas 5<sup>th</sup> grade reading comprehension was predicted by 5<sup>th</sup> grade morphological skills. In fact, 5<sup>th</sup> grade morphology was the strongest predictor of reading comprehension assessed by the Woodcock Passage Comprehension subtest (Woodcock et al., 2005) and the second strongest predictor of reading comprehension when assessed by the Gates-MacGinities Reading Comprehension Test (MacGinitie et al., 2002). Leider et al. (2013) find no effect on reading comprehension on the Gates-MacGinities Reading Comprehension Test; however, they investigate this relationship across 3<sup>rd</sup>-5<sup>th</sup> graders. Based on these findings, morphological skills might play a direct role in 5<sup>th</sup> graders' and older bilingual learners' reading comprehension but not in younger bilingual learners' reading comprehension. Thus, its impact might vary as a function of which test is used to measure reading skills. This assumption is supported by a study of 6<sup>th</sup>-8<sup>th</sup> graders where morphology did predict reading comprehension directly when assessed with a group reading assessment and diagnostic evaluation test (Williams, 2001), yet not more than vocabulary (Kieffer, Biancarosa, et al., 2013)



The prediction of *text cohesion vocabulary* of reading comprehension also differs for different reading comprehension tests. Rydland et al. (2012) use multiple regression to examine the prediction of vocabulary breadth, text cohesion vocabulary and vocabulary depth for reading comprehension across 67 *minority language* 5<sup>th</sup> graders. Text cohesion vocabulary was the predictor with the largest unique contribution to reading comprehension assessed by the Woodcock Passage Comprehension test (Woodcock et al., 2005), while only vocabulary (depth) explained any variance in a reading comprehension test administered by the researchers to measure content-area reading comprehension. In the latter test, the participants read multiple texts on a topic and then answered questions related to the texts. Crosson et al. (2008) do not investigate the relationship between reading comprehension and text cohesion vocabulary, and Droop and Verhoeven (2003) use a different version of the test. The participants in Droop and Verhoeven's (2003) study read the test items themselves and filled in the text cohesion vocabulary within a multiple-choice format to provide a meaningful sentence. This construct was then used along with two other tests to create a latent variable of reading comprehension. What we know from studies of monolingual pre-adolescents is that difficulties with text cohesion vocabulary hamper reading comprehension (Cain, Patson, & Andrews, 2005; Geva & Ryan, 1985). As the child ages and is introduced to increasingly more complex texts in schools, the frequency of text cohesion vocabulary increases (Cain et al., 2005; Geva, 2007).

*Listening comprehension* is one of the main components in the simple view of reading and can therefore be considered a unique correlate of reading comprehension (Hoover & Gough, 1990). More precisely, according to "the simple view", the same abilities are involved in listening comprehension as in reading comprehension, yet the latter process relies on graphic-based information arriving through the eye instead of sounds perceived through the ear (Hoover & Gough, 1990). Listening comprehension is the understanding of speech and meaning

provided at a normal pace and involves understanding phonemes and vocabulary to derive sentence and discourse interpretations (Hoover & Gough, 1990).

Although there is a close relationship between reading and listening comprehension, the predictive pattern of L2 listening comprehension for L2 reading comprehension has been studied less than, for instance, the predictive pattern of L2 morphology and L2 vocabulary for L2 reading comprehension (Jeon & Yamashita, 2014). Several studies find that after controlling for a range of other linguistic constructs, listening comprehension does not predict reading comprehension skills for the *minority language* pre-adolescent/adolescent reader (Burgoyne et al., 2011; Hutchinson et al., 2003; Kieffer, 2012b; Kieffer, Biancarosa, et al., 2013). This relationship has been proven across different sets of reading comprehension tests, including NARA. Most of the studies have found that vocabulary predicts reading comprehension after controlling for listening comprehension. Burgoyne et al. (2011) find a close relationship between vocabulary and listening comprehension. L2 listening comprehension in 3<sup>rd</sup> and 4<sup>th</sup> grades predicted 3<sup>rd</sup> and 4<sup>th</sup> grade L2 reading comprehension only when the variable was entered ahead of vocabulary in the analysis. This implies that these constructs, as they are measured, overlap.

In Droop and Verhoeven's (2003) study, listening comprehension was a strong predictor of 2<sup>nd</sup> grade reading comprehension and thereby indirectly predicted reading comprehension for 3<sup>rd</sup> graders. There are also studies that find listening comprehension to be the strongest predictor of L2 reading comprehension (Geva & Farnia, 2012). Geva and Farnia (2012) assess the relationship between vocabulary, listening comprehension and reading comprehension for 390 *minority language learners*. Linguistic skills measured in 2<sup>nd</sup> grade are used as auto-regressors to assess the impact of concurrent listening comprehension and vocabulary skills on 5<sup>th</sup> grade reading comprehension. Here, 5<sup>th</sup> grade listening comprehension predicted reading comprehension but not

vocabulary skills measured in 5<sup>th</sup> grade. Of the auto-regressors, only 2<sup>nd</sup> grade vocabulary was significant. However, the impact of 5<sup>th</sup> grade listening comprehension was stronger than that of 2<sup>nd</sup> grade vocabulary skills.

In summary, judging from the cited studies, there seems to be more support for vocabulary playing the most dominant role in pre-adolescents' *minority language learners'* reading comprehension not listening comprehension. Note however that most of the studies investigating the matter have used manifest variables. Furthermore, there seems to be evidence of an overlap between listening comprehension and vocabulary measures in these studies, suggesting that they might tap the same underlying language ability. This could potentially explain the similarity of correlation strength identified in the meta-analysis of Jeon and Yamashita (2014) between L2 vocabulary and L2 reading comprehension ( $r = .79$ ) and L2 listening comprehension and L2 reading comprehension ( $r = 0.77$ ).

#### **2.5.4 Research gaps**

Although prior reviews have pointed to 5-7 years of L2 exposure for bilingual learners to obtain language levels comparable to those of monolingual learners (Collier, 1989; Jim Cummins, 1984, 2017; Hakuta et al., 2000; Saunders & O'Brien, 2006), most studies comparing the language levels of monolingual and bilingual learners have not used latent invariant proven variables. The same methodological challenges exist for studies examining the predictive pattern of linguistic skills for reading comprehension across bilingual and monolingual learners. Since most prior studies have used manifest variables, most findings discussing whether the predictive pattern of linguistic skills to reading comprehension is equal across language groups as well as the unique contribution of linguistic aspects to reading comprehension are questionable (Kieffer et al., 2016). To what extent the results of prior studies are hampered by the inclusion of measurement errors or

assessments by invariant measures is unknown. There is a need for more examination of this subject to determine to what extent prior research could be duplicated by more refined statistical methods. There is a need for new studies examining language levels and predictive linguistic skills for reading comprehension across language groups using latent invariant proven variables. If the additional aim for studies of prediction of reading comprehension is to explore the unique contributions of linguistic skills to reading comprehension, dimensionalities of the linguistic variables need to be addressed. SEM analysis examining the relationship between linguistic skills and reading comprehension in monolingual learners provides some support for different linguistic aspects explaining more than one core underlying language ability (Kieffer et al., 2016), yet this issue has not been examined by a multi-model approach across monolingual and bilingual readers.

Furthermore, there is also a need for studies focusing on bilingual pre-adolescent subgroups other than *minority language learners*. Future studies should examine to what extent language levels and the relationship between language and reading comprehension of *early bilingual learners* are more similar to those observed in monolingual learners or *minority language learners*. Since studies of high-SES bilingual learners are rare and high SES is associated with reduced differences in language levels (Oller et al., 2011), future studies of high-SES pre-adolescents are of special importance.

## **2.6 Executive functions and the alleged bilingual advantage**

Despite a large amount of research on executive functions, agreement on a definition and an agreed-upon form of measurement of executive functions are still lacking (Baggetta & Alexander, 2016). This thesis understands executive functions in line with Diamond (2013), Blair (2016) Best and Miller (2010) and Zelazo, Blair, & Willoughby (2016), where the term is considered an umbrella term for the cognitive

processes that underlie goal-directed control over thoughts, behaviour and emotions.

EF consists of a set of neurocognitive skills involved in general top-down control processes such as switching, monitoring, inhibition, working memory, planning, problem-solving attention and reasoning (Diamond, 2013; Garon, Bryson, & Smith, 2008; Mezzacappa, 2004; Miyake et al., 2000). These abilities have shown strong relations to and predictive power for academic learning (Bull, Espy, & Wiebe, 2008; Duncan et al., 2007; LeFevre et al., 2013; St Clair-Thompson & Gathercole, 2006). One explanation for this is that the skills needed when one concentrates to avoid acting on impulse or behaving automatically (Diamond, 2016) are core components of self-control and self-regulation abilities with implications for everyday life (Mischel et al., 2011). How these skills are organized is, however, uncertain. Confirmatory factor analysis of EF shows conflicting results regarding the structure of the concept, ranging from uni-dimensional factor (Brydges, Reid, Fox, & Anderson, 2012; Wiebe, Espy, & Charak, 2008) to multi-factor models (Fournier-Vicente, Larigauderie, & Gaonac'h, 2008). However, a recent meta-analysis gave reason to believe that the structure of these abilities changes over time, from a uni-dimensional or a two-factor model in preschool age to a three-factor model in school age yet a three-factor or nested model in adolescence and adulthood (Karr et al., 2018). However, the authors also find evidence for publication bias towards well-fitted but potentially not replicated models with underpowered samples.

### *2.6.1 Different cognitive domains in executive functions*

Even if it is still empirically uncertain how the cognitive domains of EF are linked together, they do theoretically differ from one another to some extent. A review of the theory of EF identifies a disagreement over how many domains EF actually consists of as well as whether one and the same domain is regarded as a higher-order EF skill or a core EF skill

(Baggetta & Alexander, 2016). The review identified 39 different constructs labelled EF domains by researchers, yet 38% of these domains were referred to only once. The most frequently mentioned domain was inhibition, followed by working memory, switching, planning and attention. The description of EF domains in this over-binding is limited to the ones Baggetta & Alexander find to be most frequently mentioned, supplemented with the EF domain of monitoring. There seems, however, to be a disagreement on which processes are involved in the different cognitive domains and how the domains are separable from one another. Some of this disagreement is explained below.

In Baddely and Hitch's (1986) model of working memory, working memory and short-term memory are defined as different constructs. Short-term memory involves only repetition of a string of information in the exact order it has been presented without any manipulation, whereas working memory is the ability to hold information in one's mind while manipulating it. The results of factor analysis support the separation of working memory and short-term memory by displaying how they cluster onto separate factors (Alloway, Gathercole, Willis, & Adams, 2004). However, some researchers clearly distinguish the differences between these constructs (Diamond, 2013), while others suggest that the correlations between the two constructs are large and even overlapping, questioning whether short-term memory and working memory are indeed separable constructs (Aben, Stapert, & Blokland, 2012).

Switching is often used interchangeably with the term shifting as a measure of cognitive flexibility and involves the ability to flexibly switch between tasks or mental sets or goals (Best, Miller, & Naglieri, 2011; Friedman et al., 2006), where each set might contain more than one rule (Diamond, 2013). It has also been described as flexible attentional shifting towards goal-relevant representations (Duggan & Garcia-Barrera, 2015). Some have theorized that the ability builds on inhibition and working memory (Diamond, 2013), but it is still regarded as a separate domain (Diamond, 2013; Donnelly, 2016; Karr et al., 2018;

Miyake et al., 2000). Skills in switching emerge around school age and develop through the ages of adolescence (Karr et al., 2018). The ability to flexibly switch between tasks requires the ability to inhibit previously relevant stimuli to focus solely on the new and relevant information for the new task to be solved (Diamond, 2013). Flexibility in switching is often measured as the heightened response time required to switch and respond to the new dimension or the accuracy of correct responses in trials after a switch of the dimension to respond to has occurred (Diamond, 2013; Donnelly, 2016).

Inhibition is the ability to control one's attention, emotions, thoughts and behaviour in order to complete a task without giving in to impulses (Diamond, 2013). It seems to be agreed that different forms of inhibition exist, but various researchers define the different forms of inhibition somewhat differently. An often referred to division is to split inhibition into hot and cold components. Hot and cold inhibition are both goal-directed processes, but hot inhibition is a goal-directed process elicited in contexts that include tension between immediate gratification and long-term reward (Zelazo, Qu, & Müller, 2005). In Diamond's (2013) broad definition of inhibition, the hot components of inhibition can be described as inhibition of feelings. The handling of delayed reward and regulation of one's own social behaviour are regarded as hot components of inhibition (Grafman & Litvan, 1999; Rolls, 1995). Cold components of inhibition are more mechanical or logical without involving emotions (Grafman & Litvan, 1999; Zelazo et al., 2016) and involve inhibition of an automatic or proponent response to a neutral stimulus.

Some researchers emphasize the difference between these abilities (Diamond, 2013), while others argue that both hot and cold EF processes contribute to decision-making (Zelazo, Müller, Frye, & Marcovitch, 2003) and that it is therefore likely impossible to design a task of pure measurement of cold or hot decision-making (Séguin, Arseneault, & Tremblay, 2007).

According to Diamond (2013), there is a third component of inhibition – inhibition of attention, which concerns interference control at the level of perception. Inhibition of attention enables us to selectively focus on what we choose and suppress attention to other stimuli. This is sometimes referred to as selective attention or attention control and is most often measured by the Flanker test (Diamond, 2013). Selective attention is, however, defined differently by other researchers. Sarter, Givens, and Bruno (2001) argue that it is a fundamental component of attention characterized by the subject's readiness to detect rarely and unpredictably occurring signals over prolonged periods of time. Therefore, selectivity is seen not as equivalent to inhibition but as a construct of attention.

Attention as an overall construct is seen as a complex function that contributes to achieving and maintaining a state of alertness, orienting towards and selecting sensory input for preferred processing, and regulating thoughts and responses in a goal-directed effortful mode (Commodari, 2017; Petersen & Posner, 2012; Pozuelos, Paz-Alonso, Castillo, Fuentes, & Rueda, 2014). Attention also plays a role in the self-regulation of behaviour and emotions (Blair & Raver, 2015; Posner & Rothbart, 2000). The attention system can be divided into subsystems that perform different but interrelated functions (Petersen & Posner, 2012; Posner & Petersen, 1990). It depends on three independent networks, the orienting, the alerting and the executive network (Petersen & Posner, 2012), all measured by the ANT test (Fan, McCandliss, Sommer, Raz, & Posner, 2002).

Alerting attention is the ability to achieve and maintain alertness over long periods (Posner, 2012) and is related to sustained vigilance. In addition to the alertness index on the ANT test, another method of measuring sustained vigilance is measuring performance on a long and usually rather boring task (Petersen & Posner, 2012). Orienting attention is the ability to prioritize sensory input by selecting a modality or location (Fan et al., 2002) and refers to the capacity to change focus from



one stimulus to another and to shift and re-engage the focus of attention in response to different stimuli (Mezzacappa, 2004). Executive attention is the ability to respond to the interference of competing demands. It comprises processes involved in the execution of goal-directed behaviours, including anticipating consequences, selecting among competing demands and responses, planning actions, initiating and maintaining purposeful behaviour, etc. (Mezzacappa, 2004).

The three attentional networks, orienting, alerting and executive attention, form the basis upon which different key aspects of attention rely (Posner, Petersen, Fox, & Raichle, 1988; Pozuelos et al., 2014). Examples of such aspects are distributed attention, selective attention, focused attention and alternating attention.

However, as mentioned before, the line that different researchers draw between inhibition and attention is fuzzy at best, and the core disagreement seems to be related to the executive network of attention. It is the conflict index on the ANT test that Diamond (2013) argues is tapping inhibition of attention. It has also been argued that executive control of attention involves resolving conflict among responses on a broader level, for instance, the ability to achieve control over one's behaviour (Fan et al., 2002; Jurado & Rosselli, 2007; Posner & Rothbart, 2000). This becomes evident when measures traditionally regarded as measures of inhibition of automatic or proponent behaviour response to stimuli (Diamond, 2013; Donnelly, 2016; Lehtonen et al., 2018) are labelled examples of executive control of attention (Fan et al., 2002; Jurado & Rosselli, 2007; Posner & Rothbart, 2000). The active process involved when performing is then understood as the process of directing attention to the task demands rather than inhibition of a proponent behavioural response to a natural stimulus.

Bialystok (2017) has an even wider understanding of the term executive attention. She builds her conclusions on identifying similarities between the work done by Engle on working memory, Posener and colleagues on

executive attention networks and Botvinick, Braver, Barch, Carter and Cohen (2001) on monitoring. Engle (2002) and Engle and Kane (2004) expand the Baddeley and Hitch model of working memory by claiming that the capacity referred to in the working memory model is not a storage place but rather should be regarded as the extent to which resources are available to control attention to maintain the information relevant to a current task. According to this view, working memory is a cognitive system in which memory and attention interact to produce complex cognition. In Bialystok's (2017) understanding of the subcomponents of Posner and colleagues' attention network model, the executive attention subcomponent includes functions such as working memory, switching and inhibition. From this, she draws parallels to Botvinick, Braver, Barch, Carter, and Cohen (2001), which explains monitoring as a cognitive domain as an ability to monitor attention to conflict detection. By regarding the role of attention as the main factor in all of these abilities, she suggests that working memory, monitoring, switching and inhibition are all different expressions for the same underlying ability, which she labels executive attention.

Monitoring as a cognitive domain is included due to its place within the bilingual advantage debate (Hilchey & Klein, 2011) but is not commonly referred to as a domain within the EF field (Karr et al., 2018). Monitoring is related to inhibition of conflict; however, there is no consensus about how it is related to inhibition. Some researchers treat it as a separate domain (Hilchey & Klein, 2011; Hilchey, Saint- Aubin, & Klein, 2015; Lehtonen et al., 2018), while others view it as part of inhibition (Donnelly, 2016).

Inhibition is often measured by the reaction time of incongruent trials, and monitoring is measured as the reaction time on congruent trials across blocks of both congruent and incongruent trials (Lehtonen et al., 2018). Incongruent trials are trials that require inhibition of a natural response, thereby forcing one to deal with conflict. Congruent trials are tasks that are identical to incongruent trials yet without the conflict

(Donnelly, 2016). It is well known that tasks that require inhibition are more time-consuming than similar tasks without conflicting information that require inhibition of a natural stimulus (de Bruin & Sala, 2018). The extra reaction time is explained as inhibition costs (Donnelly, 2016). However, in mixed blocks of both congruent and incongruent trials, the reaction time to blocks that follow incongruent trials is also found to be more time consuming than the response time to a congruent trial that follows a long line of congruent trials (Hilchey & Klein, 2011). It is theorized that the reason is that it takes time to adjust the level of attention from the high level necessary to handle conflicting tasks to a level suitable for dealing with easier tasks consisting of congruent stimuli (Botvinick et al., 2001; Hilchey & Klein, 2011). How well one handles the adjustment of attention levels is described as monitoring and is generally assumed to reflect the efficiency of processing in a high-conflict environment (Donnelly, Brooks, & Homer, 2019).

Planning refers to the ability to identify and organize a sequence of steps necessary to solve a problem or accomplish a goal (Lezak, Howieson, Loring, & Fischer, 2004). Good planning is associated with the capacity to consider alternative approaches and then choose the most efficient approach (Stuss & Alexander, 2000). It is considered to be a higher-order EF (Diamond, 2013).

### *2.6.2 Measures of different cognitive domains*

EF is not just difficult to define as an overall construct and at the domain level; it is also difficult to measure. Miyake and Friedman (2012) argue that the main problem is task impurity. The reason is that many of the tasks thought to measure EF have a multi-dimensional structure, with several cognitive abilities being simultaneously active in the same performance (Duggan & Garcia-Barrera, 2015). This means that the same test often taps different cognitive domains. Furthermore, since all EF tests are embedded in a certain context, any variance attributed to non-EF processes associated with the specific task context is

automatically included in the measurement outcome (Miyake & Friedman, 2012). This makes it difficult to cleanly measure the EF component of interest.

The Trail Making Test (Armitage, 1946) is a good example. This test is often used as a measure of switching abilities (Bialystok, 2010; Bialystok & Viswanathan, 2009; Raschke, 2013; Stephens, 2013) and provides 3 outcome measures: response time on a non-switching part of the test (Part A), response time on a switching part of the test (Part B), and response time for the switching costs (Part B- Part A). The different outcome measures are uniquely predicted by different EF and non-EF abilities. For instance, part B of the test is mostly a measure of working memory, and the contribution of switching is marginal. However, for the outcome measure switching costs, where the response time on Part B of the test is corrected for the response time of the non-switching part of the test, the contributions of visuo-perceptual and working memory abilities are minimized. This provides a purer measure of EF, explained primarily by switching abilities and secondarily by working memory abilities (Sánchez-Cubillo et al., 2009). The example of the Trail Making Test is not unique.

To make this even more complicated, several tests have not been empirically investigated thoroughly enough to be sure of what they actually do measure. Moreover, the rationale for test validity is in some cases based mostly on theoretical assumptions or face validity due to similar test structures (Baggetta & Alexander, 2016; Paap & Sawi, 2014). To complicate it further, a large quantity of research has developed tests to assess EF that have been used only once for one specific study (Baggetta & Alexander, 2016). In addition, one and the same test, for instance, the Stroop test, is considered by different researchers to measure different aspects of attention (Commodari, 2017; Fan et al., 2002) and measures of inhibition (Lehtonen et al., 2018).

Empirical findings do not provide clear support for the leading theories of which tests actually measure different cognitive domains. For instance, different tests thought to measure the same cognitive domain have been found to have low correlation or even no correlation (Paap & Greenberg, 2013). This holds true for several of the cognitive domains. Fan et al. (2002) find that the ANT alerting index is uncorrelated with ANT orienting and ANT executive function, yet all indexes are theorized to measure attention skills. Similar findings are evident regarding the term inhibition, where the interference costs of Simon are uncorrelated with the Flanker test (Paap & Greenberg, 2013; Stins, Polderman, Boomsma, & de Geus, 2005) so are the interference costs of the Flanker and Stroop test (Stins et al., 2005). This makes it natural to question whether measures theorized to measure the same overall construct do indeed share the same underlying ability. The confusion creates chaos that is hard to navigate within. The chaos and uncertainty are related to both test validity and domain validity. Some researchers addressing this topic go so far as to say that there are no “pure” measures of any specific EF skill (Zelazo et al., 2016)

### *2.6.3 The alleged bilingual advantage in EF*

Bilingual learners’ language processing differs from that of monolingual learners. One reason is the activation of both bilingual learners’ languages in conversation regardless of whether the dialog between interlocutors occurs in only one language (Kroll, 2008). This requires the bilingual learner to focus on the activated language while ignoring the stimuli for the other language. For this activation to proceed fluently, the brain must possess an effective selection system (Bialystok, 2017). The selection system is claimed to be based on a domain-general system involving both language selection and non-verbal processing. Based on this assumption, bilingualism is theorised to work as an exercise of the bilingual brain and to cause enhanced EF skills (Bialystok, 2017; Hilchey & Klein, 2011). Multiple suggestions exist for the precise nature

of the relationship between bilingual learners and enhanced EF skills, which gives grounds for believing that certain EF domains might be more positively affected by bilingualism than others. One suggestion is that bilingual switching between bilingual learners' two languages exercise their ability to flexible switching in general, hence switching of non-linguist stimuli, should also be affected. This leads to a bilingual advantage in switching (Donnelly, 2016). Another suggestion is that bilingual learners inhibit the non-selected language (Bialystok, Craik, Green, & Gollan, 2009), thereby enhancing the inhibition skills of non-linguistic stimuli. A third explanation is that bilingual learners, due to their bilingual language processing, need to constantly monitor their environment for conflicting information; hence, monitoring should be enhanced (Hilchey & Klein, 2011). The newest explanation is that bilingual learners need to direct their attention to the language activated in the dialog, thereby enhancing attention skills (Bialystok, 2018).

Since enhanced EF skills are a result of bilingual learners' experiences with language processing, the alleged advantage is theorized to be larger for bilingual groups with more experience of bilingual language processing. Suggested bilingual experiences that should influence the size of the bilingual advantage are bilingual learners' AoA, L2 proficiency, SES level, age and degree of balanced bilingualism (Bialystok, 2017, 2018; Naeem, Filippi, Periche-Tomas, Papageorgiou, & Bright, 2018; Pelham & Abrams, 2014).

According to theory, these bilingual groups should exhibit a larger advantage in EF:

1. Bilingual learners with low AoA over bilingual learners with high AoA.
2. Older bilingual children over younger bilingual children.
3. Balanced bilingual learners over unbalanced bilingual learners.
4. Proficient L2 speakers over non-proficient L2 speakers

5. Conflicting hypotheses for the connection of SES and EF. One theory suggests that children with low SES have a larger advantage. An alternative suggestion mentioned in the study of De Cat, Gusnanto, and Serratrice (2018) is that SES level could function as a threshold; thus, a certain level of SES must be obtained before the bilingual advantage occurs.

See article 1 for an extended description of the theories concerning the relationship between bilingualism and the different EF domains as well as a more in-depth description of the theoretical background for why some bilingual subgroups allegedly have a greater advantage in EF than other bilingual subgroups.

Several meta-analyses have previously investigated the theory of a bilingual advantage in EF, displaying different results. Adesope et al. (2010) find an overall advantage in favour of bilingual learners ( $d = .41$ ). The bilingual advantage at the domain level ranges from  $d = .26$  for problem-solving to  $d = .96$  for attention control in favour of bilingual learners. de Bruin, Treccani, and Della Sala (2015) identify a mid-range bilingual advantage ( $d = .3$ ) but also detect signs of publication bias and question whether the alleged theory of a bilingual advantage is a case of publication bias. Donnelly (2016) finds a small to moderate effect for inhibition costs and no significant effect for switching costs. Grundy and Timmer (2016) find a small to medium effect in favour of bilingual learners in working memory ( $d = .20$ ), while the meta-analysis of Lehtonen et al. (2018) finds a small advantage in overall EF ( $d = .01$ ) and in the domains of inhibition ( $d = .1$ ), switching ( $d = .15$ ) and working memory ( $d = .07$ ). However, this advantage disappeared when controlling for the small study effect.

#### **2.6.4 Research gaps**

There is a need for more consensus on what EF as an overall construct actually is, as well as an agreement on how the different EF domains are

separated from one another. This needs to be grounded in empirical evidence to a greater degree than is the case now. There is also a need for empirical studies of different EF tests in order to address the issue of task impurity as well as to increase the reliability of the different tests.

Regarding the bilingual advantage hypothesis, there is a need to summarize the studies examining this theory in children. Possible advantages in overall EF and the full range of domain levels should then be explored. Furthermore, to place the theory under scrutiny, the relationship between the differences in EF and different dimensions of bilingual experience should be explored. There is also a need to examine whether a potential bilingual advantage in children is at the construct or task level.

## **2.7 Research questions addressed in the different articles in this thesis**

The research questions in the three papers are as follows:

- I) The research questions in paper 1:
  1. To what extent is there a bilingual advantage in overall EF?
  2. To what extent are group differences in bilingual and monolingual children in overall EF related to moderators concerning sample characteristics—such as degree of balanced bilingualism, level of L2 skills, socioeconomic status, nonverbal IQ, age, AoA, origin (of the sample) as well as moderators—related to methodology, such as sample size, publication status, publication year and lab?
  3. How do the group differences between bilingual and monolingual children vary across specific EF components, such as inhibition (hot; rewarding stimuli/ cold; neutral stimuli), attention, switching, monitoring, working memory and planning?



4. To what extent is a bilingual advantage in inhibition, attention, switching, monitoring, planning and working memory related to moderators that are significantly associated with overall EF and to the task used to measure the different EF components?
5. To what extent are the results related to small study effect and publication bias?

II) The research questions in paper 2:

1. What are the differences between monolingual toddlers' language comprehension and dual language toddlers' comprehension in the majority language when they are 2 years and 9 months?
2. What is the relationship between dual language toddlers' parental second language input at home and the toddlers Language comprehension in Norwegian?

III) The research questions in paper 3:

- 1) To what extent do bilingual 5<sup>th</sup> graders with an AoA of the instructional language from birth to 2 years old have levels of language and reading comprehension skills similar to those of their monolingual peer across different aspects of language and reading?
- 2) Are the patterns in which aspects of language comprehension, decoding skills and SES that predict reading comprehension the same for bilingual and monolingual children?



### **3 Methodological perspectives and considerations**

The three studies in this thesis all have a quantitative methodological approach but differ in the analytical approaches and statistical software programs used. To support the readers' understanding of the foundation of the findings of this thesis, the extended abstract provides a short description of the methodological choices and approaches used in the different articles. The more fine-grained methodological considerations presented in this chapter reflect methodological considerations not extensively covered in the three articles.

Study 1 is a meta-analysis. I was responsible for the literature search, coding the data, calculating the inter-rater reliability in the coding and study extractions, and conducting the analysis in CMA. The analysis in R was conducted in collaboration with two of the other authors. In study 2, all data was collected before my entry to the project. My individual contribution to this study was the project idea, the theoretical and empirical framework of the article as well as performance of the analysis. In Study 3, I planned the project, adjusted and piloted the group- tests, and was involved in collecting, scoring and punching the data. I was also responsible for training research assistants and teacher how to conduct the test assessments to ensure good data quality, as well as the analysis done in M plus.

Studies 2 and 3 share partly the same sample; however, more knowledge of the sample is available in the last study. There is therefore a need in this chapter to clarify the differences and similarities between the samples in Studies 2 and 3. Since the measures in Study 2 used to examine language comprehension originally were designed for purposes other than research, TRAS as a tool in research is also addressed here. Article 1 has a comprehensive and transparent methods section and is therefore less covered in the extended abstract.

### **3.1 Study 1**

#### *3.1.1 Preregistration/data collection*

The metaanalysis was preregistered in Prospero and can be located here: [https://www.crd.york.ac.uk/prospero/display\\_record.php?RecordID=75281](https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=75281). Study 1 contains a description of most deviations from the registration. The literature search involved key words for bilingualism (“bilingual”, “second language learner” and “dual language learner”) combined with terms related to EF (“inhibition”, “attention”, “working memory”, “switching” and “executive function”). Articles published within the timeframe of 1980 to December 2017 were searched for in the databases Eric, Medline, ProQuest dissertations, PsychInfo and Web of Science. Studies were included if they reported measures of EF skills for bilingual learners and a monolingual control group with participants aged 0-18 years. See Figure 1, article 1 for a flow diagram of the search and description of the further process in extracting the data.

It should be added that one of the anonymous reviewers questioned why the EF domains of monitoring and planning were not added as search terms in the literature search. The reviewer pointed out that this may have led to missing studies on these domains. To compensate for this, an additional search was performed using the search words “bilingual” and “benefit” or “advantage” crossed with “executive function” and “planning” or “monitoring” in the databases Web of Science, Eric and Medline. Once the first search was performed, it was evident that the search words “planning” and “monitoring” generated very different papers than the search words used in the submitted version of the paper. In this second search, most of the identified articles were related to the medical field. To reduce the number of extracted papers, new exclusion terms were added to the search. Further specifications were added to refrain from extracting studies on cancer, sickness, alcohol, drugs, medicine, and medical health. The exclusion of participants with learning disabilities from the original search was retained. The new

search yielded 450 articles across the different databases. These were assessed according to the inclusion and exclusion criteria for Study 1, resulting in the identification of no new articles.

Twenty percent of the first round of the full search was randomly selected for data extraction by a second author. The inter-rater agreement of the authors was  $\alpha = 0.891$ . To ensure coding reliability, 20% of the dataset was double-coded. The inter-coder correlation (Pearson's) for the main constructs was 0.993. In the first round of the revision, all the extracted articles were checked again to ensure that all EF measures were coded, followed by a control of coding. At this point, 100% of the data for the main constructs were double-checked to ensure coding reliability. In all, three authors were involved in quality checking the dataset.

### **3.1.2 Analysis**

The data were coded and analysed first in the comprehensive meta-analysis software (Borenstein, Rothstein, & Cohen, 2005) and then in Rubometa for R (Fisher & Tipton, 2015; Tipton & Pustejovsky, 2015). Study 1 presents only the results from the analysis in R, where dependency in the data structure was controlled by using robust variance estimates (Tanner-Smith, Tipton, & Polanin, 2016). First, samples of all control groups with multiple bilingual comparison groups were coded in such a way that sample size for the monolingual group was divided across the number of bilingual comparison groups. On the request of an anonymous reviewer, this was changed to include all monolingual children in every comparison with the bilingual samples (RVE handles this kind of dependency). However, the overall results from the analysis in CMA and the two different datasets analysed by RVE are very similar to the published version of the article. The same can be said for the corrections for the small study effect, indicating that the results remain stable across different analytical approaches. There were, however, some differences across moderator analysis, where fewer of the examined

moderators were significantly related to overall means of EF using robust variance estimates.

With CMA, non-verbal IQ was a significant moderator of overall EF, and TASK was a significant moderator for the domain of switching ( $R=0.38$ ). The differences in moderator analysis could have two causes. First, CMA handles dependency in the data structure more roughly than the RVE statistic. Basically, in CMA, multiple effect sizes must either be aggregated to the study level or one effect size must be selected over another for use in the analysis. This could have resulted in a lack of information in the CMA analysis to fully investigate the true relationship between the overall outcome of the analysis and the moderator. The other reason for these differences could be that the degrees of freedom in RVE are adjusted to handle small sample sizes (Tanner-Smith et al., 2016). In RVE, the degrees of freedom depend both on the number of studies and on the features of the covariate. Basically, with a very skewed distribution (most values within a close range yet the presence of a value that deviates from the others) or an imbalanced number of studies in the different categories (e.g., 5 in one category and 25 in another), degrees of freedom are reduced. As a result, the power of some moderator analyses in RVE is surprisingly low. There was an unbalanced number of studies in the different categories of the moderator task.

The other deviation from the results presented in article 1 is related to the control for the small study effect on the domain of switching. These results “jumped” slightly back and forth across the different sets being analysed. Switching skills were equal across language groups when controlling for the small study effect in the analysis of the first dataset in R. This dataset contained smaller sample sizes for some of the monolingual control groups. However, the analysis of the small study effect in the final dataset, as well as the analysis in CMA using trim and fill adjustments, detected a bilingual advantage in switching. These inconsistencies across analyses suggest that these results are less robust

than the rest of the results in this article. Besides these analyses, provided the three different procedures of data analyses the same outcomes.

## **3.2 Study 2**

### *3.2.1 The Stavanger Project and participants for article 2*

Study 2 of the thesis investigates bilingual and monolingual children's language comprehension in Norwegian at age 2 years and 9 months. This study includes children in The Stavanger Project born in 2005, 2006 and 2007.

All ECEC institutions in Stavanger municipality were invited to participate in The Stavanger Project. All ECEC institutions owned by the municipality and approximately 50% of those privately owned accepted the invitation. Parents of children in these ECEC institutions born between 01.07.2005 and 12.12.2007 were asked for written consent to participate in the study. Beyond the birthdate, no other criteria excluded a child's participation in the study.

The sample of Study 2 consists of 902 monolingual learners and 161 bilingual/*dual language learners*, hereafter called *early bilingual learners*. The parents of the *early bilingual learners* completed a questionnaire regarding which languages family members used when communicating with one another. Of the *early bilingual learners*, 30 of the children were exposed mainly to Norwegian at home, while 84 were exposed to both their L1 and Norwegian at home. Forty-nine bilingual children interacted with their parents mainly in L1. The monolingual and bilingual children had the same median SES background measured by parental education. The median SES was medium to high, possibly because of the oil industry in Stavanger recruiting highly educated workers from abroad.

*Methodological perspectives and considerations*

---

We knew which languages the parents and toddlers used in the bilingual homes, yet the parents' nationality was unknown. However, additional information about nationality is collected in Study 3. This information is merged with information obtained in Study 2 and presented in Table 1 below.

Table 1. The number of children with one native and one non-native Norwegian-speaking parents identified through parent questionnaire in pre-adolescent age represented in the different groups of *early bilingual learners* in toddlerhood.

Language interaction at home during toddlerhood	Missing (N= 74)	One native Norwegian-speaking parent (N= 60)	One non-native Norwegian-speaking parent (N= 31)
Mainly in Norwegian (N= 30).	(N =17)	13	0
A mix of Norwegian and L1 (N= 82).	(N =35)	43	4
Mainly L1 (N= 49).	(N= 22)	0	27
No information on language interaction during toddlerhood.		4	0

*Note.* L1 = a minority language. The 4 children with a native Norwegian-speaking parent who neglected to return the parent questionnaire in toddlerhood were included only in Study 3 and not in Study 2.

Since the samples only partly overlap, such merging produces many missing data. This brings much uncertainty to the conclusions that can be drawn based on the merged dataset. There seems, however, to be a general trend that children with one native-speaking parent are represented mainly in the *early bilingual* subgroup that interacts in both languages at home, while only children with two non-native Norwegian-



speaking parents are represented in the subgroup that interacts mainly in L1. Note, however, that the results represent a trend that could perhaps be changed if the parents in Study 2 who were not a part of Study 3 (N=74) had also contributed information about their nationality.

### **3.2.2 Selection of measures**

At the first timepoint in The Stavanger Project were data of the children's language skills mapped by the observational form TRAS. The TRAS consists of eight 9-item subscales measuring theoretically different language aspects. The form consists of one circle with three age levels representing age-related language skills: one level for 2- to 3-year-olds, one for 3- to 4-year-olds, and the last for 4- to 5-year-olds. Children's mastery of all age-related language skills was observed. See article 2 for more information on the usage and scoring of TRAS.

In the analysis of differences in bilingual and monolingual learners' language skills, only the TRAS aspect of language comprehension was chosen. The selection of this aspect was based on theoretical assumptions.

### **3.2.3 TRAS as a tool for research**

The original purpose of TRAS was not to develop an assessment tool for research. The developers' aims were twofold: first, to raise ECEC employees' awareness and knowledge of children's language development and second, to help ECEC employees identify children with poor language skills. These aims guided the selection of the assessment items. In the selection of which items to pilot, multiple items were chosen. The researchers were guided by knowledge of which items that were theoretically thought to tap into the 8 linguistic subskills assessed in TRAS and clinical knowledge of which items prior research had found to predict later language and literacy difficulties. The items were tested and selected in order to identify the children with the lowest

levels of language skills, representing .20 percent or less (Stangeland, 2018).

The statistical development of TRAS was built on classic test theory (CTT). CTT comprises a set of indicators to assess how well a proxy (the manifest variable) observes the unobserved variable (Brennan, 2010; DeVellis, 2006). These indicators are built on an understanding that in using any measurement tool, either a true score or an error term will be identified. The error represents the amount of error associated with a particular item and is assumed to be influenced by factors other than the true score. Individual error terms are also assumed to be uncorrelated to error terms in the same assessment (DeVellis, 2006). Since the true score varies across individuals and timepoints, an observed score should mirror this variation. However, since the true score is unobserved, it is not possible to identify how well the observed score and the true score covary. As a way of handling this, CTT assumes all items to be strictly parallel and uses the relationship between observed manifest items as a proxy for the unobserved variable (the true score) they share in common (DeVellis, 2006). Basically, this means that the correlations between observed variables are interpreted as the correlation between the true variable, and the correlation scores can be used as estimates of reliability of the items themselves. CCT extends the item reliability to scale (a uni-dimensional set of items) reliability, often measured by Cronbach's alpha. Cronbach's alpha is influenced by two factors: the correlations among items and the number of items. Cronbach's alpha is a measure of the proportion of variance in a scale that could be attributed to the common influence on the scores of the individual items (DeVellis, 2006).

The research group that developed TRAS did not calculate Cronbach's alpha, but according to Stangeland (2018), the reliability of TRAS overall was .9, with Cronbach's alpha ranging from .8 to .6 for the different sections of the test. TRAS consists of 8 linguistic sections. Each section is assessed by 9 items, 3 per age group, meaning that the assessment material examines each section by asking 9 questions. A

Cronbach's alpha of .7 is considered an acceptable reliability level (Nunnally, 1978); however, .8 or higher is preferable when a test is used in research (Gall, Borg, & Gall, 1996). Note, however, that Cronbach's alpha increases with test length (P. Kline, 2000) and that the Cronbach's alpha range from 0.6 to 0.8 for the different sections is based on internal reliability calculations across 9 items.

Another way of assessing internal consistency reliability is by calculating inter-rater reliability between the scores of different test administrators. This approach was chosen by the research group that developed TRAS. Pairs of ECEC teachers were asked to observe the same child and fill in the TRAS observations by judging to what degree the child mastered an ability. Based on this information, the inter-rater reliability between the two test administrators was judged.

In the case of inter-rater reliability, the indicators are correlations between human observations, not between the test items themselves. The interpretation, however, is the same- The inter-rater reliability reflects the proportion of variance between the two scores observed by the measurement tool rather than representing characteristics of the test administrators. The researchers calculated the reliability between items within each age group, resulting in a Spearman's rho ranging from .7 to .81 (Espenakk et al., 2003). Note, however, that the ECEC teachers registered the children's level of mastery without the support provided in the TRAS Handbook (Stangeland, 2018). The ECEC teachers observing the language skills of children participating in The Stavanger Project were provided with the handbook and attended a course in how to use the tool. In addition to the TRAS handbook, the ECEC teachers received a user manual developed for use in the project containing scoring examples and precision levels (Helvig & Løge, 2006). All children were independently observed by two employees. If the employees disagreed regarding the scoring of a child, new observations were warranted until agreement was achieved. Hence, the inter-rater

reliability of the observations conducted as part of The Stavanger Project is likely to be higher.

Confirmatory factor analysis of TRAS identified 4 factors when children were assessed by 33 months of age, largely resembling the different age groups represented in the assessment tool. Items from the group 2-3 years of age were represented by factors three and four, factor one represented 3-4-year-olds, and factor two included only items from the 4-5-year-olds (Stangeland, 2018). Combined, these factors indicated the variation in the material. A possible reason why the factors to such a great degree were coherent with the age groups could be how the assessment tool was designed and tested. Testing of the different items for one age group was piloted only in this age group rather than the whole range of ages the test was designed to assess. The selection of measures was driven by difficulty levels to identify children with lower levels of language skills. The difficulty levels across subsections were so similar that children performing well on one subsection performed equally well on another subsection (Stangeland, 2018). These results, combined with the high inter-rater reliability of TRAS (Espenakk et al., 2003), were interpreted by Stangeland as support for all subsections of TRAS observing one underlying language ability that changes in difficulty level as the child grows. Newer research finds that most linguistic subskills do indeed seem to load on one common underlying factor (Hjetland et al., 2018; Language and Reading Research Consortium, 2017). Note, however, that at the timepoint when TRAS was constructed, different linguistic theoretical dimensions were explored to a small extent (Kieffer & Lesaux, 2012) but were assumed on the basis of theoretical background to reflect different constructs. Hence, it was more common to rely on theory in test construction.

Although TRAS seems to observe one underlying ability, this does not necessarily mean that the underlying ability is language. A way of validating a test is to compare test results across other well-established and previously validated tests (DeVellis, 2006). Correlations between

TRAS and TROG-R (Bishop, 2003b) as well as BPVS (Dunn, Dunn, Whetton, & Burley, 1997) were examined, and the correlations were .259 and .229, respectively (Espenakk et al., 2011). See article 2 for more information on this validation test. The validation of TRAS has also been investigated by comparing correlations of TRAS to CCC-2, (Bishop, 2003a), a well-established test of pragmatic language abilities. TRAS has also been compared to the RI-5, a test of 5-year-olds assessing the risk index for developing dyslexia. The correlations between TRAS and CCC-2 and the RI-5 were .42 and -.46, respectively, which are somewhat higher than those for BPVS and TROG (Helland, Jones, & Helland, 2017). Overall, TRAS seems to tap into a language ability.

### **3.3 Study 3.**

#### *3.3.1 Participants and content forms*

The third article partly examines the same children as Study 2 yet now investigates the 10-year-old 5<sup>th</sup> graders' language and reading comprehension skills in Norwegian. The reduction of participants from Study 2 to Study 3 is caused partly by longitudinal dropouts due to participant relocations to other municipalities and partly by the selection of participants for Study 3/denial of participation in Study 3.

The tests collected as part of The Stavanger Project were supplemented with additional tests to examine the research questions of Study 3. This was done at a stage where only the birth classes of 2006 and 2007 were 5<sup>th</sup> graders, thereby reducing the number of participants from the original study. As part of enrolling their children in The Stavanger Project, parents agreed to their children's participation in group tests. To allow additional individual testing, content forms were sent to the bilingual children in birth class 2006. Study 3 received more funding by the time the 2007 birth class became 5<sup>th</sup> graders. In 2017, content forms were sent to all bilingual learners in The Stavanger Project in the

birth class of 2007 and most monolingual pupils. The monolingual learners were invited to participate if they attended schools with bilingual peers already enrolled in The Stavanger Project. The wording of the content form sent to the parents of the 2006 birth class therefore differs from that of the content form sent to parents of children in the 2007 birth class (the information letters is presented in appendix 5).

This resulted in a sample of 301 participants: 91 bilingual and 210 monolingual learners. Data from 14 of the monolingual participants were dropped due to inadequate data quality on one of the individual tests, resulting in a total sample of 287. Sixty of the bilingual children had one native Norwegian-speaking parent, while 31 had no native Norwegian-speaking parents. Information on the timepoint for the start of attendance in ECEC institutions was missing for 17 of the bilingual children without native Norwegian-speaking parents. We can only conclude that since they were enrolled in The Stavanger Project, they must *at least* have started as two-year-olds. See Table 2 below for information on ECEC start for the *early bilingual learners* without a native Norwegian-speaking parent.

Table 2. Timepoint for start of attendance in ECEC institutions for pre-adolescent bilingual children without a native language-speaking parent.

---

<i>Early bilingual learners</i> without a native Norwegian-speaking parent (N= 31)	N
Started attending ECEC institutions as 1-year-old	8
Started attending ECEC institutions as 2-year-old	6
Missing data for start of ECEC attendance	17

---

*Note.* ECEC = Early Childhood Education and Care

### **3.3.2 Selection of measures**

Measures used in Study 3 consisted of a combination of tests used for all the participants in The Stavanger Project and measures specially selected

for Study 3. This means that some of the measures, such as the measure for decoding skills and SES, were decided before my entry into the project, while others were selected based on prior research on language and reading development. There is a lack of standardized measures in Norwegian; thus, only two measures used in Study 3 are standardized and normed for Norwegian children: the vocabulary subtest from WISC-4 (Wechsler, 2003) and the word-chain test (Høyen & Tønnesen, 2008) (measuring decoding skills). Most of the other measures were adapted to Norwegian from English (Neale, 1997) or developed by employees at the Faculty of Educational Science, University of Oslo (Brinchmann, Hjetland, & Lyster, 2016). These measures have shown good psychometric qualities in prior studies of children at similar ages (Brinchmann et al., 2016; Hjetland et al., 2018; Lervåg, Hulme, & Melby-Lervåg, 2018). Two tests required adaptation to be used in this study. The adaptations were theoretically driven.

The morphological knowledge test has previously been used in an intervention study in Norway where signs of a ceiling effect were detected after the intervention; the children were then 4<sup>th</sup> graders (Brinchmann et al., 2016). In each item, children are presented with a sentence that includes a non-word and asked to identify the meaning of this non-word within a multiple-choice format. The non-word can be understood given knowledge of the meaning of the two morphemes combined into a new non-word. The test can be considered a test of derivational morphology (Kuo & Anderson, 2006). The test does not contain any misleading information and is therefore, according to Friesen and Bialystok's (2012) argument, low in executive functions demand. The children are, however, required to form the meaning of a non-word based on their knowledge of derivational morphology. The linguistic demands of this morphological knowledge test seem based on face validity similar to the linguistic demands of the Wug test (Berko, 1958), which means that the linguistic demands are assumed to be rather high. To obtain normally distributed data, the test was supplemented with

new sentences containing more complex morpheme combinations. Most of the new, more difficult Norwegian morphemes originated from Latin or English.

The text cohesion vocabulary test has a cloze test format supplemented with a list of 4 alternative text cohesion words (Crosson et al., 2008). The children were asked to pick the text cohesion word out of 4 that would provide a meaningful sentence. In the Norwegian adaptation of the test, all sentence and text cohesion words were first translated into Norwegian by an associate professor employed by the University of Stavanger who is a specialist in English. She also has prior experience with test development. To match the difficulty level of the English and Norwegian versions of the test, some of the original translated text cohesion vocabulary was replaced. The replacement was performed in three steps. First, the text cohesion vocabulary in each category (e.g., causal, contractive, additive, temporal and adversative) was matched to the number of items theoretically thought to measure categories of text cohesion vocabulary; in this way, the original distribution among the different categories held across the different language versions of the test. Second, the frequency level for each text cohesion vocabulary was examined in English and Norwegian. Direct matching of items to frequency levels was impossible across languages due to language differences between English and Norwegian. Frequency levels were therefore matched on the category level (e.g. contrastive text cohesion vocabulary) instead of the item level. Third, the associate professor who is a specialist in English modified the wordings of the text to ensure that the new Norwegian version was as close to the English version as possible yet in line with the requirements of good Norwegian language structure.

Both the morphological knowledge test and the text cohesion vocabulary test were piloted on 5<sup>th</sup> graders in two schools in Sandnes municipality. These two schools have large populations of bilingual pupils. In total, 310 pupils were included in the pilot study (280 monolingual/30



bilingual learners). Analysis of the data displayed normally distribution on both tests.

In addition to the tests, information on the parents of the bilingual participants in Study 3 was collected through a questionnaire (the parent questionnaire is presented in appendix 6).

### **3.3.3 Data collection**

The Stavanger Project collected some of the data used in Study 3. The word-chain test was conducted in October; SES data were collected later that autumn. The specific data collection for the group tests in Study 3 took place over 3 weeks in March. Teachers employed at the different schools assessed the participants with the group tests. To enable schools to set aside time for this task when planning activities for the upcoming year, the school administration office in Stavanger municipality informed the school administrations at the beginning of the school year about the exact weeks in March when the assessment should take place. Information on how to collect the data was provided for the teachers in a course two weeks before the testing. Each school in Stavanger sent one teacher representative for the 5<sup>th</sup> grade teachers of the 2006 and 2007 birth classes to this course. When a representative of the school was unable to attend the course, the teachers were visited at their schools and introduced to the tests individually. All group tests in Study 3 were in paper format.

Four research assistants and I conducted individual testing from May through the beginning of June. All the research assistants were retired teachers. Three of them had previously been employed as research assistants responsible for data collection for other projects at the National Reading Centre. The fourth research assistant had extensive test competence from her workplace. The research assistants attended a one-day course in how to conduct and score the three tests. In the course, administration of the tests, start and stop criteria of the different tests and

scoring of items were described in detail and discussed through practical examples. The research assistants were provided with one booklet per child containing instructions for the tests themselves as well as a pre-made reporting system for each test. On the front page of each test, start and stop instructions for that specific test were highlighted. The sample participants were randomly assigned at the school level to different research assistants.

### *3.3.4 Missing data*

Problems related to missing data affected the two group assignment tests, the Nara reading comprehension and listening comprehension test, and SES. These data are considered missing at random.

Test booklets measuring listening comprehension for two pupils were mixed up while anonymizing the tests. In addition, participants were missing at the test level in the group tests. The reason is that teachers' commitment to the project in the 2007 birth class varied across schools. Some schools independently scheduled new appointments for pupils who were absent on the day of testing, while others did so when they were reminded, and however small proportion of the schools did not find time to do so. Of the 287 participants, 13 did not take the text cohesion vocabulary test and 19 missed the morphological knowledge test. These data were addressed by full maximum likelihood.

The problems with missing reading comprehension data are related to two research assistants mixing up the stop criteria for the different tests and ending one test prematurely. Usually, they stopped the test after a participant failed to answer 5 consecutive questions instead of continuing until the participant failed to answer every question for two consecutive tests. A small minority of the tests were ended for no clear reason. Missing data at the item level caused by the use of inaccurate stop criteria involved a total of 124 of 310 participants (103 monolingual learners/11 bilingual learners). Participants mapped by these two research assistants

attended schools located in different areas of the municipality, ranging from high- to low-SES areas. These flaws in the dataset required a transformation of the variable in order to use the data in the desired model. To avoid a dataset with multiple imputations in the analysis conducted with Mplus, the items on Nara were re-scored, and all items after 5 consecutive failures in one and the same text were scored as 0. Participants who stopped before completing the test for no clear reason were eliminated from the sample, resulting in a total of 287 participants. Data for the new Nara variable were normally distributed for the overall sample as well as for the bilingual and monolingual subsamples (skewness: .47/.36/.59, kurtosis: .31/.59/.11). The same distribution was evident after items were dropped in order for Nara to pass the test of strict invariance (skewness; .71/.90/.64, kurtosis: .11/.86/-.09).

All parents of participants in The Stavanger Project received a questionnaire containing questions about their education level. This questionnaire was returned to the National Reading Centre before information on the extension of this project was sent out. Causes for not reporting SES data are therefore regarded as unrelated to Study 3. SES data are available for 167 of the 287 participants: 52 bilingual learners and 115 monolingual learners. In the full SEM model, data were analysed by imputations across 5 multiple datasets using R, which, in contrast to Mplus, has a package that handles imputations in combination with WLSM as an estimator. The results of the model estimation in each imputed dataset were combined using the R (Team, 2019) package *semTools* (Team, 2019). The robust Mann-Whitney U test did not lend support for any difference in SES between mono- and bilingual children ( $p = 0.26$  mother and  $p = 0.51$  father).

### **3.3.5 Analysis**

Invariance testing of each construct was conducted in Mplus (Muthén & Muthén, 1998), while the full SEM model was conducted in R (Team, 2019). These two software programs differ with regards to how to handle

missing data in small- to medium-sample SEM studies using categorical variables. In Mplus, CFA modelling with categorical variables containing a large number of items, in combination with a substantial number of items missing at random, is best handled by multiple imputations (Brown, 2015). However, the lack of a large sample size requires the use of WLSMV as an estimator in CFA with categorical factors (Brown, 2015). Unfortunately, Mplus does not handle the combination of difference testing with multiple datasets and WLSMV as an estimator (L. Muthen, personal communication, October 10, 2019); however, the package SemTools does (Team, 2019). We could have conducted the whole analysis in R yet chose to handle the different analysis sets using different software programs. The reason was twofold, based on both practical and methodological reasoning. At a practical level, I already knew Mplus, while I was unfamiliar with SemTools. On a methodological level, it was considered less risky to impute values of SES than of reading comprehension, even though a large number of imputations are involved for both variables. However, SES was used only as a control variable. Reading comprehension was, however, the dependent variable. Imputation of such a large number of scores, and just for the more difficult items of the test for reading comprehension, would have brought more uncertainty to the analysis. On the other hand, doing so could perhaps have contributed to a larger factor variance in the factor of reading comprehension.

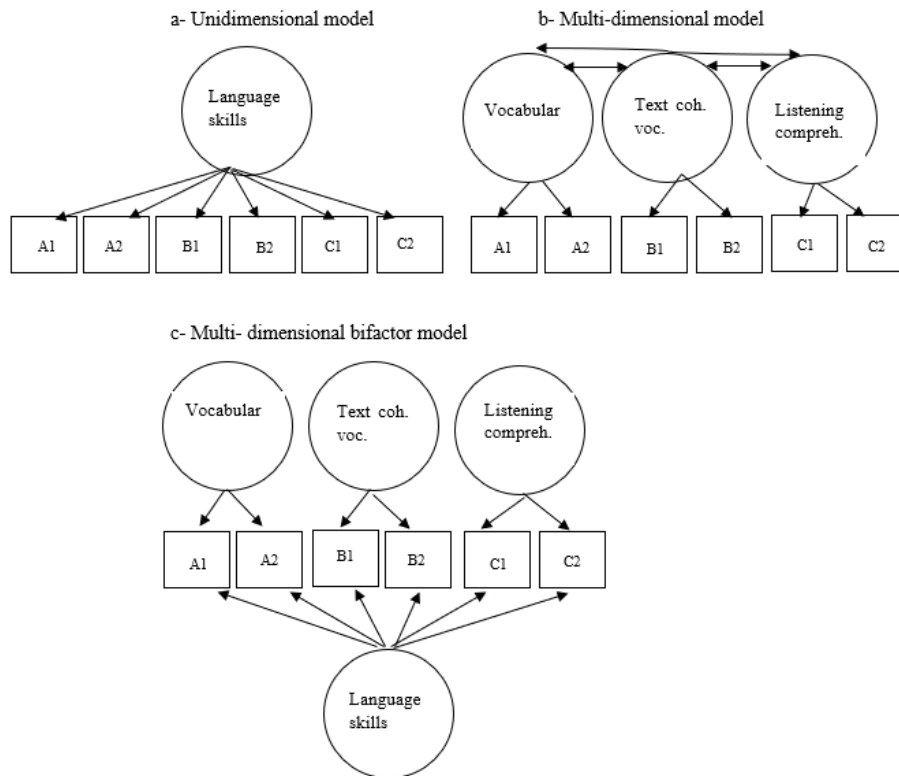
### *3.3.6 Statistical methods – choice of model*

Invariance testing of Nara was perceived on two levels. First, the extent of item bias was evaluated across the bilingual and monolingual groups by IRT using R. Thereafter, a latent variable was created based on the results of the IRT analysis. Second, the latent variable reading comprehension was tested for invariance. The reason for applying IRT analysis as a starting point was that the number of items in the test in combination with very little variance in some items and high collinearity

among other items, caused problems to nail an acceptable CFA- model. No other methodological deviations were made that are not specified in article 3.

The main purpose of examining the prediction of reading comprehension in Study 3 was to examine whether the same linguistic aspects predicted reading comprehension across the mono- and bilingual reader as well as to investigate whether the magnitude of the predictions was equal across groups. We therefore settled on the model presented in article 3. However, given a large enough dataset, it would have been interesting to additionally examine the unique prediction of specific aspects of language for reading comprehension. Such an investigation could preferably test measurement models similar to those presented in the study of Kieffer et al. (2016), thereby better accounting for the common relationship between the different linguistic constructs used in the SEM model. This would have implied testing 3 different models: first, a model where all the latent variables (text cohesion vocabulary, vocabulary and listening comprehension) loaded on one underlying common language variable; second, a correlated model similar to the one we used in article 3; and third, a bifactor model allowing text cohesion vocabulary, vocabulary, listening comprehension and a common underlying language variable to be latent factors. Figure 2, a-c illustrates a simplified version of the three measurement models. Such an approach could determine whether the models were significantly different from one another as well as identify which model had the best fit. Thus, the dimensionality of the constructs was better controlled than in Study 3, and prediction from the different latent variables displayed the unique contributions of different linguistic aspects to reading comprehension.

Figure 2. Graphic representation of the three measurement models as they could be applied at the item level.



Note. Graphical representations of the three measurement models as they can be applied to item- level data from three tasks (A= listening comprehension, B = text cohesion vocabulary, and C= vocabulary) designed to tap theoretical separate skills, including (a) unidimensional model; (b) multi-dimensional model, correlated factor model; and (c) multi- dimensional bifactor model. Number of parcels are for figure simplicity limited to only to parcels per variable.

### 3.4 Construct validity

Construct validity concerns the extent to which an instrument actually measures what it sets out to measure (Field, 2013). It concerns the validity of the inferences drawn from the indicators to a construct and thereby the operationalization of the measurement (Kleven, 2008). Handling of random measurement errors and systematic measurement

errors threatens construct validity. This threat has been handled differently in the different studies. Study 2 is based on manifest variables and is combined with Study 1 the studies that are most vulnerable to validity threats. Study 2 is vulnerable since it relies on only one measurement instrument in combination with a theoretically, not empirically, driven decision regarding which subscale to include in the analysis. Study 1 has the advantage of including results across several different tests when operationalizing constructs, which reduces the risk of construct under-representation. However, such an approach increases the chance of including something irrelevant in the constructs. The risk of having done so is perhaps greater in this meta-analysis than in many other meta-analyses. The reason is the lack of a good empirical examination of EF as an overall construct and of EF at the domain level. There are even challenges to validity at the test level given the well-known problems with task impurity within the EF field. In cases where we lacked empirical guidance on how to categorize study outcomes across EF domains, our decisions when operationalizing the constructs were guided by theory.

Study 3 used latent variables; hence, the hypothesized relationship between the observed indicators and underlying latent construct was tested by confirmatory factor analysis. Confirmatory factor analysis clarifies a construct's factor structure and removes random measurement errors, hence establishing construct validity (Little, 2013; Tabachnick & Fidell, 2013). The use of invariance testing of factors ensured construct validity across language groups (Brown, 2015). Note, however, that the dimensionality of the different linguistic constructs was not tested before they were used in the SEM model. The latent factors are assumed to be multi-dimensionally correlated, an assumption that is supported by an examination of the correlation matrix of the parcels in the model. This implies that the underlying language abilities across the different latent variables are not accounted for, which raises a validity threat. Reports in Study 3 of the percentage explained by different variables in reading

comprehension should therefore not be mistaken for the unique contribution of the different linguistic aspects of reading comprehension.

Thus, when interpreting the results presented in this PhD project, it is important to bear in mind how the constructs in the different studies were operationalized.

### **3.5 Ethical considerations**

#### **3.5.1 Research on vulnerable groups**

In this thesis, bilingual learners as a group are singled out; therefore, it is reasonable to reflect upon how the knowledge of the project could influence bilingual children as a group or the different subgroups of bilingual children investigated. The Guidelines for Research Ethics in Social Sciences, Humanities, Law and Theology clearly state that the researcher must safeguard the participants against harm and unreasonable strain (NESH-publikasjon, 2006). This is also highlighted in the Guidelines section regarding respect for vulnerable groups. This section describes the researchers' responsibility to protect the group from unreasonable strain. Judging by the current political climate in Norway, immigrants in Norway could be considered a vulnerable group. To view immigrants as a vulnerable group is also in line with Liamputtong (2007), who expands the vulnerable group to also include people from ethnic minority backgrounds. It is not clear, but it is likely that the children from minority speaking households in my project are in fact immigrants. They are certainly from ethnic minority backgrounds. This put extra strain on my ethical responsibility as a researcher.

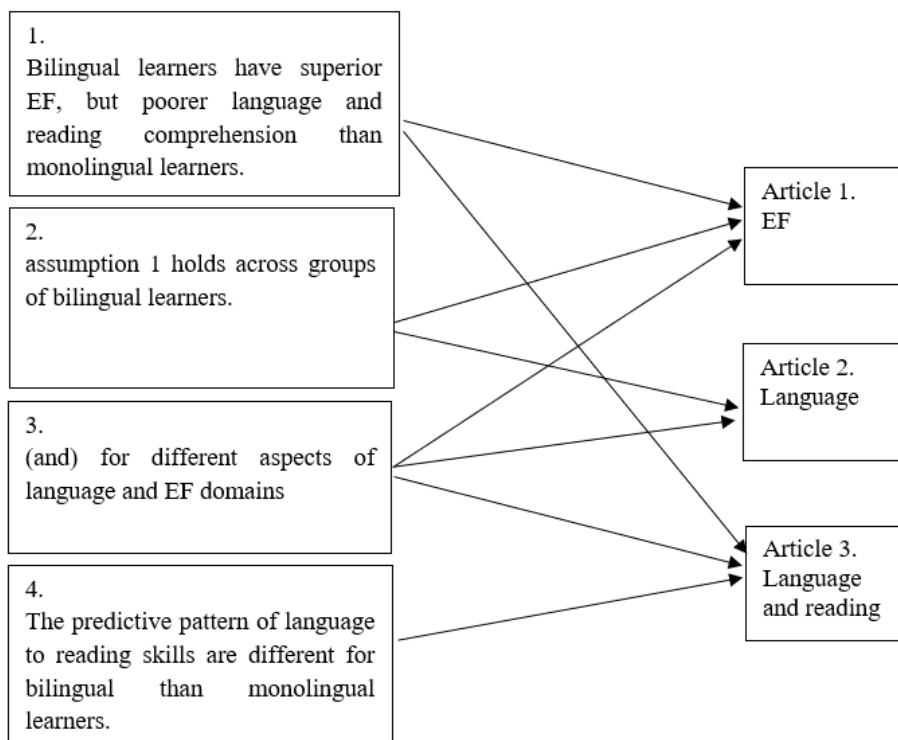
Studies 2 and 3 were approved by the Norwegian Social Science Data Service and followed national research ethical guidelines. Parents signed a written consent form and were informed that they could withdraw their child from the study at any timepoint.



## 4 Results

The aims of this thesis are investigated through the 4 hypotheses illustrated in Figure 1.

Figure 1. Visual illustration of which hypotheses are addressed in the different articles of this thesis.



#### **4.1 Research question 1: Do bilingual learners have superior EF skills and poorer language skills?**

Study 1 investigated the alleged bilingual advantage in EF by analysing data from 143 studies with 583 effect sizes. The overall effect size was marginal and in favour of bilingual learners,  $g = 0.060$ , 95% CI [0.003, 0.116]  $df = 138$ ,  $p = 0.040$ .  $I^2$  was 72.17% with a  $Tau$  value of 0.36. This means that there was much heterogeneity within the results. Article 1 provides only  $Tau$ , however,  $Tau$  can be used to calculate a prediction interval. A prediction interval displays the true variation in effect size across means for different bilingual populations and is thereby easier to interpret than  $Tau$ . The prediction interval presented in this extended abstract is calculated by the rough formula presented in Borenstein et al. (2011):  $\mu$  plus/minus  $1.96 \times Tau$ , where  $\mu$  represents the mean effect size. This formula provides an approximate prediction interval given an infinite number of studies. The prediction interval for overall EF was -0.646 to +0.66, indicating that the true effect sizes for different groups of bilingual learners lay within a range of -0.646 to +0.66 Hedges  $g$ . Furthermore, the Eggers test showed significant asymmetry ( $\beta = 1.08$ ,  $SE = 0.39$ ,  $Z = 2.80$ ,  $p = .005$ ). When controlling for small study effect by PET analysis, there was no bilingual advantage in EF (-0.16, 95% CI [-0.325, 0.005],  $p = 0.58$ ).

Since the meta-analysis of Melby-Lervåg and Lervåg (2014) found that bilingual learners, independent of their amount of L2 exposure, had lower levels of language compared to monolingual learners, this thesis compares the language levels of monolingual learners to those of *early bilingual learners*. Study 2 finds that *early bilingual learners* in toddlerhood have lower language levels than their monolingual peers. The effect size difference was  $d = 0.54$ . Study 3 finds that even after a range of 8-10 years of exposure to L2, the average *early 10-year-old bilingual learners* lag behind their average monolingual peers by  $d = 0.53$  (composite score of all linguistic aspects examined in article 3, text

cohesion vocabulary measured as 0.00  $d$  difference due to non-significant results of the analysis).

#### **4.2 Does the claim of superior EF levels and poorer language levels hold across different groups of bilingual learners?**

The main analysis in article 1 displayed a large degree of heterogeneity within overall EF. To explore the heterogeneity, 11 moderator analyses were performed, both methodological ones (sample size, publication year and laboratory) and moderators representing sample characteristics when a bilingual advantage is theorized to occur (AoA, L2 proficiency, age, degree of balanced bilingualism, and SES level [equal levels across monolingual and bilingual learners]). Since it is impossible to randomly assign bilingualism, all studies in the meta-analysis were group comparison studies. Sample characteristics (non-verbal IQ and SES) representing uneven group comparisons in the primary studies were therefore also included as moderators in the analysis.

The only moderators significantly related to differences in overall EF outcomes were the moderators laboratory and SES level. The laboratory of York was identified as a laboratory driving up the effect sizes in favour of a bilingual advantage. Only bilingual children from the medium-SES class had a bilingual advantage, in contrast to children from the low- or medium- to high-SES classes.

Articles 2 and 3 compare the language levels of different groups of *early bilingual learners* with those of a monolingual control group. In total, 4 group comparisons of *early bilingual* and monolingual learners' language levels were examined. In article 2, the bilingual toddlers' language levels were compared to the levels of a monolingual control group. The analysis in Appendix 4 shows additional analysis of a subgroup of these bilingual toddlers compared to the monolingual

control group. This subgroup was composed by merging the subgroup of bilingual toddlers who had some L2 input at home with the subgroup exposed mainly to L2 at home. (Article 2 found the language levels of these two groups of *early bilingual learners* to be equal.) However, with no direct comparison to a monolingual control group, the analysis in article 2, research question 2, shows the difference in language levels between the *early bilingual* toddlers exposed mainly to L1 and the *early bilingual* toddlers with some or mainly L2 exposure at home. When the difference in language level between the bilingual toddlers with L2 exposure at home and the monolingual group is known, an indirect comparison of the bilingual toddlers exposed mainly to L1 at home and the monolingual learners can be made. The 4<sup>th</sup> comparison of *early bilingual* and monolingual learners' levels of language is presented in article 3. The children were then 10 years old.

The results from article 2, the additional analysis in Appendix 4 and the results in article 3 all display differences that do not favour the *early bilingual learners* both in investigated age levels and almost all aspects of the pre-adolescents' language skills. For pre-adolescents, the language levels were equal across language groups for one aspect only ( $d = 0.34$ , n.s.). The effect size differences for the other linguistic aspects favour monolingual learners ( $d = .78$ ,  $d = .60$ ,  $d = 0.74$ ). The difference in language levels between the *early bilingual* toddlers and their monolingual peers was  $d = 0.54$  in favour of the monolingual learners. The smallest difference in language levels identified in toddlerhood ( $d = .39$ ) was between the subgroup of bilingual toddlers who were exposed to some or mainly Norwegian input at home and the monolingual control group. Note, however, that the majority of these children might well have had a native majority language-speaking parent at home (see Table 1 in section 3.2.1). The language levels of the *early bilingual* toddlers exposed mainly to L1 at home were  $d = 0.56$  lower than those of *early bilingual* toddlers with some L2 exposure at home; hence, this subgroup also had lower language levels than their monolingual peers. Table 1 in

section 3.2.1 indicates that this subpopulation of bilingual learners most likely did not have a native Norwegian-speaking parent at home, although considering the amount of missing data, it is difficult to draw such conclusions.

### **4.3 Does the claim of superior EF levels and poorer language levels hold across different cognitive domains and linguistic aspects?**

The additional analysis presented in Appendix 4 and article 2 shows that the *early bilingual learners* had poorer language comprehension in toddlerhood. In the pre-adolescents, the *early bilingual learners* also exhibited poorer vocabulary levels ( $d = 0.74$ ,  $p < 0.001$ ), listening comprehension ( $d = 0.60$ ,  $p = 0.002$ ) and reading comprehension ( $d = 0.78$ ,  $p = 0.009$ ). Morphology skills could not be investigated in the preadolescents due to invariant test results. The pre-adolescent bilingual learners had text cohesion vocabulary skills equal to those of the monolingual learners ( $0.34$   $d$ ,  $p = 0.102$ ).

Regarding the EF domains, the moderator analysis showed that different EF domains significantly moderated the outcome of overall EF ( $F = 4.59$ ,  $df = 23.2$ ,  $P = 0.002$ ). The results of monitoring ( $g = 0.24$ , 95% CI [0.058, 0.428]), switching ( $g = 0.329$ , 95% CI [0.192, 0.446]) and cold inhibition ( $g = 0.196$ , 95% CI [0.034, 0.358]) were in favour of bilinguals, in contrast to hot inhibition, inhibition of attention, attention, planning and working memory were the results were either similar for the monolingual and bilingual learners or unreliable. The Egger's regression test showed a significant asymmetry for the domain of monitoring ( $\beta = 2.580$ ,  $SE = 0.77$ ,  $Z = 3.35$ ,  $p = 0.001$ ). Partialing out the small study effect with PET analysis showed that the effect sizes were then nonsignificant. For cold inhibition, a regression analysis of published versus unpublished studies showed a larger effect for published studies, in spite of a symmetric funnel plot. In fact, the mean effect of unpublished papers of cold

inhibition was in favour of monolinguals. Switching was not influenced by publication bias or the small study effect. Thus, switching was the only domain in which a bilingual advantage was detected after partialing out the small study effect or controlling for publication bias. An analysis of the mean effect size for the domain of switching showed a bilingual advantage of  $d = 0.27$ , but there was a large observed variation in effect sizes (ranging from -0.517 to 1.667).

The analysis of the relationship between task on overall mean at the domain level was either unreliable (switching and cold inhibition) or not significant (monitoring).

#### ***4.4 Is the predictive pattern of language and decoding skills to reading comprehension different for bilingual and monolingual learners?***

The prediction of linguistic skills and decoding to reading comprehension was equal across groups,  $F(5.0) = 0.555$ ,  $p = 0.734$ . When controlling for all literacy variables, vocabulary did not predict reading comprehension ( $p = 0.872$ ). Listening comprehension had the greatest impact on reading comprehension and explained 26.01% of the variance in reading comprehension, followed by text cohesion vocabulary, which explained 18.49% of the variance. Decoding skills explained only a marginal variation in the 5<sup>th</sup> graders' reading comprehension ( $R = 0.009\%$ ).

## 5 Discussion

The present thesis aims to explore cognitive profiles related to language and EF in bilingual children compared to monolingual children and whether there are differences in predictive patterns for reading comprehension between the groups.

Thus, the overall hypotheses examined are as follows:

1. Bilingual learners have superior EF skills and poorer language skills than monolingual learners.
2. Hypothesis 1 holds across different groups of bilingual learners.
3. Hypothesis 1 holds for different aspects of language and EF domains.
4. The predictive patterns of language and decoding to reading comprehension are different for bilingual and monolingual learners.

### **5.1 *Bilingual learners have superior EF levels and poorer language levels than monolingual learners***

Regarding superior EF skills, the results from Study 1 showed little support for a bilingual advantage in overall EF. The overall effect size was  $g = 0.060$ , 95% CI [0.003, 0.116], displaying a marginal advantage in favour of bilingual learners. There were, however, signs of publication bias, and when controlling for the small study effect, there were no bilingual advantages. This finding is in conflict with the results of the meta-analysis of Adesope et al. (2010), who found an advantage in overall EF of  $d = 0.41$ . However, there are multiple differences between these two studies. The analysis of Adesope et al. (2010) relied on a rather unconventional definition of EF, where 62% of the effect sizes contained measures not traditionally regarded as EF measures (e.g., metalinguistic awareness, metacognitive awareness and abstract and symbolic

representations). Adesope et al.'s (2010) meta-analysis is also older, and more recent studies have challenged the hypothesis of a bilingual advantage; therefore, the primary articles included in Adesope et al. (2010) and Study 1 differ (39 studies and 63 effect sizes versus 143 studies and 583 effect sizes). However, differences in papers due to publication year cannot alone explain the different outcomes of the present meta-analysis and that of Adesope et al. (2010). Even when the two analyses extracted the same studies for coding, different effect sizes from these studies were coded. The present meta-analysis coded all measures of working memory reported in the primary studies, and Adesope et al. (2010) neglected to include such measures on several occasions. One example is the study of Bialystok, Luk and Kwan (2005). Here, Adesope et al. (2010) coded measures of metalinguistic awareness, which are not included in the present analysis, while neglecting to code measures of working memory. In fact, of the 5 primary studies reporting measures of working memory included in both analyses, Adesope et al. (2010) included working memory measures from only one study. Note also that Bialystok (2009) labelled working memory the EF domain unlikely to be affected by bilingualism.

Another possible reason for the different outcomes is that Adesope et al. (2010) included primary studies of participants of all ages, while the present study included only studies of children (0-18 years). Based on this, one could wonder whether a bilingual advantage exists in adulthood but not in childhood. However, this is unlikely; a meta-analysis by Lehtonen et al. (2018) found marginal support for a bilingual advantage in overall EF in adulthood ( $d=0.01$ ), but the advantage disappeared when controlling for publication bias. Where Adesope et al. (2010) included only a limited number of effect sizes (63 effect sizes), aggregated to the study level, Lehtonen et al. (2018) included an impressive number of effect sizes (891). Furthermore, as in the present meta-analysis, Lehtonen et al. (2018) controlled for the dependency of multiple outcomes within studies, thereby presenting results less hampered by measurement errors



compared to a dataset of aggregated effect sizes (Tanner-Smith & Tipton, 2014). Hence, the present meta-analysis and that by Lehtonen et al. (2018) provide a more comprehensive number of effect sizes as well as more methodologically solid analyses than the analysis by Adesope et al. (2010). Additionally, the results of both meta-analyses present a coherent picture, thereby providing convincing evidence against the likelihood of the existence of a bilingual advantage in overall EF.

One caveat should, however, be mentioned. Since these meta-analyses examine a bilingual advantage in EF, the validity of EF as an overall concept comes under scrutiny. Unfortunately, there are surprisingly few empirical studies of EF as an overall construct, EF at the domain level, and EF measures at the test level. Most discussions and definitions of these terms have largely relied on assumed theoretical differences (Baggetta & Alexander, 2016). Based on the current knowledge of EF, EF is an overall construct consisting of different but interchangeable EF domains (Miyake & Friedman, 2012; Miyake et al., 2000). Furthermore, the most mentioned and agreed-upon EF domains are attention, WM, inhibition, switching and planning (Baggetta & Alexander, 2016). It is on the basis of this definition, in addition to measures of domain monitoring, that the present meta-analysis and the one by Lehtonen et al. (2018) built their case. However, it should be highlighted that several empirical studies have identified lower correlations among EF measures than the suggested correlation of .6 (e.g., Cirino et al., 2018; Karr et al., 2018; Miyake et al., 2000) or no correlation (Paap & Greenberg, 2013). This is why correlations in the present meta-analysis were set at .3 when addressing dependency in the dataset. However, low correlations among EF measures provide problems that cannot be solved only by fixing the correlations between the outcomes of effect sizes. These low correlations provide a reason to question to what extent the current understanding of EF as an overall construct is accurate. Hence, the conclusions of a meta-analysis investigating a bilingual advantage in EF are vulnerable to construct validity. Should the definition of overall EF change, the effect

sizes of different measures would perhaps be included/excluded; thus, the results of the analysis could possibly alternate. Under such circumstances, a bilingual advantage in EF might occur. However, based on the current knowledge of EF and the results of the present meta-analysis of children and the meta-analysis by Lehtonen et al. (2018) of adults, a bilingual advantage in overall EF seems unlikely.

Regarding the hypothesis of lower language levels, since the meta-analysis by Melby-Lervåg and Lervåg (2014) found lower language levels of bilingual learners as an overall group compared to monolingual learners, this thesis compares the language levels of monolingual children with those of bilingual children with sample characteristics theoretically suggested to positively influence L2 development. Many of the factors influencing L2 development seem to be intertwined and separating the specific impact of one factor from the impact of other factors is difficult. However, early AoA, having one native majority language-speaking parent and length, amount and quality of L2 exposure seem to positively drive L2 development, and high-SES children seem to benefit more than low-SES children (Jim Cummins, 2017; Hammer et al., 2014; Kovelman et al., 2008; Oller et al., 2011; Place & Hoff, 2011, 2016; Unsworth, 2013, 2016; Unsworth et al., 2014). As the evidence stands now, children who have all these advantages are more likely to obtain L2 levels equal to those of monolingual learners.

These sample characteristics are coherent with the sample characteristics of the *early bilingual learners* in Study 3. Two-thirds of the bilingual learners in Study 3 had a native Norwegian-speaking parent. All bilingual learners had started in ECEC institutions at least by the age of 2. Even though attending ECEC institutions ensures that the amount of L2 exposure for *minority language learners* rises, this does not necessarily influence growth in L2 skills. In fact, attendance in ECEC institutions unregulated by a curriculum does not in itself promote young bilingual learners' L2 skills (Hoff et al., 2018). Norwegian ECEC institutions do not follow a set curriculum, but the teachers follow a

social pedagogic tradition and are obligated to nurse interest and learning through formal and informal settings (Moser, 2010). Furthermore, The Norwegian Framework Plan for the Content and Tasks (The Norwegian Directorate for Education and Training, 2017) highlights the responsibility of ECEC teachers to ensure all children's participation in learning activities that can promote their language levels. Hence, there are some standards for the management of ECEC institutions aiming to ensure high-quality language input to all children enrolled in the institutions.

Taken together, the *early bilingual learners* in Study 3 were richly exposed to Norwegian for 8-10 years. Furthermore, the monolingual and bilingual samples came from comparable SES backgrounds, and both samples came mainly from medium- to high-SES backgrounds. Study 3 finds that even after a range of 8-10 years of exposure to L2, the average *early 10-year-old bilingual learner* lags behind the average monolingual peer by  $d = .53$  (composite score of all linguistic aspects examined in article 3, text cohesion vocabulary measured as  $d = .00$ , difference due to non-significant results of the analysis). Interestingly, this gap in language levels in pre-adolescence is similar to the gap identified in toddlerhood. This might suggest that the gap observed in toddlerhood remains stable until the age of 10. Note, however, that the two studies assessed language skills with different tests, the analytical approaches differed, and, most importantly, the participants in these two studies only partly overlapped (article 2,  $N = 161$  bilingual learners; article 3,  $N = 91$  bilingual learners). Furthermore, growth in L2 skills was not examined. To draw conclusions regarding the stability of the gap based only on these two effect sizes is therefore premature. It is safer to generalize the results of Study 3 to other bilingual samples. Since the *early bilingual learners* in Study 3 had poorer language levels than the monolingual learners, most groups of bilingual children that include *minority language learners* probably do so if their language skills are tested by sensitive enough tests.

Hence, the results of this thesis support the claims of poorer language levels, while the proof of a bilingual advantage in overall EF is not convincing.

## **5.2 The assumption of superior EF and poorer language levels holds across different groups of bilingual learners**

The meta-analysis found little support for a bilingual advantage in overall EF. However,  $I^2$  was 72.17%, and the *Tau* value was 0.36. The prediction interval for overall EF was -0.646 to +0.66, indicating that the true effect sizes for different groups of bilingual learners lie within a range of -0.646 to +0.66 Hedges *g*. Notably, there was no marginal bilingual advantage for all bilingual children; some had a moderate to large disadvantage in EF, while other groups of bilingual learners had a moderate to large advantage. This implies that a bilingual advantage might exist for the bilingual groups theorized to have the largest advantage in EF.

Since enhanced EF skills are a result of bilingual learners' experiences with language processing, the alleged advantage is theorized to be larger for bilingual groups that have more experience with bilingual language processing. It has been suggested that an equal degree of fluency in both languages, frequent changes between the two languages, and early AoA are factors that positively affect bilingual learners' level of attention control (Bialystok, 2017). While AoA indicates when a child is introduced to L2, it says little about the amount of bilingual language experience the child has had. L2 proficiency could act as a proxy and hence could promote a bilingual advantage (Pelham & Abrams, 2014). Furthermore, the bilingual advantage takes time to develop; hence, it strengthens during the years of childhood (Bialystok, 2017). SES levels are also theorized to cause a bilingual advantage, but there is less agreement on how they influence EF. One theory suggests that children with low SES would have a larger advantage (Naeem et al., 2018). An alternative suggestion mentioned in the study of De Cat et al. (2018) is

that SES level could function as a threshold where a certain level of SES must be obtained before the bilingual advantage occurs.

In summary, according to theory, the following bilingual groups should exhibit a larger advantage in EF:

1. Bilingual learners with low AoA over bilingual learners with high AoA.
2. Older bilingual children over younger bilingual children.
3. Balanced bilingual learners over unbalanced bilingual learners.
4. Proficient L2 speakers over non-proficient L2 speakers.
5. Low-SES bilingual learners over medium- and high-SES bilingual learners.
6. Alternatively, medium- to high-SES bilingual learners over low-SES bilingual learners.

Moderator analysis was used in the meta-analysis (Study 1) to examine whether differences in the overall outcome were related to differences in sample characteristics or methodological choices. Comparisons in primary studies of groups that are uneven in terms of third factors that could influence the examined variable could affect the results of a meta-analysis. Hence, the relationship between differences in overall EF and effect size differences across samples in primary studies on non-verbal IQ and SES was also explored (Friedman et al., 2006; Lawson, Hook, & Farah, 2018).

The results of the meta-analysis showed that differences in age, the degree of balanced bilingualism, L2 proficiency and AoA were unrelated to differences in overall EF. There was, however, a relationship between SES level and overall EF ( $F(15.5) = 4.49, p = 0.029$ ), where medium-SES bilingual children had an advantage over their medium-SES monolingual peers ( $g = 0.175, 95\% \text{ CI } [0.070, 0.280]$ ). If SES level worked as a moderator of the bilingual advantage, where bilingualism especially promoted a bilingual advantage in low-SES children (Naeem et al., 2018), a larger advantage for low-SES bilingual learners should

have been detected in the present meta-analysis. The results, however, indicate the opposite. There is a larger bilingual advantage in medium-SES children than in low-SES children ( $\beta = 0.235$ ;  $p = .017$ ). In fact, when examining the 95% CI of low-SES children, it is evident that there is not even a significant bilingual advantage for the low-SES bilingual children,  $-0.060$ , 95% CI  $[-0.218, 0.099]$ .

Regarding the hypothesis of an SES threshold, where a certain level of SES is necessary for a bilingual advantage to emerge, one should expect that high- and medium-SES bilingual children would also have a bilingual advantage. Here, the results show a larger effect for the medium-SES children than for the high- to medium-SES samples ( $\beta = 0.230$ ;  $p = .046$ ), combined with a non-significant result of the relationship between high-SES children and overall EF ( $g = -0.055$ , 95% CI  $[-0.258, 0.147]$ ). Hence, the results show no support for the hypothesis that SES level is related to an overall EF advantage for certain groups of bilingual learners.

The heterogeneity identified in the analysis, and displayed by the prediction interval, could not be explained by uneven comparisons of the bilingual and monolingual groups in the primary studies concerning measures of non-verbal IQ or SES or by the bilingual learners' language experiences. It is therefore more likely that some of the heterogeneity within the results could be ascribed to methodology. In fact, when groups of laboratories were created by comparing the studies of the laboratory of York to a group comprising all the other laboratories, laboratory moderated the results in overall EF ( $F = 6.89$ ,  $df = 76.2$ ,  $p = .011$ ). There was a larger effect of the laboratory of York than of other laboratories ( $\beta = 0.153$ ;  $p = .011$ ). In fact, the laboratory of York reported a bilingual advantage ( $g = 0.168$ , 95% CI  $[0.073, 0.263]$ ), while this was not the case for the other laboratories ( $g = 0.015$ , 95% CI  $[-0.054, 0.084]$ ). Overall, the results do not suggest that bilingual learners' superior EF skills can be generalized across different groups of bilingual learners.

Regarding the hypothesis of poorer language levels across different groups of *early bilingual learners*, the results of this thesis are mainly supportive. The *early bilingual toddlers* in Study 2 had lower levels of language than the monolingual control group. The additional analysis in Appendix 4 examines a different subsample of the bilingual participants in Study 2 and the monolingual control group. These analyses find the same result. The monolingual learners had better language levels than the *early bilingual toddlers* with some or mainly Norwegian exposure at home ( $d = .39$ ). The latter group may well consist mainly of bilingual children with a native language majority language-speaking parent at home (see Table 1). If this is the case, then most of these bilingual learners were likely introduced to Norwegian at birth. Thus, the AoA is the same for this subgroup of bilingual learners and their monolingual peers. However, since they also received some input in a minority language, the amount of Norwegian input is less than for the monolingual control group. Given similar AoA and the presence of native language input in the majority language, these children should, according to De Houwer (2009a, 2009b), have comparable language levels. De Houwer acknowledges that you can find differences in language levels between a particular monolingual and a particular bilingual child but states that the magnitude of these differences is no greater than the variation within groups of monolingual children and groups of bilingual children. Clearly, a difference in language levels of  $d = .39$  at the group level is within the normal variation of monolingual children. However, if there were no differences across the *bilingual first learners* and the monolingual controls, the differences in language levels should not be significant and preferably close to an effect size difference of 0. Hence, the results imply that the claims of De Houwer are not supported.

The results of the present thesis are more in line with the hypothesis of this dissertation and Bialystok and Feng's (2011) findings. There are, however, larger differences in the observed effect size between *bilingual first children* and monolingual learners in Bialystok and Feng's (2011)

study and the present study. Bialystok and Feng (2011) find a large difference in favour of monolingual learners, while the effect size difference in the present study is small to medium. One possible reason could be differences in examined samples. The present sample is of mainly medium- to high-SES children, and SES is a factor that limits the gap in language levels between monolingual and bilingual learners (Oller et al., 2011). Bialystok and Feng (2011) does not provide information of their sample's SES background or the parents' nationality. It is therefore unclear whether the two samples SES levels are comparable as well as whether the *bilingual first children* in Bialystok and Feng's (2011) study received native language input at home.

There is, however, a large drawback to this line of argumentation. Table 1 shows that all *bilingual first children* (identified by having one native majority language-speaking parent at home) belong to the groups of *mostly Norwegian* exposure at home and *some Norwegian* exposure at home. Parental Norwegian exposure in toddlerhood, in combination with having one native Norwegian-speaking parent and being raised in Norway, points to the likelihood of a bilingual child's AoA at birth. One can thereby assume that they indeed are *bilingual first children*. However, it is more problematic to assume that the case is the same for the 4 bilingual toddlers with some parental Norwegian exposure at home from non-native Norwegian speakers. These parents could have started interaction with the child in Norwegian at a later timepoint than from birth. The number of participants who belong to the group with some Norwegian exposure at home, where certain knowledge of the parents' nationality is missing, is therefore worrisome (35 of 82 participants). Hence, it is possible that the numbers of children with a possible later onset of Norwegian than from birth in this subgroup are much larger than assumed. If that is the case, then the results of the present analysis cannot be ascribed as non-supportive of De Houwer's (2009a, 2012, 2018) claims of equal language levels of *bilingual first* and monolingual learners. The lower L2 levels would then be representative of a subgroup



of *early bilingual learners*. Note, however, that the gap to monolingual learners then differs for different groups of *early bilingual learners*. There is a gap in language level of .56 *d* between *early bilingual learners* mainly exposed to L1 at home and *early bilingual learners* exposed to some/mainly exposed to Norwegian at home.

Study 3 examines a subsample of the *early bilingual learners* examined in Study 2; however, the participants then had become 10 years of age. At this point, we knew the nationality of the *early bilingual learners*' parents. Two-thirds of them had a native majority language-speaking parent. Native language speaking input enhances growth in language skills (Hoff et al., 2018; Place & Hoff, 2016). However, in spite of low AoA, medium to high SES, and long and rich exposure to Norwegian, the *early bilingual learners* in preadolescence still lagged behind their monolingual peers.

The gap in pre-adolescents varies in size for different linguistic and literacy subskills and ranges, from  $d = .60$  to  $d = .78$ , with the exception of one linguistic aspect and decoding skills. This is less of a gap than has been found in some of the often-cited reviews of *minority language children* exposed to L2 at age 5-7 starting from around school entry. These children's language skills are often found to be approximately 1 SD below the mean of the skills of monolingual children (Jim Cummins, 1984, 2017). The fact that the gap is less for *early bilingual learners* than for some groups of *minority language learners* is perhaps not surprising, but there is still a gap between them and their monolingual peers. The gap identified in the present study is in line with other studies of *early bilingual learners* that have found lower levels of language in *early bilingual learners* than in monolingual learners in spite of early AoA and rich and long exposure to L2 (Bonifacci & Tobia, 2016; Kovelman et al., 2008; Vernice & Pagliarini, 2018). Note also that some of the gaps in *early bilingual learners*' language levels are even larger than the one identified in the present study, which is also even larger than the gap in

the more optimistic reviews of *minority language learners* (Jim Cummins, 1984, 2017; Hakuta et al., 2000).

In contrast to the present study, Hsu et al. (2019) find equal language levels between monolingual and *early bilingual* 3<sup>rd</sup> graders. The sample of the present study and the sample of Hsu et al. (2019) were both introduced to L2 from at least the age of 2. Given the similarities of AoA and length of exposure, the differences in the results of these studies are interesting. The differences are, however, most likely caused by differences in extracting bilingual participants in the studies. The present thesis included bilingual children based on their AoA, whereas the bilingual children in Hsu et al. (2019) were invited to participate in the study only if their language levels were above the cut-off point of norms for monolingual children. Hence, the bilingual learners were pre-selected to have language levels comparable to those of their monolingual learners. Based on the results of the present study and the prior literature on this topic, true differences in language levels most likely exist in pre-adolescent *early bilingual* and monolingual learners. Such differences are then likely to arise in comparing unselected *early bilingual learners'* language levels to monolingual learners' language levels. Furthermore, even if high SES is suggested to be a protective factor that reduces the size of the gap in language levels (Oller et al., 2011), it does not seem to be sufficient to ensure equal language levels in pre-adolescent *early bilingual* and monolingual learners. Language levels of *bilingual first* pre-adolescents might, however, not be affected.

The main conclusion regarding the hypothesized bilingual advantage in EF does not hold across groups – not even the bilingual groups theorized to have the largest advantage in EF. Regarding a bilingual disadvantage in language, reaching a conclusion is even more complicated. At present, it seems likely that poorer language levels can be generalized to different groups of *early bilingual learners* but perhaps not to *bilingual first children*.

### **5.3 Bilingual learners have superior EF levels and poorer language levels across different EF domains and linguistic aspects**

The results from the present analysis, using RVE to examine the extent to which there is a bilingual advantage at the domain level, were mixed. The present thesis is the first study to investigate a possible bilingual advantage for the domains of *planning* and *hot inhibition*. The moderator analysis of the different EF domains on overall EF were however unreliable for these domains due low degrees of freedom. Regarding *attention*, there were no significant differences between monolingual and bilingual learners in the present study, which is in line with Lehtonen et al. (2018).

The present meta- analysis did not detect a bilingual advantage for the domain of *working memory either*. Grundy and Timmer (2016), however, do find a small to medium effect size in favour of bilingual learners ( $d = .20$ ). Even though their analysis is of adults and children, it comprises fewer extracted studies and effect sizes than the analysis in the present study ( $k = 88, m = 27$  versus  $k = 155, m = 94$ ); thus, the results of the present study seem more likely. This assumption is supported by the results of Lehtonen et al. (2018), who find a marginal advantage for adult bilingual learners in working memory ( $d = 0.07$ ), which disappears after controlling for publication bias.

For the domain of *monitoring*, the moderator analysis of domains on overall EF found support for an advantage in favour of bilingual learners ( $g = 0.243, 95\% \text{ CI } [0.058, 0.428]$ ). The task moderator analysis was not significant; hence, the use of different tasks did not moderate the overall effect size difference in monitoring. The Eggers test was positive, indicating publication bias. After controlling for the small study effect, the advantage in monitoring disappeared. The lack of existence of a bilingual advantage in monitoring is supported by Lehtonen et al. (2018),

who find a small advantage in favour of bilingual learners ( $d = 0.15$ ) that vanishes after controlling for publication bias.

For *cold inhibition*, the moderator analysis EF domains on overall EF indicated an advantage in favour of bilingual learners ( $g = 0.196$ , 95% CI [0.034, 0.358]). Task moderator analysis of cold inhibition was unreliable; hence, the heterogeneity within the results could not be explained by differences in task used to measure cold inhibition. A moderator analysis of published versus non-published studies found that the results were influenced by publication bias. The results of other meta-analyses vary for this domain. Donnelly (2016) examines inhibition costs in computerized tasks and finds a small to medium effect in favour of bilingual learners ( $d = .24$ ). Note that this is a much more restricted examination of inhibition. There is evidence of publication bias, but Donnelly (2016) does not attempt to control for it. In contrast to Donnelly (2016), Lehtonen et al. (2018) find a small advantage in favour of bilingual learners ( $d = 0.11$ ) that disappears after controlling for inhibition. Lehtonen et al. (2018) include measures coded as inhibition of attention in the present analysis as part of their construct of inhibition; hence, a close comparison across domains is somewhat difficult. In the present analysis, there were no differences between monolingual and bilingual learners in *inhibition of attention*. Thus, inclusion of measures of inhibition of attention in the construct of inhibition could potentially have reduced the magnitude of the overall effect size in favour of bilingual learners in Lehtonen et al.'s (2018) study. Due to signs of publication bias in all three analyses, the prediction interval identified in the present study, and the low effect size difference in the comprehensive meta-analysis by Lehtonen et al. (2018), the likelihood of a bilingual advantage in inhibition is low.

Taken together, the results do not support a bilingual advantage for the domains of attention, inhibition of attention, cold inhibition, hot inhibition, planning, working memory or monitoring. However, the results for *switching* were somewhat different. The overall effect size for

the domain switching on overall EF was  $g = 0.329$ ,  $p < 0.001$ . The Eggers test did not detect signs of publication bias in the final dataset. Note, however, that the results for the publication bias analysis for switching appear less robust than the publication bias analysis of the other domains (see section 3.1.2). When publication bias was examined using the dataset where the sample size was corrected for multiple comparison groups, the Eggers test was positive. After controlling for the small study effect by the PET- PEESE analysis in this dataset, switching was no longer a bilingual advantage. Hence, the result of switching appears less robust than for the results of the other EF domains.

Since the results of switching remained significant after examination of publication bias and small study effect in the present dataset, the overall mean of switching was examined. The mean effect size for switching showed a small advantage for bilinguals ( $d = 0.27$ ). The observed scores of switching did however indicate large observed variation, ranging from  $-0.517$  to  $1.667$ , and moderate true variation in the results. Even though a prediction interval displaying the true range of scores for different bilingual groups in switching is not calculated here, the observed scores in combination with the moderate true variation in the results still indicates that there is a possible advantage in switching for some bilingual groups or under certain circumstances, yet not for all bilingual groups in general.

The association between the overall mean of switching and task was examined in order to explore whether task could explain any of the heterogeneity within the results. Due to low  $df$ , the task moderator analysis was unreliable. However, the power of some moderator analyses in RVE is surprisingly low, especially if there is an imbalance of effect sizes in the different levels of categorical variables, which was the case for the moderator analysis of the domain of switching. In this case, the 8 levels of switching tasks mainly contained 6 to 8 effect sizes per task, while two tasks had 18 and 25 effect sizes. Hence, even though the total number of effect sizes was large (84 effect sizes) and thereby

strengthened the statistical power, the unevenness of effect sizes across the different categorical levels could have caused the unreliable outcome of the relationship between task and overall effect size difference in switching. Thus, given more power, task might have moderated the overall results in switching. In fact, task moderator analysis on switching, using an aggregated dataset in CMA, showed that task explained 38% of the variance in effect sizes in switching. Although these results do not control the dependency in the dataset, the task of switching was one of the few differences observed across the different approaches to analysing the data in CMA (Borenstein et al., 2005) and Robumeta (Fisher & Tipton, 2015).

Taken together, the inconsistency across analyses is restricted mainly to the domain of switching, which suggests that the results concerning switching in article 1 are less robust than the rest of the results. Although the variations in the observed scores of switching indicate that there is a bilingual advantage for some bilingual groups, it remains unclear which bilingual groups are affected and under what circumstances the advantage emerges. It is also unclear whether such an advantage is at the construct level or task level.

The results of the present analysis and the other meta-analysis examining a possible bilingual advantage on domain levels must, however, be interpreted in light of theoretical and empirical knowledge of the different EF domains. While one researcher labels a process inhibition of attention (Diamond, 2013), the same process is called executive attention by another researcher (Mezzacappa, 2004). Inhibition of behaviour (cold inhibition) is argued to be part of the construct for executive attention (Fan et al., 2002; Jurado & Rosselli, 2007; Posner & Rothbart, 2000) and inhibition (Diamond, 2013; Donnelly, 2016; Lehtonen et al., 2018). A third theory extends executive attention to higher-level EF constructs that build on WM, switching and inhibition (Bialystok, 2017). Monitoring is sometimes acknowledged to be a separate domain (Hilchey & Klein, 2011), while at other times it is

claimed to be part of inhibition (Donnelly, 2016). The same distinction is relevant at the test level (Paap & Sawi, 2014). The main reason is possibly that EF is a multi-dimensional construct in which several cognitive domains are simultaneously active (Baggetta & Alexander, 2016). This leads to task impurity, where the same test taps several processes simultaneously (Baggetta & Alexander, 2016). Empirical investigations of different test outcomes assumed to tap into the same underlying domain are sometimes poorly correlated and sometimes not correlated at all (Paap & Sawi, 2014). This leads to uncertainty of both test validity and domain validity. Hence, the tests consistently used to assess the possibility of a bilingual advantage in executive functions lack construct validity (Paap & Sawi, 2014). Paap and Sawi (2014) suggest that the low level of convergent validity implies that these measures reflect task-specific mechanisms rather than the construct of EF domains. They therefore urge researchers examining the bilingual advantage to identify and use measures that show better convergent validity. To safeguard against misleading conclusions, in the present analysis, it has been important to be as transparent as possible regarding how the different domains are defined, as well as which test outcomes are coded in the different domains. Note, however, that a different definition of EF domains, or selection of tasks belonging to the different domains, could have affected the results in the present analysis. A clear consensus on how different EF domains should be understood is still lacking (Baggetta & Alexander, 2016). Should consensus of a new understanding of the EF domains be achieved at a later point in time, the effect sizes of the present analysis coded in conflict with this consensus could easily be altered in line with the new understanding of EF domains. Under such circumstances, a bilingual advantage in some EF domains could occur or vanish. The need for more reliable test outcomes of EF is however more difficult to fulfil (Paap & Sawi, 2014). If a meta-analysis is based on primary studies that have examined a bilingual advantage using self-made or otherwise unreliable tests, there is no way to avoid bringing the uncertainty of task impurity to the results of the analysis. This cannot be

fixed by categorizing study outcomes in line with a different definition of EF domains. Thus, some uncertainty in the results of the present analysis would remain even if consensus were achieved regarding the different EF domains. However, as the knowledge of EF domains stands now, it is very unlikely that a new categorization of domains would lead to a bilingual advantage across all EF domains. Hence, the claim of a bilingual advantage across all EF domains is not supported.

Regarding the claim of poorer language levels for *early bilingual learners* holding across different linguistic aspects, the results of this thesis are mainly supportive. For *language comprehension*, Study 2 finds that *early bilingual learners* had poorer levels than monolingual learners in toddlerhood ( $d = 0.54$ ). Most other studies of *early bilingual learners* in this age group have assessed vocabulary skills by CDI (Fenson et al., 2007) because a shortage of studies examining levels in language comprehension makes comparisons of this construct across studies somewhat difficult. Note, however, that in contrast to the other linguistic constructs examined in this thesis, the construct of language comprehension is not empirically tested. This might imply that some of the observed differences across language groups might be caused by comparisons across unequal constructs or use of an assessment instrument with items favouring one of the language groups despite otherwise equal language levels. Furthermore, the operationalization of language comprehension affects the validity of the construct. The construct, as it is operationalized in article 2, might therefore be interchangeable with other linguistic aspects.

In pre-adolescence, the *early bilingual learners* in the present thesis had lower levels of *vocabulary* ( $d = 0.74$ ,  $p < 0.001$ ). Prior studies of *early bilingual learners'* vocabulary levels are mixed, ranging from a large difference in vocabulary levels ( $d = 1.39$ ) to no difference (Hsu et al., 2019; Vernice & Pagliarini, 2018). The non-significant differences in the sample of early bilingual 3<sup>rd</sup> graders in Hsu et al.'s (2019) study are interesting. As with the sample of the present thesis, *early bilingual*



*learners* had been exposed to L2 by the age of 2. However, since they were 3<sup>rd</sup> graders, the years of exposure were fewer than for the pre-adolescents examined in the present thesis. One should therefore perhaps anticipate a larger gap in vocabulary skills in Hsu et al.'s (2019) study simply because the length of L2 exposure is less than for the early pre-adolescents in the present thesis. However, as mentioned earlier, the inclusion criteria for the two studies differs. While the bilingual sample in the present study was recruited simply on the basis of AoA, allowing full variance in L2 skills in the bilingual sample, Hsu et al. (2019) aimed to match the bilingual and monolingual sample in vocabulary skills in the instructional language. This was done by recruiting bilingual children with a minimum of an 85 standard score in L2 performance. The language skills of the monolingual children were, however, not examined. Hence, if the inclusion criterion of an 85 standard score in L2 had not been used, a gap in language levels in favour of monolingual learners would likely have been identified. Moreover, studies that find lower levels of vocabulary in bilingual children find, in line with the present study, that the gap in vocabulary skills is larger than gaps in other linguistic aspects (Vernice & Pagliarini, 2018). The same pattern can be found in several studies of *minority language learners* (Farnia & Geva, 2013; Hutchinson et al., 2003; Lipka & Siegel, 2012), indicating that vocabulary might represent the largest linguistic challenge for bilingual learners.

Regarding *listening comprehension*, the early bilingual pre-adolescents in the present thesis lagged behind their monolingual peers with a medium effect size ( $d = 0.60$ ,  $p = 0.002$ ). Bonifacci and Tobia (2016), however, find no significant differences in listening comprehension between monolingual learners and a sample of *early bilingual* 1<sup>st</sup>-5<sup>th</sup> graders. Interestingly, the AoA of the mixed sample of *early bilingual learners* in Bonifacci and Tobia (2016) was from birth until the age of 4. Hence, the length of exposure of this sample was shorter than that of the pre-adolescent bilingual learners in the present thesis. Bonifacci and Tobia

(2016), however, do not compare group performance across latent means and control SES only at the residence level. Thus, to what extent these methodological decisions influenced the results of their analysis remains unclear.

Differences in *morphology* across groups could not be examined in the present thesis due to an invariant test result. Since very few studies have compared the morphological levels of bilingual learners and monolingual learners across an invariant tested construct, it is difficult to know whether a prior result that indicates relative bilingual strength in morphology is correct (e.g., Barac & Bialystok, 2012; Friesen & Bialystok, 2012; Lipka & Siegel, 2012). Given the amount of theory suggesting a bilingual advantage in morphology (Bialystok, 2001b; James Cummins, 1978; Jim Cummins, 1987; Vygotsky, 1964), along with the results of the meta-analysis of Adesope et al. (2010), who identified a bilingual advantage in metalinguistic awareness ( $g = 0.33$ ), it is quite worrisome that so few studies have secured the quality of the morphological construct across which bilingual and monolingual learners are compared.

Regarding *text cohesion vocabulary*, the pre-adolescent bilingual learners in the present study had text cohesion vocabulary levels equal to those of the monolingual learners ( $d = 0.34$ ,  $p = 0.102$ ). Note, however, that the Cronbach's alpha of text cohesion vocabulary was .631. Low Cronbach's alpha is a sign of limited true variation in the manifest variable. Although low-scale radiality is less problematic in latent variables than in observed variables, the latent variables still form the basis of the manifest variables; hence, low-scale reliability could have impacted the results of the group analysis. Note, however, that a low Cronbach's alpha is sometimes caused by uneven distribution in the strength of factor loadings rather than reflecting how well the Cronbach's alpha captures the true scale reliability (McNeish, 2018). This danger could be limited to calculations of Cronbach's alpha for factors based on items that as a mean do not load sufficiently high on the variable

(Raykov, 1997; Raykov & Marcoulides, 2019). The factor loadings of text cohesion vocabulary ranged from .337 to .761; hence, they were both unevenly distributed, and the mean of the factor loadings was not specifically high. A calculation of the upper bound of the population discrepancy between coefficient alpha and scale reliability could have been performed to evaluate to what extent the Cronbach's alpha of text cohesion provided a suitable measure of the true scale reliability. This would have provided better grounds for evaluating to what extent the Cronbach's alpha of .631 is indeed problematic, hence indirectly evaluating whether there is reason to be concerned about low scale reliability influencing the results of the group analysis. Despite extensive searching, no other studies comparing text cohesion across *early bilingual* and monolingual pre-adolescents have been identified. Studies of pre-adolescent *minority language learners*, however, find a large gap in text cohesion vocabulary in favour of monolingual learners (Droop & Verhoeven, 2003).

The levels of *reading comprehension* between *early bilingual* and monolingual pre-adolescents were not similar. There was a monolingual advantage with a medium to high effect size ( $d = 0.78, p = 0.009$ ). Other studies of *early bilingual learners* had similar findings (Bonifacci & Tobia, 2016; Kovelman et al., 2008). Bonifacci and Tobia (2016) examine 1<sup>st</sup>- 5<sup>th</sup> graders (mean age = 8.97 years) and find a medium effect size in favour of monolingual learners ( $d = .69$ ). Kovelman et al. (2008) examine 2<sup>nd</sup> and 3<sup>rd</sup> graders (mean age = 8.04) and find a large gap in favour of monolingual learners ( $d = 1.47$ ).

Taken together, studies comparing linguistic aspects and reading comprehension across bilingual and monolingual learners have focused mainly on *minority language learners*. Studies of *minority language learners* with 4-7 years of L2 exposure have often found large gaps in reading comprehension and linguistic skills in favour of monolingual learners (Droop & Verhoeven, 2003; Hutchinson et al., 2003; Verhoeven & van Leeuwe, 2012). Due to a limited number of studies of *early*

*bilingual* pre-adolescents, conclusions about their language and reading comprehension levels must be drawn with caution. As it stands now, it seems likely that *early bilingual learners* have lower levels of vocabulary, reading comprehension and listening comprehension than monolingual learners, in line with hypothesis 3 of this thesis. It is, however, not clear whether *early bilingual* and monolingual learners have comparable morphological levels, but their levels of text cohesion vocabulary are equal. Given early AoA, long-term exposure to L2, and medium to high SES, it could be possible for *early bilingual learners* on average to close the gap between them and their average monolingual peers in some linguistic aspects. The claim that *early bilingual learners* have poorer language levels than monolingual learners does not hold across different aspects of language and reading comprehension.

Taken together, the results addressing hypothesis 3 are not supported. The claim of bilingual learners' superior levels in EF and poorer levels in language cannot be generalized across different domains of EF and aspects of language.

#### **5.4 The predictive patterns of language and decoding skills are different for (early) bilingual than for monolingual learners**

The results from Study 3 found the prediction of linguistic skills and decoding for reading comprehension to be equal across the *early bilingual* and monolingual readers,  $F(5.0) = 0.555$ ,  $p = 0.734$ . In spite of extensive searching, Grant et al. (2011) was the only identified study of pre-adolescent *early bilingual learners* that examined how L2 skills predicted L2 reading comprehension. However, that study examines *bilingual first 3<sup>rd</sup> graders* with larger vocabulary gaps than was the case for the *early bilingual 5<sup>th</sup> graders* in Study 3. Differences in the samples might explain the differences in the results. Study 3 found that the linguistic variables predicting reading comprehension were the same

across language groups, and Grant et al. (2011) found a deviating pattern. In Grant et al. (2011), decoding predicted reading comprehension for both groups, and vocabulary predicted reading comprehension for *bilingual first children*. Interestingly, the results of Grant et al. (2012) are similar to those of several studies of *minority language learners* indicating differences in predictive patterns for reading comprehension across language groups (Burgoyne et al., 2011; Droop & Verhoeven, 2003; Geva & Farnia, 2012; Hutchinson et al., 2003; Limbird et al., 2014). These studies represent a mix of methods, including both studies using latent variables in SEM analysis and those using sum scores in multiple hierarchical regressions. The gap in the instructional language levels also varies across the samples of these studies. The gap varies from a smaller gap than was found in the present thesis to a larger gap than was identified in the *bilingual first* sample of Grant et al. (2011). Furthermore, the participants' age also varies. Hence, it is difficult to detect any reason why some studies of bilingual learners find equal predictive patterns (Babayiğit, 2015; Proctor et al., 2012), in line with the present study examining *early bilingual learners*, thereby indicating that the results could apply to bilingual learners in general, while others do not (Burgoyne et al., 2011; Geva & Farnia, 2012; Grant et al., 2011; Hutchinson et al., 2003; Limbird et al., 2014).

Due to the lack of studies of *early bilingual* pre-adolescents examining the prediction of reading comprehension, it is difficult to determine how well the results of the present study can be generalized to other samples of *early bilingual learners*. Furthermore, the mixed results of studies of *minority language learners* provide no support in detecting a pattern that could explain differences in results across studies regarding predictive patterns of reading comprehension.

Regarding the magnitude of the prediction of linguistic skills to reading comprehension, the present study finds the model parameters examined in Study 3 to be equal. This contradicts hypothesis 4 in this thesis. The hypothesis was based on prior research indicating that linguistic skills

seem to play a more central role in reading comprehension for *minority language learners* than for monolingual learners (Hutchinson et al., 2003; Kieffer, 2012b; Proctor et al., 2012; Silverman et al., 2015). This is also the general trend in the systematic review of Proctor and Louick (2018) examining the impact of vocabulary on reading comprehension, although they point out the need for meta-analytic work to determine with certainty that vocabulary is a stronger predictor of reading comprehension for bilingual learners than for monolingual learners. Hence, linguistic skills are likely a stronger predictor for reading comprehension for *early bilingual learners*. as well.

However, studies that have found that language is more crucial for *minority language learners* have often examined growth (e.g., Kieffer, 2012b; Proctor et al., 2012; Silverman et al., 2015). Hence, rather than differences in concurrent prediction, it is possible that linguistic skills play a superior role in the growth of reading comprehension for bilingual learners than for monolingual learners. An equal magnitude of concurrent prediction of linguistic skills for reading comprehension is supported by prior research using SEM modelling to test this relationship (Babayigit, 2015; Proctor et al., 2012). Such models have the advantages of testing whether different regression coefficients actually are significantly different from one another, which ordinary multiple regression and hierarchical regression cannot do (R. B. Kline, 2015). The results of the present thesis therefore conflict with hypothesis 4. The *early bilingual learners* in Study 3 do have lower language levels than their monolingual peers; however, this gap did not result in different magnitudes of the regression coefficients of linguistic skills to reading comprehension.

Regarding the predictive relationship of linguistic aspects to reading comprehension, listening comprehension had the largest impact on reading comprehension. Listening comprehension explained 26.01% of the variance in reading comprehension. This finding is in contrast to most other studies, which have identified vocabulary as the main variable to

predict L2 reading comprehension in *minority language learners* (Droop & Verhoeven, 2003; Hutchinson et al., 2003; Kieffer, 2012b; Leider et al., 2013; Proctor et al., 2012; Silverman et al., 2015). However, in the present study, the relationship between vocabulary and reading comprehension was not significant ( $p = 0.872$ ). Text cohesion vocabulary explained an additional 18.49% of the variance in reading comprehension. This is in line with Rydland et al. (2012), who also examined the reading comprehension of 5<sup>th</sup> grade bilingual learners. The results of the present study must be interpreted with caution. Despite the use of invariant latent variables, the dimensionality of the latent construct has not been tested. Hence, several of the latent variables might be better represented by merging them into one underlying language ability. In fact, most studies that have examined the specific impact of different linguistic subskills have not tested the dimensionality of the constructs used in regressions (e.g., Burgoyne et al., 2011; Geva & Farnia, 2012; Hutchinson et al., 2003; Leider et al., 2013; Rydland et al., 2012). Thus, little is known about whether some constructs truly explain more L2 reading comprehension than others. To truly examine this issue, it would be critical to examine whether a one-factor model, a correlated model and a bifactor model differed significantly from one another.

Taken together, the results of Study 3 do not support hypothesis 4 of this thesis.

## **5.5 Ethical reflections**

Over the years, the attitudes of Norwegians towards immigration have changed. The changes can be identified by a resurgence of right-wing anti-immigrant websites and journals, where hateful statements about immigrants flourish in the article comment fields (Skybakmoen, Klungtveit, Berg, & Nordseth, 2017). One of the many arguments against immigration is the cost of the welfare state funding facilities for immigrants. In addition to this hostile attitude towards immigrants is a difficult and polarized debate regarding the time spent assessing

children's abilities in ECEC institutions and schools (Essahli Vik, 2019; Gjervan & Svolsbru, 2010; Klem & Hagtvvet, 2018; Martinsen, 2016; Pettersvold & Østrem, 2018; Straume, 2015).

By presenting research that explicitly states that ECEC teachers should pay special attention to the language skills of bilingual children and suggests the need for more assessment and facilitation of bilingual children's language skills – both in ECEC institutions and in schools – this thesis indirectly contributes to this debate. Assessment costs time and thereby indirectly money. Furthermore, by providing information about long-term differences in language skills between monolingual and bilingual learners, this thesis risks not safeguarding the group investigated well enough. It could be argued that if the immigrant children most likely to succeed (due to high SES and massive exposure to Norwegian) have not caught up, then how much will unprivileged immigrants cost the Norwegian welfare state? Arguments such as this would create a hostile climate for immigrant children. This risk, most considered in light of the benefit this thesis could contribute.

Bilingual learners as a group, and specifically some subpopulations, are theorized to have EF skills superior to those of monolingual learners (Bialystok, 2017, 2018), which was the theme of examination in Study 1. The results of this study are unlikely to be misused. The reason is that the research findings presented here do not suggest more expenditure by the welfare state. However, this thesis does contribute new knowledge regarding bilingualism. Based on the meta-analysis, we can now assume that there are no differences between bilingual and monolingual learners in EF skills, which could be clinically important. The results of Studies 2 and 3 are more at risk of misuse, but the results could also benefit the investigated group. This thesis is the only study focusing on the long-term outcomes of *early bilingual learners* in a broad range of literacy skills and finds that *early bilingual learners* are outperformed by monolingual learners in toddlerhood and also in pre-adolescence. They did, however, perform within the range of normality according to the



norms for Norwegian children, although close to -1 SD for some linguistic aspects. This shows that even with massive exposure to Norwegian over a long period of time, the average *early bilingual* child's skills in the instructional language are not at the same level as the skills of his or her monolingual peers. In an evaluation of special needs, this knowledge is important, as well as the knowledge of which linguistic aspects are most likely to lag behind. This means that in a normal distribution of the L2 skills of *early bilingual learners* massively exposed to Norwegian, some of the children will score under the cut-off of normality for their native Norwegian-speaking peers in certain linguistic aspects. This must not be misread as an indicator of learning disabilities; rather, it is a sign that ECEC institutions and schools should facilitate their lessons in such a way that children can better benefit from the education provided for them. The model of Egeberg (2016) could then be applied to investigate children's further language development in light of the new and better-fitting educational facilitation. Therefore, in light of the knowledge this thesis contributes, I consider its clinical value to be greater than the risk of possible misuse of the results.

*Discussion*

---

## **6 Practical implications**

The thesis contributes additional knowledge of similarities and differences between bilingual and monolingual learners' skills in language, reading comprehension and executive function. The results indicate that bilingual learners have lower language levels than their monolingual peers but similar EF levels. The findings hold across different groups of bilingual learners and across most linguistic skills and EF domains. There is some support for a bilingual advantage in switching, but not for all groups of bilingual learners. *Early bilingual learners* have levels of text cohesion vocabulary similar to those of their monolingual peers.

When bilingual children's competence in each of their languages is investigated, the results show that bilingual children have lower levels of language skills in each of their languages than their monolingual peers (Bialystok, Luk, et al., 2010; Hammer, Lawrence, & Miccio, 2008; Oller, 2005). How the language development of bilingual children occurs is to a large degree influenced by the amount of input they receive in each of their languages. Some bilingual children can acquire language skills within the normal variation of monolingual children in the most stimulated language (Gatt, O'Toole, & Haman, 2015). This reason is, according to Hoff et al. (2012), the large variation in language skills within typically developed monolingual children. Depending on the language input received, some bilingual children could have equal developmental language curves in both their languages or a steep growth curve in one language combined with poor growth in the other language (David & Wei, 2008). After attendance in schools, and thus exposure to an academic language in L2, a shift in dominance from L1 to L2 is also common (Mancilla-Martinez & Lesaux, 2011b; Oller et al., 2007; Thurmann-Moe, Bjerkan, & Monsrud, 2012). Based on this, one can conclude that bilingual children's language skills in their first and second languages are varied. The variation in skills within a group of bilingual

learners is often larger than the variation within a group of monolingual learners (Barne- likestillings- og inkluderingsdepartementet, 2012). Hence, the normal language development of bilingual children at the group level is hard to define. This lack of homogeneity is seconded by Genesee, Lindholm-Leary, Saunders, and Christian's (2006) synthesis of research evidence of English-language learners. The synthesis points out differences between subpopulations of bilingual learners and urges that bilingual student diversity warrants increased attention in future research. Knowledge of normally developed skills for different subgroups of bilingual learners is therefore of great importance. This thesis presents knowledge of the language skills of *early bilingual learners* mainly from medium- to high-SES backgrounds, a bilingual subgroup that is theorized to have the utmost possibility of achieving language levels similar to those of its monolingual peers. *Early bilingual learners* are an understudied subgroup of bilingual learners, especially concerning their long-term language outcomes and the relations of different linguistic aspects and decoding skills to L2 reading comprehension. The results of this thesis show that *early bilingual learners* lag behind their peers at different ages and in most linguistic aspects. The difference in effect size varies from one linguistic aspect to another. However, after a minimum of 8 years of exposure, the gap across linguistic aspects is  $d = .53$ , ranging from levels of text cohesion vocabulary similar to those of their monolingual peers to a medium to large gap in reading comprehension and vocabulary ( $d = 0.78$  and  $0.74$ , respectively). However, the predictive strength of linguistic skills is similar across language groups. Furthermore, in toddlerhood, differences in language skills across different groups of bilingual learners were detected. Bilingual children with primary L1 input at home had lower second language skills than bilingual learners with some exposure to the instructional language at home and bilingual learners exposed mainly to the instructional language at home. Language skills in the instructional language in the latter two groups of bilingual learners were interchangeable, implying that language input in the instructional

language above a certain threshold might be enough for them to develop good second language skills. Such knowledge is important when following the advice of Egeberg (2016) to always bear in mind the learning experiences bilingual children have taken part in when evaluating their learning outcomes.

This thesis also contributes knowledge that could be clinically important when assessing bilingual children who have been less exposed to L2 than *early bilingual learners*. We know that typical language-developed bilingual learners with limited knowledge of L2 show language profiles similar to those of their monolingual peers with language disorders concerning morphosyntax, vocabulary levels and narrative skills (Armon-Lotem, 2018; Paradis, 2010). In such cases, the need for a thorough L1 assessment is urgent (Bishop, Snowling, Thompson, Greenhalgh, & the Catalise-2 consortium, 2017). However, there is often a lack of available diagnostic tests to examine bilingual children's L1 skills (Tuller, 2015), and if such tests are available, they are often normed for monolingual children in this language rather than bilingual children (Huang, Fang Kan, & Fang Kan, 2019). Tests normed for monolingual learners are not suitable for evaluating the size of a potential language gap in the average bilingual child (Gatt et al., 2015; Schelletter, 2019). In such cases, the bilingual child often ends up with an evaluation of his or her language skills by diagnostic tests in his or her second language only. When the possibilities of assessing bilingual learners' L1 skills with formal tests are scarce, mapping the child's history of L1 development, input and exposure is recommended, as well as an indirect assessment of the child's L1 mastery through parental interviews (Egeberg, 2016; Paradis, Emmerzael, & Duncan, 2010; Tuller, 2015). Although such interviews provide useful information about L1 functioning, conclusions about language skills based solely on parental interviews could lead to over- and under-identification of children with language delays (Grimm & Schulz, 2014; Paradis et al., 2010). Assessment of the bilingual child's executive function could then be

argued to be an important supplement in order to determine whether a language disorder exists. The reason is that EF deficits are thought to cause difficulties in “uptake” of linguistic input and are theorized to partly or wholly underlie the language development of children with language impairments (Paradis & Govindarajan, 2018). This theory is supported by several empirical studies that find that EF skills are associated with many kinds of learning disabilities as well as literacy and language skills (Archibald & Gathercole, 2006; Potocki, Sanchez, Ecalle, & Magnan, 2017). The relationship between executive function and language disabilities has been extensively studied through several meta-analyses. The results of such meta-analyses show that children with specific learning impairments have a large deficit in non-word repetition (1.27 *SD*). Some researchers claim that non-word repetition taps working memory (Grundy & Timmer, 2016), while others find non-word repetition to be more closely related to language skills (Melby-Lervåg et al., 2012). Given the close relationship between non-word repetition and language skills, it is not surprising that children with specific language impairments have lower levels of non-word repetition than their typical developed (TD) peers (Estes, Evans, & Else-Quest, 2007). However, another meta-analysis targeting visual working memory in children with specific language disorders finds a gap in visual working memory compared with their TD peers (Vugs, Cuperus, Hendriks, & Verhoeven, 2013), although the gap was only of a moderate effect size ( $g = 0.63$ ). Gaps in other EF domains, such as the domains of attention, inhibition and switching, have also been detected. Through meta-analysis, children with language impairments have been identified as exhibiting lower levels of sustained attention ( $g = 0.69$ ), inhibition ( $g = 0.56$ ) and switching skills ( $g = 0.27$ ) than their TD peers (Ebert & Kohnert, 2011; Pauls & Archibald, 2016). Altogether, there are grounds for believing that children with language impairments have lower levels of EF. How does this affect bilingual children with language disorders? Bilingual learners as a group are thought to have enhanced EF skills (Adesope et al., 2010; Bialystok, 2018). If test norms of monolingual learners are unsuitable for

evaluating the size of a potential language gap in bilingual children (Schelletter, 2019), are norms of monolingual learners' EF skills fit for evaluating the size of a bilingual child's gap in EF? Given the support for a connection between EF and language disorders (Ebert & Kohnert, 2011; Pauls & Archibald, 2016; Vugs et al., 2013), such questions could become important when assessing a bilingual child with EF skills that fall just within the normal variation of a monolingual child. Therefore, from this perspective, the results of the meta-analysis of this thesis also contribute to knowledge that could be clinically important when assessing subgroups other than *early bilingual learners*. Since this meta-analysis found similar levels of EF skills across overall EF and for most domains of EF, it seems unnecessary to develop special norms to assess bilingual children's EF skills.

There is still a question regarding how to interpret the switching skills of bilingual children who are suspected to suffer from language disabilities. Whereas children with language disabilities have somewhat lower levels of switching than typical language developed children, equivalent to  $d = 0.27$  (Pauls & Archibald, 2016), the meta-analysis provided some support for a bilingual advantage in switching ( $d = 0.27$ ). However, based on the observed variation in switching, not all bilingual groups have an advantage in switching, and at this moment, it is not clear whether it should be interpreted as an advantage at the construct level. Therefore, it is too early to state whether there is a need for more finely tuned cut-off norms for bilingual learners in this area. However, importantly, with the exception of switching, one can now assume that a bilingual child with EF skills within the lower levels of distribution of monolingual norms has normal EF skills. This cannot be read as a symptom of language disorders or any other type of disorder.

Assessment of learning disabilities in bilingual children is generally complicated, which has resulted in over- and under-identification of bilingual children with learning disabilities within the education system (Geva & Wiener, 2015). Although bilingual learners are over-

represented in special pedagogics education, there is no reason to believe that they as a group are more prone to learning disability than monolingual learners (Egeberg, 2016). In fact, bilingual children have over decades been misdiagnosed as mentally retarded by the Educational Psychological Services after assessment of WISC testing in their second language (Jim Cummins, 1984; Pihl, 2010). The problems with separating typical bilingual development and learning difficulties (also more minor than mental retardation) could thereby lead to bilingual learners incorrectly receiving special education (Geva, 2000). Hence, more information on similarities and differences in language, reading comprehension and EF for specific subgroups of bilingual and monolingual children is important. The knowledge this thesis contributes could be useful for Educational Psychological Services employees, teachers and politicians. The knowledge of typical development of executive functions in bilingual learners as well as information about *early bilingual* language skills could work as clinical guidelines for employees in Educational Psychological Services when assessing and interpreting bilingual children's test results. Thus, the knowledge could potentially be helpful in reducing over- and under-identification of bilingual children with learning disabilities. It also provides grounds for teachers and politicians to understand how complex second language development truly is. When *early bilingual learners* from medium to high SES on average have not caught up with their monolingual peers within 8 years of exposure to L2, the average bilingual learner with a later AoA most likely will not have done so either. There are therefore grounds for believing that bilingual children in general, perhaps excepting *bilingual first* children, have a long-term need for L2 support.



## **7 Limitations**

There are several limitations for the results of the present thesis. Some of the limitations apply to the whole thesis; others are directly related to particular articles or the additional analysis in Appendixes 4.

Regarding limitations that apply to the whole thesis, the operationalization of the term bilingual learners is perhaps the most prominent one. Bilingualism has in all three articles been treated as a categorical term. There are some attempts to test different dimensions of bilingual experience in Study 1, but the dimensions are tested individually and not as a combination of different dimensions as a true continuous dimension (e.g., Bialystok & Barac, 2012; De Cat, 2020; Luk & Bialystok, 2013). Studies 2 and 3 relied on comparisons of different bilingual subgroups against a monolingual control group. Such comparisons could be problematic since it is difficult to design non-overlapping groups (Genesee, 2010; Luk & Bialystok, 2013). Hence, the use of bilingualism as a continuous variable could have led to the creation of different and more suitable subgroups of bilingual learners. It would also have been possible to use the continuous variable in a regression to determine how the degree of bilingualism was related to skills in a combined group of monolingual and bilingual learners. Thus, if bilingualism were measured as a continuous variable, the conclusions regarding the examined hypotheses of the present thesis could have differed.

Since bilingualism was treated as a categorical variable in this thesis, more characteristics of the samples in Studies 2 and 3 would have been preferable. As mentioned in the discussion of the results of Study 2, more knowledge of the bilingual parents' nationality would have strengthened the paper. In the discussion of Studies 2 and 3 in this thesis, I have assumed that bilingual learners with a native majority language-speaking parent growing up in Norway, with both languages being used at home

in toddlerhood, are *bilingual first learners*. It would have improved the thesis if we had indeed asked the parents about the precise AoA of their bilingual child. The method of sample selection in The Stavanger Project also limits the conclusions of the hypotheses examined in this thesis. The Stavanger Project was designed to examine the development of skills such as social functioning reading and language in all children. Hence, there were no exclusion criteria for children displaying atypical development. This could be a validity challenge to the group comparison studies conducted in articles 2 and 3. One way to reduce this threat would have been to ensure that the compared groups were equal on all other third factors that could influence the examined factor (in this case language level). Hence, group performance across the examined groups is more likely caused by the grouping factor (here monolingual or bilingual status) than by a hidden third factor. The lack of exclusion of children with atypical development in The Stavanger Project is therefore not optimal. There is a chance that children with atypical development are unevenly distributed across the different groups examined in the different articles. An uneven distribution of potential learning disability across groups could then work as a hidden third factor influencing the language levels of the monolingual or bilingual group beyond what can be expected on the basis of monolingual and bilingual status. This is particularly problematic for the results of Study 2 and the analysis in Appendix 4. In Study 3, the children were older; hence, signs of learning disabilities would most likely have emerged. When invitations to participate in Study 3 were sent, some parents returned a consent form stating that their child unfortunately could not take part in it. The parents then explicitly explained that although they valued the aims of the project, their child was not a good candidate for the project since the child had been diagnosed with learning disabilities. The child would therefore not be representative of typical developed monolingual or bilingual children. It is not clear whether all parents of children with learning disabilities provided such information. The type of assessment used in Study 3 provided, however, additional support in excluding

children who had obvious learning disabilities. The language and reading comprehension levels of all participants were individually evaluated by teachers with long experience with children. The completion of the individual tests took 1 hour; hence, it was possible to detect age-inappropriate behaviour. In one case, a bilingual child was unable to finish the tests, and his test scores were then excluded from the study. Even though such an evaluation of the children was performed, it is important to acknowledge that the parents of participants in Study 3 were not specifically asked whether their child had learning disabilities. Thus, there is still a possibility that an uneven distribution of learning disability between groups influenced the results of the analysis in Study 3.

Another important limitation is the lack of a systematic search for relevant articles regarding the hypothesis of poorer language levels. Such a search, followed by analysis of the data in a meta-analysis, would have strengthened the quality of this thesis. Even though the time spent searching for relevant articles was extensive (articles about pre-adolescent *minority language learners* with 4-7 years of L2 exposure, pre-adolescent *early bilingual learners*, pre-adolescent *bilingual first learners*), a systematic search could have identified a range of studies that contradict the results of the identified studies and the present analysis. Hence, in coding and analysing all relevant studies, the conclusions regarding these hypotheses could potentially differ from what was possible based on a narrative review. Furthermore, a meta-analytic approach would have enabled us to form a conclusion regarding to what extent the levels of reading comprehension and language of *early bilingual learners* resemble the levels of *minority language learners*, or their monolingual peers. Casting light on this problem is beyond the scope of for the present approach.

Regarding the measurement used to examine language comprehension in Study 2, the results could have been hampered by comparing groups across sum scores. It would have strengthened the validity of the results if data from all the TRAS sections were examined by explorative and

confirmatory factor analysis, followed by securing an invariant construct across the bilingual and monolingual groups. Such an approach, instead of a theory-driven approach, would have allowed for comparisons of true scores. As it stands now, one cannot be sure how much of the observed differences across language groups could be ascribed to comparisons across different constructs, or even whether some items in the measurement instrument are biased for one language group despite otherwise equal language levels. The true differences in levels of language comprehension might therefore be either larger or smaller than the observed differences identified in article 2.

For Study 3, as with many other invariance studies, there is a limitation regarding these analysis in the article. The chosen procedure in article 3 relies on an assumption of a normal distributed latent variable underlying the ordered categorical items. Whether these variables in study 3 indeed are normally distributed is however not empirically investigated. If the latent response variates that underly the categorical variables are not multivariate normal, invariance in the thresholds and factor model parameters cannot guaranty measurement invariants for the measured variables (Millsap & Yun- Tein 2010). Another limitation is the sample size in this article. A larger sample size would have been preferable for many reasons. This would have allowed a more complicated model, either testing the dimensions in constructs or examining a bifactor model. These more complicated models would have enabled an examination of the individual contributions of different linguistic constructs to reading comprehension, controlled for the common underlying language ability. Such knowledge would be preferable when designing intervention studies for weak monolingual and bilingual readers.

Finally, there is a limitation regarding how well the different papers address the overarching hypothesis of poorer language levels in bilingual than in monolingual learners. The articles themselves could have provided a better examination of how well poorer language levels can be generalized across different groups of bilingual learners. In hindsight, it

### *Limitations*

---

would have been preferable for article 2, research question 2 to compare the different bilingual subgroups to the monolingual control group instead of solely comparing the levels of the three different bilingual subgroups. The reason for this limitation is that at the stage in the process where article 2 took shape, the overarching aims of the thesis were not fully developed. The overarching aims thus could not work as guidance for the formation of article 2. (Despite being labelled article 2, it was the first completed article of the thesis.) Thus, in order to fully address this overarching hypothesis, in article 2, the language levels of a composite group of bilingual learners exposed to some or mainly L2 at home are compared to the language levels of monolingual learners. The results of this analysis are presented in Appendix 4.

*Limitations*

---

## **8 Recommendations for future research**

Each individual article presents suggestions for how to move the field forwards by addressing the research questions in those papers. The study-specific suggestions will not be mentioned here again; thus, recommendations for future research in this section are limited to areas of research that to a larger extent address the hypotheses of this thesis.

First, Hammer et al. (2014) call for more detailed descriptions of the sample characteristics of bilingual learners in primary studies. They state that this is necessary to be able to draw conclusions across studies. A better way to enable comparisons of results across studies would be to develop a universal standard for how to measure and report bilingualism. Such a standard would make comparisons across studies more fruitful. Such a standard should aim to achieve consensus on how to operationalize bilingualism, preferably as a continuous variable. At present, different studies operationalize bilingualism as a continuous variable somewhat differently (e.g., Bialystok & Barac, 2012; De Cat, 2020; Luk & Bialystok, 2013).

Second, there is a need for more research on a possible bilingual advantage in switching for children. A determination of whether this is an advantage at the construct level would move the field forwards. Due to construct validity threats, such a possible bilingual advantage in switching should be examined using latent variables. This would strengthen the possibility of determining whether this is an advantage at the construct level. In addition to the use of latent variables, upcoming studies should report the means and SD scores of bilingual and monolingual learners on the different switching tests. Most prior studies have used the DCCD test, which resulted in an unbalanced task moderator in article 1. The statistical power of unbalanced moderator analysis in RVE statistics is surprisingly low; hence, it is preferable for future studies to assess switching with tests other than the DCCD. This

would cast additional light on the extent to which an advantage in switching is task dependent. Furthermore, in order to address the theory beyond a bilingual advantage in EF, there is a need to assess the relationship between typical bilingual language experiences and switching. Due to a lack of studies reporting the frequencies of bilingual participants switching between L1 and L2, this particular dimension of bilingualism was not examined in this thesis. To ground in theory an empirically supported rationale for a bilingual advantage in switching, new primary studies of a bilingual advantage in switching should include measures of bilingual learners' frequency of daily language switching. However, perhaps more importantly, before one addresses the possible advantage in switching, there is a need to empirically determine what EF actually is as well as what separates the different EF domains. If this is not addressed, the same threats to construct validity that existed in this meta-analysis will apply to new primary studies and meta-analyses examining the alleged bilingual advantage in switching.

Studies 2 and 3 find that *early bilingual learners* lag behind their peers in toddlerhood, and the gap is not closed in 5<sup>th</sup> grade. This shows that bilingual learners are in need of more support than what they generally receive just by being enrolled for a certain amount of time in ECEC institutions and school. It seems that their need for long-term language support is not met. Since this need is not met for *early bilingual learners* with medium to high SES, the need is most likely not met for *minority language learners* either. There could be many reasons for this. One reason is that bilingual children might lose their right to extra support to develop proficient L2 skills too early to receive the full benefit of ordinary education. US studies find large variations in criteria for determining when L2 proficiency is sufficiently good to rely on ordinary education only (Abedi, 2008; Umansky & Reardon, 2014), in combination with findings that *minority language learners* lose their L2 support at an earlier timepoint than the timeframe it takes to develop CALP (Slama, 2014). There are no such studies examining how long



*minority language learners* keep their L2 support in Norwegian schools nor an evaluation of how Norwegian *minority language learners* then score on normative tests. The *early bilingual learners* in the present thesis have a minimum of 4 years of L2 exposure before school entry, and 2/3 have even been exposed to L2 from birth. When these children still have lower language and reading comprehension skills than their monolingual peers, there is a need to address the way teachers educate children with low language skills. The way to improve their academic results seems to lie in their language skills. Instructional programmes invented to support language skills do not seem to have an effect on reading comprehension of general texts (Rogde, Hagen, Melby-Lervåg, & Lervåg, 2019). Hence, there is a need to develop interventions targeting all bilingual (and monolingual) children with low levels of language skills in the instructional language.

*Recommendations for future research*

---

## References

- Abedi, J. (2008). Classification System for English Language Learners: Issues and Recommendations. *Educational Measurement: Issues and Practice*, 27(3), 17-31.
- Aben, B., Stapert, S., & Blokland, A. (2012). About the Distinction between Working Memory and Short-Term Memory. *Frontiers in Psychology*, 3, 301-301.
- Abrahamsson, N., & Hyltenstam, K. (2009). Age of onset and nativelikeness in a second language: Listener perception versus linguistic scrutiny. *Language learning*, 59(2), 249-306.
- Adesope, O. O., Lavin, T., Thompson, T., & Ungerleider, C. (2010). A Systematic Review and Meta-Analysis of the Cognitive Correlates of Bilingualism. *Review of Educational Research*, 80(2), 207-245.
- Alloway, T. P., Gathercole, S. E., Willis, C., & Adams, A.-M. (2004). A structural analysis of working memory and related cognitive skills in young children. *Journal of Experimental Child Psychology*, 87(2), 85-106.
- Anderson, R. C., & Freebody, P. (1981). Vocabulary knowledge. In *Comprehension and teaching: Research reviews* (pp. 77-117). Newark, DE: International Reading association.
- Archibald, & Gathercole. (2006). Short term and working memory in specific language impairment. *International Journal of Communication Disorder*, Vol 41(No. 6), s. 675- 693.
- Armitage, S. G. (1946). An analysis of certain psychological tests used for the evaluation of brain injury. *Psychological monographs*, 60(1).
- Armon-Lotem, S. (2018). SLI in bilingual development: How do we approach assessment? In A. Bar-On, D. Ravid, & E. Dattner (Eds.), *Handbook of Communications Disorders: Theoretical, Empirical, and Applied Linguistic Perspectives*. Boston, United States: De Gruyter, Inc.
- Artiles, A. J., & Ortiz, A. A. (2002). *English Language Learners With Special Education Needs. Identification, Assessment and Instruction*: Centre for Applied Linguistics, Delta Systems Co. Inc

## References

---

- August, D., & Shanahan, T. (2006). *Developing literacy in second-language learners: report of the National Literacy Panel on Language Minority Children and Youth*. Mahwah, N.J: Lawrence Erlbaum.
- Babayigit, 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.
- Baddely, S. D., & Hitch, G. I. (1986). *Working memory*. New York: Oxford University Press.
- Baggetta, P., & Alexander, P. A. (2016). Conceptualization and Operationalization of Executive Function. *Mind, Brain, and Education*, 10(1), 10-33.
- Bakken, A. (2003). *Minoritetsspråklig ungdom i skolen: reproduksjon av ulikhet eller sosial mobilitet?* (Vol. 15/2003).
- Bakken, A., & Elstad, J. I. (2012). *For store forventninger?: Kunnskapsløftet og ulikhetene i grunnskolekarakterer* (Vol. 7/2012).
- 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.
- Barac, R., Bialystok, E., Castro, D. C., & Sanchez, M. (2014). The cognitive development of young dual language learners: A critical review. *Early Childhood Research Quarterly*, 29(4), 699-714.
- Barne- likestillings- og inkluderingsdepartementet. (2012). *Meld. st 6. En helhetlig integreringspolitikk*.
- Barrena, A., Ezeizabarrena, M. J., & Garcia, I. (2008). Influence of the Linguistic Environment on the Development of the Lexicon and Grammar of Basque Bilingual Children. In C. P. Vidal, M. Juan-Garau, & A. Bel (Eds.), *A portrait of the young in the new multilingual Spain*. Clevedon, Buffalo, Toronto Multilingual Matters LTD.
- Bellocchi, S., Tobia, V., & Bonifacci, P. (2017). Predictors of reading and comprehension abilities in bilingual and monolingual children: a longitudinal study on a transparent language. *Reading and Writing*, 30(6), 1311-1334.

## References

---

- Berko, J. (1958). The child's learning of English morphology. *Word*, (14), 150-177.
- Berthiaume, R., Daigle, D., & Desrochers, A. (2018). *Morphological Processing and Literacy Development: Current Issues and Research*: Routledge.
- Best, J. R., & Miller, P. H. (2010). A developmental perspective on executive function. *Child Development*, 81(6), 1641-1660
- Best, J. R., Miller, P. H., & Naglieri, J. A. (2011). Relations between executive function and academic achievement from ages 5 to 17 in a large, representative national sample. *Learning and Individual Differences*, 21(4), 327-336.
- Bialystok, E. (2001a). *Bilingualism in development: Language, literacy, and cognition*: Cambridge University Press.
- Bialystok, E. (2001b). Metalinguistic aspects of bilingual processing. *Annual review of applied linguistics*, 21, 169-181.
- Bialystok, E. (2009). Bilingualism: The good, the bad, and the indifferent. *Bilingualism: Language and Cognition*, 12(1), 3-11.
- Bialystok, E. (2010). Global-Local and Trail-Making Tasks by Monolingual and Bilingual Children: Beyond Inhibition. *Developmental Psychology*, 46(1), 93-105.
- Bialystok, E. (2017a). The bilingual adaptation: How minds accommodate experience. *Psychological Bulletin*, 143(3), 233-262.
- Bialystok, E. (2018). Bilingualism and executive functions: What is the connection? In D. Miller, F. Bayram, J. Rothman, & L. Serratrice (Eds.), *Bilingual Cognition and Language* (pp. 283-306). Amsterdam/ Philadelphia: John Benjamins Publishing Company.
- Bialystok, E., & Barac, R. (2012). Emerging bilingualism: Dissociating advantages for metalinguistic awareness and executive control. *Cognition*, 122(1), 67-73.
- Bialystok, E., Barac, R., Blaye, A., & Poulin-Dubois, D. (2010). Word mapping and executive functioning in young monolingual and bilingual children. *Journal of Cognition and Development*, 11(4), 485-508.
- Bialystok, E., Craik, F. I., Binns, M. A., Osher, L., & Freedman, M. (2014). Effects of bilingualism on the age of onset and progression of MCI and AD: Evidence from executive function tests. *Neuropsychology*, 28(2), 290.

## References

---

- Bialystok, E., Craik, F. I. M., Green, D. W., & Gollan, T. H. (2009). Bilingual Minds. *Psychological Science in the Public Interest*, 10(3), 89-129.
- Bialystok, E., & Feng, X. (2011). Language proficiency 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). New York: The Guildford Press.
- Bialystok, E., Luk, G., Peets, K., F., & Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism-Language and Cognition*, 13(4), 525-531.
- Bialystok, E., Luk, G., & Kwan, E. (2005). Bilingualism, Biliteracy, and Learning to Read: Interactions Among Languages and Writing Systems. *Scientific Studies of Reading*, 9(1), 43-61.
- Bialystok, E., Peets, K. F., & Moreno, S. (2014). Producing bilinguals through immersion education: Development of metalinguistic awareness. *Applied Psycholinguistics*, 35(1), 177-191.
- Bialystok, E., & Viswanathan, M. (2009). Components of executive control with advantages for bilingual children in two cultures. *Cognition*, 112(3), 494-500.
- Biedroń, A., & Birdsong, D. (2019). Highly proficient and gifted bilinguals. In A. De Houwer & L. Ortega (Eds.), *The Cambridge Handbook of Bilingualism* (pp. 307-0323): Cambridge University Press.
- Biemiller, A., & Slonim, N. (2001). Estimating root word vocabulary growth in normative and advantaged populations: Evidence for a common sequence of vocabulary acquisition. *Journal of Educational Psychology*, 93(3), 498.
- Birdsong, D. (2005). Interpreting age effects in second language acquisition. *Handbook of bilingualism: Psycholinguistic approaches*, 109, 127.
- Birdsong, D. (2006). Age and second language acquisition and processing: A selective overview. *Language learning*, 56, 9-49.
- Bishop, D. V. M. (2003a). *The children's communication checklist-CCC-2* (2.ed.). London, England: Psychological Corporation.
- Bishop, D. V. M. (2003b). *Test for Reception of Grammar. Version 2. (TROG -2) Manual*. London: Harcourt Assessment.

## References

---

- Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T., & the CATALISE-2 consortium. (2017). Phase 2 of CATALISE: a multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *Journal of Child Psychology and Psychiatry*, 58(10), 1068-1080.
- Bjerkan, K. M., Monsrud, M.-B., & Thurmann-Moe, A. C. (2013). *Ordforråd hos flerspråklige barn: pedagogiske og spesialpedagogiske utfordringer*. Oslo: Gyldendal akademisk.
- Black, M. M., Walker, S. P., Fernald, L. C. H., Andersen, C. T., DiGirolamo, A. M., Lu, C., . . . Grantham-McGregor, S. (2017). Early childhood development coming of age: science through the life course. *The Lancet*, 389(10064), 77-90.
- Blair, C. (2016). Developmental science and executive function. *Current directions in psychological science*, 25(1), 3-7.
- Blair, C., & Raver, C. C. (2015). School readiness and self-regulation: A developmental psychobiological approach. *Annual Review of Psychology*, 66, 711-731.
- Blom, E., & Bosma, E. (2016). The sooner the better? An investigation into the role of age of onset and its relation with transfer and exposure in bilingual Frisian–Dutch children. *Journal of Child Language*, 43(3), 581-607.
- Bohman, T. M., Bedore, L. M., Peña, E. D., Mendez-Perez, A., & Gillam, R. B. (2010). What you hear and what you say: Language performance in Spanish–English bilinguals. *International Journal of Bilingual Education and Bilingualism*, 13(3), 325-344.
- 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.
- Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2011). *Introduction to meta-analysis*: John Wiley & Sons.
- Borenstein, M., Rothstein, D., & Cohen, J. (2005). *Comprehensive meta-analysis: A computer program for research synthesis* [computer software]. Englewood, NJ: Biostat.
- Bornstein, M. H., & Bradley, R. H. (2014). *Socioeconomic status, parenting, and child development*: Routledge.

## References

---

- Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *108*(3), 624.
- Bowers, E. P., & Vasilyeva, M. (2011). The relation between teacher input and lexical growth of pre-schoolers. *Applied Psycholinguistics*, *32*(1), 221-241.
- Brennan, R. L. (2010). Generalizability Theory and Classical Test Theory. *Applied Measurement in Education*, *24*(1), 1-21.
- Bridges, K., & Hoff, E. (2014). Older sibling influences on the language environment and language development of toddlers in bilingual homes. *Applied Psycholinguistics*, *35*(2), 225-241.
- 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.
- Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. New York: Guilford Publications.
- Brydges, C. R., Reid, C. L., Fox, A. M., & Anderson, M. J. I. (2012). A unitary executive function predicts intelligence in children. *40*(5), 458-469.
- Bull, R., Espy, K. A., & Wiebe, S. A. (2008). Short-term memory, working memory, and executive functioning in preschoolers: Longitudinal predictors of mathematical achievement at age 7 years. *Developmental neuropsychology*, *33*(3), 205-228.
- 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.
- Cain, K., Patson, N., & Andrews, L. (2005). Age-and ability-related differences in young readers' use of conjunctions. *Journal of Child Language*, *32*(4), 877-892.
- Capps, R., Fix, M., Murray, J., Ost, J., Passel, J. S., & Herwanto, S. (2005). The new demography of America's schools: Immigration and the No Child Left Behind Act. *Urban Institute (NJI)*.
- Carlisle, J. F. (2000). Awareness of the structure and meaning of morphologically complex words: Impact on reading. *Reading and writing*, *12*(3), 169-190.



## References

---

- Carlisle, J. F., & Feldman, L. (1995). Morphological awareness and early reading achievement. In Feldman (Ed) *Morphological aspects of language processing* (pp. 189-210) Hillsdale, New Jersey, Hove, UK: Lawrence Erlbaum Associates, Publishers.
- Carroll, J. B. (1963). A model of school learning. *Teachers college record*, 64(8), 723-733.
- Chen, X., Ramirez, G., Luo, Y., Geva, E., & Ku, Y.-M. (2012). Comparing vocabulary development in Spanish- and Chinese-speaking ELLs: the effects of metalinguistic and sociocultural factors. *Reading and Writing*, 25(8), 1991-2020.
- Chondrogianni, V. (2018). Child L2 acquisition. In D. Miller, F. Bayram, J. Rothman, & L. Serratrice (Eds.), *Bilingual Cognition and Language. The state of the science across its subfields*. (pp. 103-126). Amsterdam, Philadelphia: John Benjamins Publishing Company.
- Chondrogianni, V., & Marinis, T. (2011). Differential effects of internal and external factors on the development of vocabulary, tense morphology and morpho-syntax in successive bilingual children. *Linguistic Approaches to Bilingualism*, 1(3), 318-345.
- Cirino, P. T., Ahmed, Y., Miciak, J., Taylor, W. P., Gerst, E. H., & Barnes, M. A. (2018). A framework for executive function in the late elementary years. *Neuropsychology*, 32(2), 176.
- Collier, V. P. (1987). Age and Rate of Acquisition of Second Language for Academic Purposes. *TESOL quarterly*, 21(4), 617-641.
- Collier, V. P. (1989). How long? A synthesis of research on academic achievement in a second language. *TESOL quarterly*, 23(3), 509-531.
- Commodari, E. (2017). Novice readers: the role of focused, selective, distributed and alternating attention at the first year of the academic curriculum. *i-Perception*, 8(4), 1-18.
- Craik, F. I., Bialystok, E., & Freedman, M. (2010). Delaying the onset of Alzheimer disease: Bilingualism as a form of cognitive reserve. *Neurology*, 75(19), 1726-1729.
- Crosson, A. C., Lesauc, N. K., & Martiniello, M. (2008). Factors that influence comprehension of connectives among language minority children from Spanish-speaking backgrounds. *Applied Psycholinguistics*, 29, 603-625.

## References

---

- Cummins, J. (1978). Bilingualism and the development of metalinguistic awareness. *Journal of Cross-Cultural Psychology*, 9(2), 131-149.
- Cummins, J. (1979). Linguistic Interdependence and the Educational Development of Bilingual Children. *Education & Educational Research, spring* 49(2), 222-251.
- Cummins, J. (1984). *Bilingualism and special education. issues in assessment and pedagogy*. Clevedon: Multilingual Matters.
- Cummins, J. (1987). Bilingualism Language Proficiency, and Metalinguistic Development. In P. Homel, M. Palić, & D. Aronson (Eds.), *Childhood bilingualism: Aspects of Linguistic Cognitive and Social development* (pp. 57-74). Hillsdale, New Jersey. London: Lawrence Erlbaum Associates, Publishers.
- Cummins, J. (2017). *Flerspråkiga elever: effektiv undervisning i en utmanande tid*: Natur & Kultur Akademisk.
- David, A., & Wei, L. (2008). Individual differences in the lexical development of French-English bilingual children. *International Journal of Bilingual Education and Bilingualism*, 11(5), 598-618.
- Davis-Kean, P. E. (2005). The Influence of Parent Education and Family Income on Child Achievement: The Indirect Role of Parental Expectations and the Home Environment. *Journal of Family Psychology*, 19(2), 294-304.
- de Bruin, A., & Sala, S. D. (2018). Effects of age on inhibitory control are affected by task-specific features. *Quarterly Journal of Experimental Psychology*, 71(5), 1219-1233.
- de Bruin, A., Treccani, B., & Della Sala, S. (2015). Cognitive advantage in bilingualism: an example of publication bias? *Psychological Science*, 26(1), 99-107.
- De Cat, C. (2020). Predicting language proficiency in bilingual children. *Studies in Second Language Acquisition*, 42(2), 279-325.
- De Cat, C., Gusnanto, A., & Serratrice, L. (2018). Identifying a threshold for the executive function advantage in bilingual children. *Studies in Second Language Acquisition*, 40(1), 119-151.
- De Houwer, A. (2009a). *Bilingual first language acquisition*: Multilingual Matters

## References

---

- De Houwer, A. (2009b). *An introduction to bilingual development: Multilingual Matters*.
- De Houwer, A. (2012). Milestones in bilingual children's development. In *Encyclopaedia of Language and Literacy Development* (pp. 1-8). London, Canada: Western Ontario University.
- De Houwer, A. (2018). Input, context and early child bilingualism: Implications for clinical practice. In A. Bar-On, D. Ravid, & E. Dattner (Eds.), *Handbook of Communications Disorders: Theoretical, Empirical, and Applied Linguistic Perspectives*. Boston, United States: De Gruyter, Inc.
- De Houwer, A., Bornstein, M. H., & Putnick, D. L. (2014). A bilingual–monolingual comparison of young children's vocabulary size: Evidence from comprehension and production. *Applied Psycholinguistics*, 35(6), 1189-1211.
- de Zeeuw, E. L., de Geus, E. J., & Boomsma, D. I. (2015). Meta-analysis of twin studies highlights the importance of genetic variation in primary school educational achievement. *Trends in Neuroscience and Education*, 4(3), 69-76.
- DeKeyser, R. M. (2013). Age effects in second language learning: Steppingstones toward better understanding. *Language learning*, 63, 52-67.
- DeKeyser, R. M., & Larson-Hall, J. (2005). What does the critical period really mean? *Handbook of bilingualism: Psycholinguistic approaches*, 88, 108.
- DeVellis, R. F. (2006). Classical test theory. *Medical Care*, S50-S59.
- Diamond, A. (2013). Executive Functions. *Annual Review of Psychology* (64), 135- 168.
- Diamond, A. (2016). Why improving and assessing executive functions early in life is critical. In J. A. Griffin, P. McCardle, & L. S. Freund (Eds.), *Executive Function in Preschool- Age Children: Integrating Measurement, Neurodevelopment, and Translational Research: The American Psychological Association*.
- Donnelly, S. (2016). *Re-examining the bilingual advantage on interference-control and task-switching tasks: A meta-analysis*. (10009829 Ph.D.). City University of New York, Ann Arbor.
- Donnelly, S., Brooks, P. J., & Homer, B. D. (2019). Is there a bilingual advantage on interference-control tasks? A multiverse meta-

## References

---

- analysis of global reaction time and interference cost. *Psychonomic Bulletin & Review*, 26(4), 1122-1147.
- Droop, M., & Verhoeven, L. (2003). Language proficiency and reading ability in first- and second-language learners. *Reading Research Quarterly*, 38(1), 78-103.
- Duggan, E. C., & Garcia-Barrera, M. A. (2015). Executive functioning and Intelligence. In S. Goldstein, J. A. Naglieri, & D. Princiotta (Eds.), *Handbook of intelligence: Evolutionary theory, historical perspective, and current concepts* (pp. 435-458). New York: NY: Springer.
- Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., . . . Brooks-Gunn, J. (2007). School readiness and later achievement. *Developmental Psychology*, 43(6), 1428.
- Dunn, L., Dunn, L., Whetton, C., & Burley, J. (1997). *The British Picture Vocabulary Scale* (2 ed.). London: nferNelson.
- Ebert, K. D., & Kohnert, K. (2011). Sustained attention in children with primary language impairment: A meta-analysis. *Journal of Speech, Language, and Hearing Research*.
- Egeberg, E. (2016). *Minoritetsspråk og flerspråklighet: en håndbok i utredning og vurdering* (2. utg. ed.). Oslo: Cappelen Damm akademisk.
- Engle, R. W. (2002). Working memory capacity as executive attention. *Current directions in psychological science*, 11(1), 19-23.
- Engle, R. W., & Kane, M. J. (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. *Psychology of learning and motivation*, 44, 145-200.
- Espenakk, U., Frost, J., Færevaag, M. K., Horn, E., Løge, I. K., Solheim, R. G., & Wagner, Å. K. H. (2003). *TRAS- håndbok [TRAS handbook]*. Bergen: Tras gruppen.
- Espenakk, U., Frost, J., Færevaag, M. K., Horn, E., Løge, I. K., Solheim, R. G., & Wagner, Å. K. H. (2011). *TRAS: observasjon av språk i daglig samspill [Early observation of language development in natural settings]*. Stavanger: Nasjonalt senter for leseopplæring og leseforskning, Universitetet i Stavanger.
- Essahli Vik, N. (2019). Den føyelige pedagogen – kartlegging som disiplinering av pedagogisk praksis. 18(1). *Tidsskrift for Nordisk Barnehegeforskning*, 18(1).

## References

---

- Estes, K. G., Evans, J. L., & Else-Quest, N. M. (2007). Differences in the nonword repetition performance of children with and without specific language impairment: A meta-analysis. *Journal of Speech, Language, and Hearing Research*.
- Fan, J., McCandliss, B. D., Sommer, T., Raz, A., & Posner, M. I. (2002). Testing the efficiency and independence of attentional networks. *Journal of Cognitive Neuroscience*, 14(3), 340-347.
- 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.
- Fenson, L., Bates, E., Dale, P. S., Marchman, V. A., Reznick, J. S., & Thal, D. J. (2007). *MacArthur-Bates communicative development inventories*: Paul H. Brookes Publishing Company.
- Field, A. (2013). *Discovering statistics using IBM SPSS statistics*: Sage.
- Fisher, Z., & Tipton, E. (2015). robumeta: An R-package for robust variance estimation in meta-analysis.
- Fodor, J. A. (1983). *The modularity of mind*: MIT press.
- Follmer, D. J. (2018). Executive function and reading comprehension: A meta-analytic review. *Educational Psychologist*, 53(1), 42-60.
- Fournier-Vicente, S., Larigauderie, P., & Gaonac'h, D. J. A. p. (2008). More dissociations and interactions within central executive functioning: A comprehensive latent-variable analysis. *129(1)*, 32-48.
- Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and performance 12: The psychology of reading* (pp. 559-586). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.
- Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., DeFries, J. C., & Hewitt, J. K. (2006). Not all executive functions are related to intelligence. *Psychological Science*, 17(2), 172-179.
- Friesen, D. C., & Bialystok, E. (2012). Metalinguistic Ability in Bilingual Children: The Role of Executive Control. *Rivista di psicolinguistica applicata*, 12(3), 47-56.
- Gabszewicz, J., Ginsburgh, V., & Weber, S. (2011). Bilingualism and Communicative Benefits. *Annals of Economics and Statistics* (101/102), 271-286.
- Gall, M. D., Borg, W. R., & Gall, J. P. (1996). *Educational Research. An introduction* (6th ed.): Longman Publishers USA.

## References

---

- García, J. R., & Cain, K. (2014). Decoding and Reading Comprehension: A Meta-Analysis to Identify Which Reader and Assessment Characteristics Influence the Strength of the Relationship in English. *Review of Educational Research, 84*(1), 74-111.
- Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: a review using an integrative framework. *Psychological Bulletin, 134*(1), 31-60.
- Gathercole, V.C. M., & Thomas, E. M. (2009). Bilingual first-language development: Dominant language takeover, threatened minority language take-up. *Bilingualism: Language and Cognition, 12*(2), 213-237.
- Gatt, D., O'Toole, C., & Haman, E. (2015). Using Parental Reports to Assess Early Lexical Production in Children Exposed to More Than One language. In S. Armon-Lotem, J. de Jong, & N. Meir (Eds.), *Assessing Multilingual Children. Disentangling Bilingualism from Language Impairment*. Bristol, Buffalo, Toronto: Multilingual matters.
- Genesee, F. (2010). Dual language development in preschool children. *Young English language learners: Current research and emerging directions for practice and policy, 59-79*.
- Genesee, F., Lindholm-Leary, K., Saunders, W. M., & Christian, D. (2006). *Educating English language learners: A synthesis of research evidence*. Cambridge: Cambridge University Press.
- Geva, E. (2000). Issues in the assessment of reading disabilities in L2 children—beliefs and research evidence. *Dyslexia (10769242), 6*(1), 13-28.
- Geva, E. (2006). Second-Language Oral Proficiency and Second-Language Literacy. In *Developing literacy in second-language learners: Report of the National Literacy Panel on Language-Minority Children and Youth*. (pp. 123-139). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.
- Geva, E. (2007). *Conjunction use in School Children's Oral Language and Reading*. In R. Horowitz (Ed.), *Talking Texts. How Speech and Writing Interact in School Learning* (1st ed.).
- Geva, E., & Farnia, F. (2012). Developmental changes in the nature of language proficiency and reading fluency paint a more complex

## References

---

- view of reading comprehension in ELL and EL1. *Reading and Writing*, 25(8), 1819-1845.
- Geva, E., & Ryan, E. B. (1985). Use of conjunctions in expository texts by skilled and less skilled readers. *Journal of Reading Behavior*, 17(4), 331-346.
- Geva, E., & Wiener, J. (2015). *Psychological assessment of culturally and linguistically diverse children and adolescents: A practitioner's guide*: Springer Publishing Company.
- Gisela, H., Eva-Kristina, S., & Ulrika, N. (2003). Measuring language development in bilingual children: Swedish- Arabic children with and without language impairment. *Linguistics*, 41(2), 255-288.
- Gjervan, M., & Svolsbru, G. (2010). Behov for nyansering i kartleggingsdebatten. *Barnhagefolk*, 2, 87-89.
- Grafman, J., & Litvan, I. (1999). Importance of deficits in executive functions. *The Lancet*, 354(9194), 1921-1923.
- 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.
- Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, 1(2), 67-81.
- Grimm, A., & Schulz, P. (2014). Specific language impairment and early second language acquisition: the risk of over- and underdiagnosis. *Child Indicators Research*, 7(4), 821-841.
- Grundy, J. G., & Timmer, K. (2016). Bilingualism and working memory capacity: A comprehensive meta-analysis. *Second Language Research*, 33(3), 325-340.
- Guajardo, N. R., & Cartwright, K. B. (2016). The contribution of theory of mind, counterfactual reasoning, and executive function to pre-readers' language comprehension and later reading awareness and comprehension in elementary school. *Journal of Experimental Child Psychology*, 144, 27-45.
- Haft, S. L., Caballero, J. N., Tanaka, H., Zekelman, L., Cutting, L. E., Uchikoshi, Y., & Hoeft, F. (2019). Direct and indirect contributions of executive function to word decoding and reading comprehension in kindergarten. *Learning and Individual Differences*, 76, 101783.

## References

---

- Hakuta, K., Butler, Y. G., & Witt, D. (2000). *How Long Does It Take English Learners To Attain Proficiency? Policy Report 2000-2001*. California University. The Linguistic Minority Research Institute, Santa Barbara, CA.
- 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.
- Hammer, C. S., Davison, M. D., Lawrence, F. R., & Miccio, A. W. (2009). The Effect of Maternal Language on Bilingual Children's Vocabulary and Emergent Literacy Development During Head Start and Kindergarten. *Scientific Studies of Reading, 13*(2), 99-121.
- Hammer, C. S., Hoff, E., Uchikoshi, Y., Gillanders, C., Castro, D. C., & Sandilos, L. E. (2014). The language and literacy development of young dual language learners: A critical review. *Early Childhood Research Quarterly, 29*(4), 715-733.
- Hammer, C. S., Lawrence, F. R., & Miccio, A. W. (2008). Exposure to English before and after Entry into Head Start: Bilingual Children's Receptive Language Growth in Spanish and English. *International Journal of Bilingual Education and Bilingualism, 11*(1), 30-56.
- Han, W.-J. (2012). Bilingualism and Academic Achievement. *Child Development, 83*(1), 300-321.
- Harkio, N., & Pietilä, P. (2016). The role of vocabulary breadth and depth in reading comprehension: A quantitative study of Finnish EFL learners. *Journal of Language teaching and research, 7*(6), 1079-1088.
- Hart, B., & Risley, T. R. (2003). The Early Catastrophe. The 30 Million Word Gap by Age 3. *American Educator, 27*(1), 4-9.
- Helland, T., Jones, L. Ø., & Helland, W. (2017). Detecting preschool language impairment and risk of developmental dyslexia. *Journal of Research in Childhood Education, 31*(2), 295-311.
- Heller, M. C., Lervåg, A., & Grøver, V. (2019). Oral Language Intervention in Norwegian Schools Serving Young Language-Minority Learners: A Randomized Trial. *Reading Research Quarterly*.



## References

---

- Helvig, A., & Løge, I. K. (2006). *Presiseringer av spørsmålene i TRAS*. Stavanger: University of Stavanger.
- Hermansen, A. S. (2009). *Unmaking the vertical Mosaic? Occupational Class Attainment among Second- Generation Immigrants in Norway*. (Master of Philosophy Thesis. Sociology). University of Oslo, Department of Sociology, and Human geography, faculty of Social Sciences, University of Oslo.
- Hermansen, A. S. (2013). Occupational Attainment Among Children of Immigrants in Norway: Bottlenecks into Employment—Equal Access to Advantaged Positions? *European Sociological Review*, 29(3), 517-534.
- Hermansen, A. S. (2016). Moving Up or Falling Behind? Intergenerational Socioeconomic Transmission among Children of Immigrants in Norway. *European Sociological Review*, 32(5), 675-689.
- Hilchey, M. D., & Klein, R. M. (2011). Are there bilingual advantages on nonlinguistic interference tasks? Implications for the plasticity of executive control processes. *Psychonomic Bulletin & Review*, 18(4), 625-658.
- Hilchey, M. D., Saint- Aubin, J., & Klein, R. M. (2015). Does bilingual exercise enhance cognitive fitness in non-linguistic executive processing tasks. In *Cambridge handbook of bilingual processing*.
- Hipfner-Boucher, K., Milburn, T., Weitzman, E., Greenberg, J., Pelletier, J., & Girolametto, L. (2015). Narrative abilities in subgroups of English language learners and monolingual peers. *International Journal of Bilingualism*, 19(6), 677-692.
- Hjetland, H. N., Lervåg, A., Lyster, S.-A. H., Hagtvet, B. E., Hulme, C., & Melby-Lervåg, M. (2018). Pathways to reading comprehension: A longitudinal study from 4 to 9 years of age. *Journal of Educational Psychology*. 111(5), 751-763.
- Hoff, E. (2006). How social contexts support and shape language development. *Developmental Review*, 26(1), 55-88.
- Hoff, E. (2013). Interpreting the early language trajectories of children from low-SES and language minority homes: implications for closing achievement gaps. *Developmental Psychology*, 49(1), 4.

## References

---

- Hoff, E., Core, C., Place, S., Rumiche, R., Senior, M., & Parra, M. (2012). Dual language exposure and early bilingual development. *Journal of Child Language*, 39(1), 1-27.
- Hoff, E., Giguere, D., Quinn, J., & Lauro, J. (2018). The development of English and Spanish among children in immigrant families in the United States. *Pensamiento educativo: revista de investigacion educacional Latinoamericana*, 55(2), 1.
- Hoff, E., Rumiche, R., Burridge, A., Ribot, K. M., & Welsh, S. N. (2014). Expressive vocabulary development in children from bilingual and monolingual homes: A longitudinal study from two to four years. *Early Childhood Research Quarterly*, 29(4), 433-444.
- Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing*, 2(2), 127-160.
- 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.
- Huang, S., Fang Kan, M., & Fang Kan, P. (2019). Working with Culturally and Linguistic Diverse Populations. In M. W. Hudson & m. DeRuiter (Eds.), *Professional Issues in Speech- Language pathology and Audiology* (5th ed.).
- 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.
- Huttenlocher, J., Vasilyeva, M., Cymerman, E., & Levine, S. (2002). Language input and child syntax. *Cognitive Psychology*, 45(3), 337-374.
- Huttenlocher, J., Waterfall, H., Vasilyeva, M., Vevea, J., & Hedges, L. V. (2010). Sources of variability in children's language growth. *Cognitive Psychology*, 61(4), 343-365.
- Høien, T., & Tønnesen, G. (2008). *Introduksjonshefte til Ordkjedeprøven*. Bryne: Logometrica AS.

## References

---

- Im-Bolter, N., & Cohen, N. J. (2007). Language impairment and psychiatric comorbidities. *Pediatric Clinics of North America*, 54(3), 525-542.
- Ingre, M. (2013). Why small low-powered studies are worse than large high-powered studies and how to protect against “trivial” findings in research: Comment on Friston (2012). *NeuroImage*, 81, 496-498.
- Jeon, E. H., & Yamashita, J. (2014). L2 reading comprehension and its correlates: A meta-analysis. *Language learning*, 64(1), 160-212.
- Jurado, M. B., & Rosselli, M. (2007). The elusive nature of executive functions: a review of our current understanding. *Neuropsychology review*, 17(3), 213-233.
- Justis- og beredskapsdepartementet. (2015). *Meld. St. 30 (2015- 2016) Fra mottak til arbeidsliv- en effektiv integreringspolitikk*. Regjeringen.no
- Karr, J. E., Areshenkoff, C. N., Hofer, S. M., Rast, P., Iverson, G. L., & Garcia- barrera, M. A. (2018). The unity and diversity of executive functions: A systematic review and re-analysis of latent variable studies. *Psychological Bulletin*, 144, 1147-1185.
- Kaushanskaya, M., & Prior, A. (2015). Variability in the effects of bilingualism on cognition: It is not just about cognition, it is also about bilingualism. *Bilingualism: Language and Cognition*, 18(1), 27-28.
- 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.
- Keijzer, M., & De Bot, K. (2019). Unlearning and relearning of languages from childhood to later. In A. De Houwer & L. Ortega (Eds.), *The Cambridge Handbook of Bilingualism* (pp. 267-286): The Cambridge University press.
- Kieffer, M. J. (2008). Catching up or falling behind? Initial English proficiency, concentrated poverty, and the reading growth of language minority learners in the United States. *Journal of Educational Psychology*, 100(4), 851-868.
- Kieffer, M. J. (2012a). Before and after third grade: Longitudinal evidence for the shifting role of socioeconomic status in reading growth. *Reading and Writing*, 25(7), 1725-1746.

## References

---

- Kieffer, M. J. (2012b). 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.
- Kieffer, M. J. (2014). Morphological awareness and reading difficulties in adolescent Spanish-speaking language minority learners and their classmates. *Journal of Learning Disabilities, 47*(1), 44-53.
- Kieffer, M. J., Biancarosa, G., & Mancilla-Martinez, J. (2013). Roles of morphological awareness in the reading comprehension of Spanish-speaking language minority learners: Exploring partial mediation by vocabulary and reading fluency. *Applied Psycholinguistics, 34*(4), 697-725.
- Kieffer, M. J., & Lesaux, N. K. (2008). The role of derivational morphology in the reading comprehension of Spanish-speaking English language learners. *Reading and Writing, 21*(8), 783-804.
- Kieffer, M. J., & Lesaux, N. K. (2012). Knowledge of words, knowledge about words: Dimensions of vocabulary in first and second language learners in sixth grade. *Reading and Writing, 25*(2), 347-373.
- Kieffer, M. J., Petscher, Y., Proctor, C. P., & Silverman, R. D. (2016). Is the Whole Greater Than the Sum of Its Parts? Modeling the Contributions of Language Comprehension Skills to Reading Comprehension in the Upper Elementary Grades. *Scientific Studies of Reading, 20*(6), 436-454.
- Kieffer, M. J., & Vukovic, R. K. (2013). Growth in reading-related skills of language minority learners and their classmates: more evidence for early identification and intervention. *Reading and Writing, 26*(7), 1159-1194.
- Kieffer, M. J., Vukovic, R. K., & Berry, D. (2013). Roles of attention shifting and inhibitory control in fourth-grade reading comprehension. *Reading Research Quarterly, 48*(4), 333-348.
- Kirby, J. R., & Hogan, B. (2008). Family literacy environment and early literacy development. *Exceptionality Education International, 18*(3), 112-130.
- Klem, M., & Hagtvet, B. E. (2018). Tidlig språkkartlegging, til barnets beste? *Norsk tidsskrift for logopedi, 4*, 12-24.
- Klem, M., Melby-Lervåg, M., Hagtvet, B., Lyster, S. A. H., Gustafsson, J. E., & Hulme, C. (2015). Sentence repetition is a measure of

## References

---

- children's language skills rather than working memory limitations. *Developmental Science*, 18(1), 146-154.
- Kleven, T. A. (2008). Validity and validation in qualitative and quantitative research. *Nordic Studies in Education*, 28(03), 219-233.
- Kline, P. (2000). *The handbook of psychological testing*: London, New York: Routledge.
- Kline, R. B. (2015). *Principles and practice of structural equation modeling*: New York, London: Guilford press.
- Kohnert, K. J., & Bates, E. (2002). Balancing bilinguals II: lexical comprehension and cognitive processing in children learning Spanish and English. *Journal of Speech, Language & Hearing Research*, 45(2).
- Kopke, B. (2019). First language attrition: from bilingual to monolingual proficiency? In A. De Houwer & L. Ortega (Eds.), *The Cambridge Handbook of Bilingualism* (pp. 349- 366): Cambridge University Press.
- 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
- Kroll, J. F. (2008). Juggling two languages in one mind. *Psychological Science Agenda, American Psychological Association*, 22(1).
- Kuo, L.-j., & Anderson, R. C. (2006). Morphological awareness and learning to read: A cross-language perspective. *Educational Psychologist*, 41(3), 161-180.
- Language and Reading Research Consortium. (2017). Oral language and listening comprehension: same or different constructs? *Journal of Speech, Language & Hearing Research* (60), 1273- 1284.
- Lawson, G. M., Hook, C. J., & Farah, M. J. (2018). A meta-analysis of the relationship between socioeconomic status and executive function performance among children. *Developmental Science*, 21(2), e12529.
- Lee Salvatierra, J., & Rosselli, M. (2011). The effect of bilingualism and age on inhibitory control. *International Journal of Bilingualism*, 15(1), 26-37.
- LeFevre, J.-A., Berrigan, L., Vendetti, C., Kamawar, D., Bisanz, J., Skwarchuk, S.-L., & Smith-Chant, B. L. (2013). The role of

## References

---

- executive attention in the acquisition of mathematical skills for children in Grades 2 through 4. *Journal of Experimental Child Psychology*, 114(2), 243-261.
- Lehtonen, M., Soveri, A., Laine, A., Järvenpää, J., de Bruin, A., & Antfolk, J. (2018a). Is bilingualism associated with enhanced executive functioning in adults? A meta-analytic review. *Psychological Bulletin*.
- Leider, C. M., Proctor, C. P., Silverman, R. D., & Harring, J. R. (2013). Examining the role of vocabulary depth, cross-linguistic transfer, and types of reading measures on the reading comprehension of Latino bilinguals in elementary school. *Reading and Writing*, 26(9), 1459-1485.
- 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.
- Lesaux, N. K., Rupp, A. A., & Siegel, L. S. (2007). Growth in reading skills of children from diverse linguistic backgrounds: Finding from a 5- year longitudinal study. *Journal of Educational Psychology*, 99, 821-834.
- Levesque, K. C., Kieffer, M. J., & Deacon, S. H. (2017). Morphological awareness and reading comprehension: Examining mediating factors. *Journal of Experimental Child Psychology*, 160, 1-20.
- Lezak, M. D., Howieson, D. B., Loring, D. W., & Fischer, J. S. (2004). *Neuropsychological assessment*: Oxford University Press, USA.
- Li, M., & Kirby, J. R. (2014). The Effects of Vocabulary Breadth and Depth on English Reading. *Applied Linguistics*, 36(5), 611-634.
- Liamputtong, P. (2007). *Researching the vulnerable: A guide to sensitive research methods*. London, Thousand oaks, New Delhi: Sage Publications.
- Limbird, C. K., Maluch, J. T., Rjosk, C., Stanat, P., & Merkens, H. (2014). Differential growth patterns in emerging reading skills of Turkish-German bilingual and German monolingual primary school students. *Reading and Writing*, 27(5), 945-968.

## References

---

- Lipka, O., & Siegel, L. (2012). The development of reading comprehension skills in children learning English as a second language. *Reading and Writing, 25*(8), 1873-1898.
- Little, T. D. (2013). *Longitudinal structural equation modeling*. New York: Guilford press.
- Liu, Y., Sun, H., Lin, D., Li, H., Yeung, S. S. s., & Wong, T. T. Y. (2018). The unique role of executive function skills in predicting Hong Kong kindergarteners' reading comprehension. *British Journal of Educational Psychology, 88*(4), 628-644.
- Luk, G., & Bialystok, E. (2013). Bilingualism is not a categorical variable: Interaction between language proficiency and usage. *Journal of Cognitive Psychology, 25*(5), 605-621.
- Lyster, S.-A. H., Lervåg, A. O., & Hulme, C. (2016). Preschool morphological training produces long-term improvements in reading comprehension. *Reading and writing, 29*(6), 1269-1288.
- MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods, 7*(1), 19
- MacGinitie, W. H., MacGinitie, R. K., Maria, K., & Dreyer, L. G. (2002). *Gates- MacGinitie reading test* (4th ed.). Itasca, IL: Riverside publishing.
- Mancilla-Martinez, J., & Lesaux, N. K. (2011a). The gap between Spanish speakers' word reading and word knowledge: A longitudinal study. *Child Development, 82*(5), 1544-1560.
- Mancilla-Martinez, J., & Lesaux, N. K. (2011b). Early home language use and later vocabulary development. *Journal of Educational Psychology, 103*(3), 535.
- Marchman, V. A., Martínez, L. Z., Hurtado, N., Grüter, T., & Fernald, A. (2017). Caregiver talk to young Spanish-English bilinguals: comparing direct observation and parent-report measures of dual-language exposure. *Developmental Science, 20*(1), e12425.
- Martinsen, L. B. (2016). *Diskurser om kartlegging: en diskursanalytisk studie av massemedienes debatt om kartlegging i barnehagen*. NTNU.
- McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods, 23*(3), 412.

## References

---

- Meisel, J. M. (2018). Early child second language acquisition: French gender in German children. *Bilingualism: Language and Cognition*, 21(4), 656-673.
- 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.
- Melby-Lervåg, M., Lervåg, A., Lyster, S.-A. H., Klem, M., Hagtvet, B., & Hulme, C. (2012). Nonword-Repetition Ability Does Not Appear to Be a Causal Influence on Children's Vocabulary Development. *Psychological Science*, 23(10), 1092-1098.
- Mercer, J. R. (1973). *Labelling the mental retarded: Clinical and social system perspectives on mental retardation*. Berkley: University of California Press.
- Mezzacappa, E. (2004). Alerting, Orienting, and Executive Attention: Developmental Properties and Sociodemographic Correlates in an Epidemiological Sample of Young, Urban Children. *Child Development*, 75(5), 1373-1386.
- Midtbøen, A. H. (2019). *Etterkommere av innvandrere i Norge. Mobilitet, assimilering, diskriminering*.
- Millsap, R. E., & Yun-Tein, J. (2004). Assessing Factorial Invariance in Ordered-Categorical Measures. *Multivariate Behav Res*, 39(3), 479-515
- Mischel, W., Ayduk, O., Berman, M. G., Casey, B., Gotlib, I. H., Jonides, J., . . . Zayas, V. (2011). 'Willpower' over the life span: decomposing self-regulation. *Social cognitive and affective neuroscience*, 6(2), 252-256.
- Miyake, A., & Friedman, N. P. (2012). The Nature and Organization of Individual Differences in Executive Functions: Four General Conclusions. *Current directions in psychological science*, 21(1), 8-14.
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, 41(1), 49-100.
- Monsrud, M.-B., Rydland, V., Geva, E., Thurmann-Moe, A. C., & Lyster, S.-A. H. (2019). The advantages of jointly considering first and second language vocabulary skills among emergent



## References

---

- bilingual children. *International Journal of Bilingual Education and Bilingualism*, 1-17.
- Moser, T. (2010). Syn på læring i barnehageloven og rammeplanen (I Barnas barnehage. Målsetninger, føringer og rammer for barnehagen)[The View of Learning in The Kindergarten Act and in the Norwegian Framework Plan for the Content and Tasks]. *Oslo: Gyldendal Akademisk*.
- Muthén, L., & Muthén, B. (1998). Mplus (Version 8.4) [computer software]. Los Angeles, CA: Author: Statmodel.com.
- Naeem, K., Filippi, R., Periche-Tomas, E., Papageorgiou, A., & Bright, P. (2018). The importance of socioeconomic status as a modulator of the bilingual advantage in cognitive ability. *Frontiers in Psychology*, 9, 1818.
- Nagy, W. E., & Anderson, R. C. (1984). How many words are there in printed school English? *Reading Research Quarterly*, 304-330.
- Nagy, W. E., & Herman, P. A. (1987). Breadth and depth of vocabulary knowledge: Implications for acquisition and instruction. *The nature of vocabulary acquisition*, 19, 35.
- Nagy, W. E., & Scott, J. (2000). Vocabulary processes. In Kamil, M., Mosenthal, P., Pearson, P.D. & Barr, R (Eds) *Handbook of reading research*, 3, (pp. 269-284). Mahwah, N.J.: Lawrence Erlbaum Associates.
- Nation, K. (2006). Assessing children's reading comprehension. In M. J. Snowling & J. Stackhouse (Eds.), *Dyslexia, Speech and Language. A Practitioner's Handbook*. (2. ed.).
- Neale, M. D. (1997). *Neale Analysis of Reading Ability- Revised. Second revised British Edition*. London: GL Assessment.
- NESH-publikasjon. (2016). Guidelines for Research Ethics in the Social Science, Humanities, Law and Theology. (4. Ed): The Norwegian National Research Ethics Committees.
- Nicoladis, E. (2018). Simultaneous child bilingualism. In D. Miller, F. Bayram, J. Rothman, & L. Serratrice (Eds.), *Bilingual Cognition and Language. The state of the science across its subfields* (pp. 81-101). Amsterdam/ Philadelphia: John Benjamins publishing company.
- Nippold, M.A. (2016). *Later language development: School-age children, adolescents, and young adults: (4<sup>th</sup> Ed.)*. Austin, Texas: PRO-ED, Inc.

## References

---

- Nordahl, T., & Overland, T. (1998). *Idealer og realiteter: Evaluering av spesialundervisningen i Oslo kommune* (Vol. 20/98). Oslo: Norsk institutt for forskning om oppvekst, velferd og aldring.
- Nunnally, J. C. (1978). *Psychometric theory* (2. Ed.). New York: McGraw\_Hill.
- 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. (2005). The distributed Characteristic in Bilingual Learning. *International Symposium on Bilingualism, 1744- 1749*.
- 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 Durgunoglu & Goldberg (Eds.), *Language and literacy development in bilingual settings*. New York: The Guilford Press.
- Oller, D. K., Pearson, B. Z., & Cobo-Lewis, A. B. (2007). Profile effects in early bilingual language and literacy. *Applied Psycholinguistics, 28*(02), 191-230.
- Olson, R. K., Keenan, J. M., Byrne, B., & Samuelsson, S. (2014). Why do children differ in their development of reading and related skills? *Scientific Studies of Reading, 18*(1), 38-54.
- Ouellette, G. P. (2006). What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology, 98*(3), 554-566.
- Paap, K. R., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology, 66*(2), 232-258.
- Paap, K. R., & Sawi, O. (2014). Bilingual advantages in executive functioning: problems in convergent validity, discriminant validity, and the identification of the theoretical constructs. *Frontiers in Psychology, 5*, 962-962.
- Paradis, J. (2010). The interface between bilingual development and specific language impairment. *Applied Psycholinguistics, 31*(2), 227-252.
- Paradis, J., Emmerzael, K., & Duncan, T. S. (2010). Assessment of English language learners: Using parent report on first language

## References

---

- development. *Journal of Communication disorders*, 43(6), 474-497.
- Paradis, J., & Govindarajan, K. (2018). Bilingualism and children with developmental language and communication disorders. *Bilingual cognition and language: The state of the science across its subfields*, 347-370.
- Paradis, J., & Kirova, A. (2014). English second-language learners in preschool: Profile effects in their English abilities and the role of home language environment. *International Journal of Behavioral Development*, 38(4), 342-349.
- Paradis, J., & Ruiting, J. (2017). Bilingual children's long-term outcomes in English as a second language: language environment factors shape individual differences in catching up with monolinguals. *Developmental Science*, 20(1).
- Pauls, L. J., & Archibald, L. M. (2016). Executive functions in children with specific language impairment: A meta-analysis. *Journal of Speech, Language, and Hearing Research*, 59(5), 1074-1086.
- Pearson, B. Z., Fernandez, S. C., & Oller, D. K. (1993). Lexical development in bilingual infants and toddlers: Comparison to monolingual norms. *Language learning*, 43(1), 93-120.
- Pelham, S. D., & Abrams, L. (2014). Cognitive advantages and disadvantages in early and late bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(2), 313.
- Petersen, S. E., & Posner, M. I. (2012). The attention system of the human brain: 20 years after. *Annual review of neuroscience*, 35, 73-89.
- Pettersvold, M., & Østrem, S. (2018). *Profesjonell uro: Barnehagelæreres ansvar, integritet og motstand*: Fagbokforlaget.
- Pihl, J. (2010). *Etnisk mangfold i skolen: det sakkyndige blikket*: Universitetsforlaget.
- Place, S., & Hoff, E. (2011). Properties of Dual Language Exposure That Influence 2-Year-Olds' Bilingual Proficiency. *Child Development*, 82(6), 1834-1849.
- Place, S., & Hoff, E. (2016). Effects and noneffects of input in bilingual environments on dual language skills in 2 1/2-years-olds. *Bilingualism: Language and Cognition*, 19(5), 1023-1041.

## References

---

- Posner, M. I. (2012). Attentional networks and consciousness. *Frontiers in Psychology, 3*, 64.
- Posner, M. I., & Petersen, S. (1990). The attention system of the human brain *Annual Review of Neuroscience, 13*, 25-42.
- Posner, M. I., Petersen, S. E., Fox, P. T., & Raichle, M. E. (1988). Localization of cognitive operations in the human brain. *Science, 240*(4859), 1627-1631.
- Posner, M. I., & Rothbart, M. K. (2000). Developing mechanisms of self-regulation. *Development and psychopathology, 12*(3), 427-441.
- Potocki, A., Sanchez, M., Ecalte, J., & Magnan, A. (2017). Linguistic and cognitive profiles of 8-to 15-year-old children with specific reading comprehension difficulties: The role of executive functions. *Journal of Learning Disabilities, 50*(2), 128-142.
- Pozuelos, J. P., Paz-Alonso, P. M., Castillo, A., Fuentes, L. J., & Rueda, M. R. (2014). Development of attention networks and their interactions in childhood. *Developmental Psychology, 50*(10), 2405.
- 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*. Boston, United States: De Gruyter, Inc.
- Proctor, C. P., Montecillo, C., Silverman, R. D., & Haring, 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.
- Puglisi, M. L., Hulme, C., Hamilton, L. G., & Snowling, M. J. (2017). The Home Literacy Environment Is a Correlate, but Perhaps Not a Cause, of Variations in Children's Language and Literacy Development. *Scientific Studies of Reading, 21*(6), 498-514.
- Raghubar, K. P., Barnes, M. A., & Hecht, S. A. (2010). Working memory and mathematics: A review of developmental, individual difference, and cognitive approaches. *Learning and Individual Differences, 20*(2), 110-122.

## References

---

- Raschke, V. R. (2013). *Processes underlying syntactic control: Evaluating linguistically diverse children*. (Doctor of Philosophy). Loyola University Chicago, ProQuest Dissertations & Theses Global database.
- Raykov, T. (1997). Scale reliability, Cronbach's coefficient alpha, and violations of essential tau-equivalence with fixed congeneric components. *Multivariate Behavioral Research*, 32(4), 329-353.
- Raykov, T., & Marcoulides, G. A. (2019). Thanks Coefficient Alpha, We Still Need You! *Educational and psychological measurement*, 79(1), 200-210.
- Reynolds, R. E. (2000). Attentional resource emancipation: Toward understanding the interaction of word identification and comprehension processes in reading. *Scientific Studies of Reading*, 4(3), 169-195.
- Ringblom, N. (2010). An introduction to bilingual development. *International Journal of Bilingual Education and Bilingualism*, 13(6), 749-751.
- Rogde, K., Hagen, Å. M., Melby-Lervåg, M., & Lervåg, A. (2019). The effect of linguistic comprehension instruction on generalized language and reading comprehension skills: A systematic review. *Campbell Systematic Reviews*, 15(4), e1059.
- 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.
- Rolls, E. T. (1995). A theory of emotion and consciousness, and its application to understanding the neural basis of emotion. In Gazzaniger, M.S.(Ed) *The cognitive Neurosciences: The MIT press*.
- Rowe, M. L. (2012). A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child Development*, 83(5), 1762-1774.
- Rumbaut, R. G. (2014). English plus: Exploring the Socioeconomic benefits of bilingualism in South California. In R. M. Callahan & P. C. Gándara (Eds.), *The Bilingual Advantages. Language, Literacy and the US Labor Market* (pp. 182- 210). Bristol, Buffalo, Toronto: Multilingualism matters.

## References

---

- 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.
- Rydland, V., Aukrust, V. G., & Fulland, H. (2013). Living in neighborhoods with high or low co-ethnic concentration: Turkish-Norwegian-speaking students' vocabulary skills and reading comprehension. *International Journal of Bilingual Education and Bilingualism, 16*(5), 657-674.
- Rydland, V., Grøver, V., & Lawrence, J. (2014). The second-language vocabulary trajectories of Turkish immigrant children in Norway from ages five to ten: the role of preschool talk exposure, maternal education, and co-ethnic concentration in the neighborhood. *Journal of Child Language, 41*(02), 352-381.
- Sánchez-Cubillo, I., Periàñez, J. A., Adrover-Roig, D., Rodríguez-Sánchez, J. M., Ríos\_Lago, M., Tirpau, J., & Barcelò, F. (2009). Construct validity of the trail making test: Role of task-switching, working memory, inhibition/interference control, and visuomotor abilities. *Journal of the International Neuropsychological Society: JINS, 15*(3), 438-450.
- Sarter, M., Givens, B., & Bruno, J. P. (2001). The cognitive neuroscience of sustained attention: where top-down meets bottom-up. *Brain Research Reviews, 35*(2), 146-160.
- Saunders, W. M., & O'Brien, G. (2006). Oral language. In F. Genesee, K. Lindholm-Leary, W. M. Saunders, & D. Christian (Eds.), *Educating English language learners: A synthesis of research evidence* (pp. 14-63).
- Scheele, A. F., Leseman, P. P. M., & Mayo, A. Y. (2010). The home language environment of monolingual and bilingual children and their language proficiency. *Applied Psycholinguistics, 31*(1), 117-140.
- Schelleter, C. (2019). *An Introduction to Bilingualism*. London: Red Globe Press.
- Schwartz, M., & Katzir, T. (2012). Depth of lexical knowledge among bilingual children: the impact of schooling. *An Interdisciplinary Journal, 25*(8), 1947-1971.

## References

---

- Sebastián-Gallés, N., Echeverría, S., & Bosch, L. (2005). The influence of initial exposure on lexical representation: Comparing early and simultaneous bilinguals. *Journal of Memory and Language*, 52(2), 240-255.
- Séguin, J. R., Arseneault, L., & Tremblay, R. E. (2007). The contribution of “cool” and “hot” components of decision-making in adolescence: Implications for developmental psychopathology. *Cognitive Development*, 22(4), 530-543.
- Silverman, R. D., Proctor, C. P., Harring, J. R., Hartranft, A. M., Doyle, B., & Zelinke, S. B. (2015). Language skills and reading comprehension in English monolingual and Spanish–English bilingual children in grades 2–5. *Reading and Writing*, 28(9), 1381-1405.
- Sirin, S. R. (2005). Socioeconomic Status and Academic Achievement: A Meta-Analytic Review of Research. *Review of Educational Research*, 75(3), 417-453.
- Skybakmoen, J., Klungtveit, H. S., Berg, T. L., & Nordseth, P. (2017). Dette er de norske nettavisene du bør være på vakt mot i valgkampen. Retrieved from <https://filternyheter.no/dette-er-de-ni-norske-nettstedene-du-bor-vaere-pa-vakt-mot/>
- Slama, R. B. (2014). Investigating Whether and When English Learners Are Reclassified Into Mainstream Classrooms in the United States: A Discrete-Time Survival Analysis. *American Educational Research Journal*, 51(2), 220-252.
- St Clair-Thompson, H. L., & Gathercole, S. E. (2006). Executive functions and achievements in school: Shifting, updating, inhibition, and working memory. *The Quarterly Journal of Experimental Psychology*, 59(4), 745-759.
- Stahl, S. A., & Nagy, W. E. (2006). *Teaching word meanings*. Mahwah, N.J: Lawrence Erlbaum.
- Stangeland, E. B. (2018). *Språklig mestring, lek og sosial kompetanse hos de yngste barnehagebarna. En studie av variasjon og sammenhenger*. [Toddlers` Language proficiency, Play and Social competence within the ECEC-institution. A study of variation and relationships] (ph.d.). University of Stavanger, Stavanger.

## References

---

- Steffensen, K., & Ziade, S. E. (2009). *Skoleresultater 2008: en kartlegging av karakterer fra grunnskolen og videregående skoler i Norge* (Vol. 2009/23).
- Stephens, C. (2013). *Executive function development: A comparison of monolingual and bilingual children in Ireland*. (U620379 Ph.D.). Queen's University Belfast (United Kingdom), Ann Arbor.
- Stins, J. F., Polderman, J., Boomsma, D. I., & de Geus, E. J. (2005). Response interference and working memory in 12-year-old children. *Child Neuropsychology*, *11*(2), 191-201.
- Storch, S. A., & Whitehurst, G. J. (2002). Oral language and code-related precursors to reading: Evidence from a longitudinal structural study. *Developmental Psychology*, *38*, 934-947.
- Strand, O., Wagner, Å. K. H., & Foldnes, N. (2017). 4 Flerspråklige elevs leseresultater. In *Klar framgang!* (pp. 75-95).
- Straume, H. R. (2015). Leserinnlegg: Kartlegging av barn. *Norsk tidsskrift for logopedi* (3), 46-47.
- Stuss, D. T., & Alexander, M. P. J. P. r. (2000). Executive functions and the frontal lobes: a conceptual view. *Psychological Research* *63*(3-4), 289-298.
- Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th). Boston: Pearson.
- Tanner-Smith, E. E., Tipton, E., & Polanin, J. R. (2016). Handling Complex Meta-analytic Data Structures Using Robust Variance Estimates: a Tutorial in R. *Journal of Developmental and Life-Course Criminology*, *2*(1), 85-112.
- Tanner-Smith, E. E., & Tipton, E. (2014). Robust variance estimation with dependent effect sizes: Practical considerations including a software tutorial in Stata and SPSS. *Research synthesis methods*, *5*(1), 13-30.
- Team, R. C. (2019). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>.
- Teepen, J. (2004). On the relationship between aptitude and intelligence in second language acquisition. *Studies in Applied Linguistics and TESOL*, *4*(3).



## References

---

- The Norwegian Directorate for Education and Training. (2017). *Regulations on the framework plan for the kindergarten's content and tasks*. Utdanningsdirektoratet.
- Thompson, K. D. (2017). English learners' time to reclassification: An analysis. *Educational Policy*, 31(3), 330-363.
- Thurmann-Moe, C., Bjerkan, K. M., & Monsrud, M.-B. (2012). Utvikling av ordforståelse for ulike kategorier ord hos flerspråklige elever på morsmål og norsk. *NOA-Norsk som andrespråk*, 28(1).
- Tipton, E., & Pustejovsky, J. E. (2015). Small-sample adjustments for tests of moderators and model fit using robust variance estimation in meta-regression. *Journal of Educational and Behavioral Statistics*, 40(6), 604-634.
- Treffers- Daller, J. (2019). The Measurement of Bilingual Abilities. In A. De Houwer & L. Ortega (Eds.), *The Cambridge Handbook of Bilingualism* (pp. 289-306): Cambridge Press University.
- Tse, L. (2000). The Effects of Ethnic Identity Formation on Bilingual Maintenance and Development: An Analysis of Asian American Narratives. *International Journal of Bilingual Education and Bilingualism*, 3(3), 185-200.
- Tuller, L. (2015). Clinical use of parental questionnaires in multilingual contexts. In S. Armon-Lotem, J. de Jong, & N. Meir (Eds.), *Assessing multilingual children: Disentangling bilingualism from language impairment* (p. 301-330). Bristol, Buffalo, Toronto: Multilingual Matters.
- U.S. Department of Education (2012). *The condition of education 2012* (NCES 2012-045). Retrieved from <http://nces.ed.gov/pubs2012/2012045.pdf>
- Umansky, I. M., & Reardon, S. F. (2014). Reclassification Patterns Among Latino English Learner Students in Bilingual, Dual Immersion, and English Immersion Classrooms. *American Educational Research Journal*, 51(5), 879-912.
- Umbel, V. M., Pearson, B. Z., Fernández, M. C., & Oller, D. K. (1992). Measuring bilingual children's receptive vocabularies. *Child Development*, 63(4), 1012-1020.
- Unsworth, S. (2013). Assessing the role of current and cumulative exposure in simultaneous bilingual acquisition: The case of

## References

---

- Dutch gender. *Bilingualism: Language and Cognition*, 16(1), 86-110.
- Unsworth, S. (2016). Quantity and quality of language input in bilingual language development. In *Bilingualism across the lifespan: Factors moderating language proficiency*. (pp. 103-121). Washington, DC, US: American Psychological Association.
- Unsworth, S., Argyri, F., Cornips, L., Hulk, A., Sorace, A., & Tsimpli, I. (2014). The role of age of onset and input in early child bilingualism in Greek and Dutch. *Applied Psycholinguistics*, 35(4), 765-805.
- Vagh, S. B., Pan, B. A., & Mancilla-Martinez, J. (2009). Measuring growth in bilingual and monolingual children's English productive vocabulary development: The utility of combining parent and teacher report. *Child Development*, 80(5), 1545-1563.
- Verhallen, M., & Schoonen, R. (1993). Lexical knowledge of monolingual and bilingual children. *Applied linguistics*, 14(4), 344-363.
- Verhoeven, L. (2000). Components in early second language reading and spelling. *Scientific Studies of Reading* (4), 313-330.
- Verhoeven, L., & van Leeuwe, J. (2012). The simple view of second language reading throughout the primary grades. *Reading and Writing*, 25(8), 1805-1818.
- 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.
- Vugs, B., Cuperus, J., Hendriks, M., & Verhoeven, L. (2013). Visuospatial working memory in specific language impairment: A meta-analysis. *Research in Developmental Disabilities*, 34(9), 2586-2597.
- Vygotsky, L. S. (1964). Thought and language. *Annals of dyslexia*, 14(1), 97-98.
- Wagner, Å. K. H. (2004). *Hvordan Leser minoritetsspråklige elever i Norge? En studie av minoritetsspråklige og majoritetsspråklige*

## References

---

- 10 åringers leseresultater og bakgrunnsfaktorer i den norske delen av PIRLS 2001*. Stavanger: Høgskolen i Stavanger.
- Wechsler, D. (2003). *WISC- IV: Wechsler Intelligence Scale for Children*: NCS Pearson inc. Norwegian translation copyright (2009). Stockholm: Katarina Tryck AB.
- Wiebe, S. A., Espy, K. A., & Charak, D. J. D. p. (2008). Using confirmatory factor analysis to understand executive control in preschool children: I. Latent structure. *44*(2), 575.
- Williams, K. T. (2001). *Group Reading Assessment and Diagnostic Evaluation (GRADE). Technical manual: All levels*. Circle Pines, MN: AGS Publishing.
- Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample Size Requirements for Structural Equation Models: An Evaluation of Power, Bias, and Solution Propriety. *Educational and psychological measurement*, *76*(6), 913-934.
- Woodard, E., & Rodman, A. (2007). *Socio-economic status as a mediating factor in the relationship between bilingualism and the cognitive development of pre-school children*. Paper presented at the annual meeting of the National Communication Association 93rd Annual Convention.
- Woodcock, R. W., Muñoz-Sandoval, A. F., Ruef, M. L., Alvarado, C. G., & Wendling, B. J. (2005). *Introducing Woodcock-Muñoz language survey, revised edition: Administration and scoring training*.
- Woods, E. A. (2013). *The Development of Translation Equivalents in Bilingual Preschool Children*. (Doctoral). University of Houston.
- Yoshida, H., Tran, D. N., Benitez, V., & Kuwabara, M. (2011). Inhibition and Adjective Learning in Bilingual and Monolingual Children. *Frontiers in Psychology*, *2*.
- Zelasko, N., & Antunez, B. (2000). *If Your Child Learns in Two Languages. A Parent's Guide for Improving Educational Opportunities for Children Acquiring English as a Second Language*. Washington: National clearinghouse for Bilingual Education.
- Zelazo, P. D., Blair, C., & Willoughby, M. T. (2016). Executive Function: Implications for Education. NCER 2017-2000. *National Center for Education Research*.

### References

---

- Zelazo, P. D., Müller, U., Frye, D., & Marcovitch, S. (2003). The development of executive function. *Monographs of the Society for Research in Child Development*, 68(3), 93-119.
- Zelazo, P. D., Qu, L., & Müller, U. (2005). Hot and cool aspects of executive function: Relations in early development. In W. Schneider, R. Schumann-Hengsteler, & B. Sodian (Eds.), *Young children's cognitive development: Interrelationships among executive functioning, working memory, verbal ability, and theory of mind* (p. 71–93). Lawrence Erlbaum Associates Publishers.

## Article I

Gunnerud, H. L., ten Braak, D., Reikerås, E. K. L., Donolato, E., & Melby-Lervåg, M. (2020). Is bilingualism related to a cognitive advantage in children? A systematic review and meta-analysis. *Psychological Bulletin*, 146(12), 1059-1083.  
<http://dx.doi.org/10.1037/bul0000301>

Unfortunately, due to the copyright, this article cannot be made available in Brage.

## Article II

Gunnerud, H.L., Reikerås, E. & Dahle, A.E. (2018) The influence of home language on dual language toddlers' comprehension in Norwegian, *European Early Childhood Education Research Journal*, 26:6, 833-854, DOI: 10.1080/1350293X.2018.1533704

Unfortunately, due to the copyright, this article cannot be made available in Brage.

---

**Article III**



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

**Abstract**

It has been claimed that bilingualism is related to low language skills and reading comprehension in the instructional language. However, it has been suggested that starting with two languages at an early age can be a protective factor against such problems. Most studies of school-aged bilingual children include minority-language learners with 3-7 years of second language (L2) exposure. This study compares 91 early bilingual 5<sup>th</sup> graders with 196 monolingual peers on a range of linguistic skills and their relationships with reading comprehension. All bilingual learners in the sample have had rich exposure 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 equal text-cohesion vocabulary and decoding skills and equal predictive patterns from linguistic skills to reading comprehension. The theoretical and practical implications of the findings are discussed.

**Key phrases:** Early bilingual children, school- aged children, low age of acquisition (AoA) vocabulary, morphology, listening comprehension, reading comprehension

### **Introduction**

Several large-scale studies suggest that bilingual preadolescents (8 to 12 years) with 5-7 years of exposure to their second language (L2) still have not caught up with monolingual peers in oral language skills, such as different aspects of vocabulary, morphology and listening comprehension (Bialystok, Luk, Peets, & Yang, 2010; Droop & Verhoeven, 2003; Farnia & Geva, 2013; Oller, Pearson, & Cobo-Lewis, 2007). The same is also the case for reading comprehension skills (Herbert, Massey-Garrison, & Geva, 2020; O'Connor, Geva, & Koh, 2019). This can seriously hamper bilingual children's ability to achieve academic success and employability (Halle, Hair, Wandner, McNamara, & Chien, 2012; Han, 2012; Kieffer, 2008). Regarding predictive patterns of reading comprehension, it has been suggested that oral language is more strongly related to reading comprehension for bilingual than for monolingual children (Proctor & Louick, 2018); hence, bilingual children's reading comprehension seems to rely more heavily on their levels of instructional language than the reading comprehension of monolingual children does.

Notably, in previous studies, the selection of samples from language minority school-aged children is often convenience based, and children are not selected or grouped based on their age of acquisition (AoA). This is problematic, since the lack of these precise details on the timing of second-language acquisition prevents examination regarding the extent to which AoA, in combination with rich and long-term exposure to the instructional language, by itself could even out levels of language and literacy outcomes for monolingual and bilingual learners. Such information is vital to develop effective instruction programmes for bilingual children. Although some attempts have been made to address the relationship between AoA and bilingual children's language and literacy outcomes, studies of long-term language outcomes for early bilingual learners (AoA from birth to 3 years old) are scarce. In general, prior work examining early bilingual learners' second language and literacy levels is mostly limited to longitudinal case studies (e.g., Cruz-

Ferreira, 2006) or studies of language skills in young (0- to 4-year-old) early bilingual learners (e.g., De Houwer, Bornstein, & Putnick, 2014; Hoff, Rumiche, Burrige, Ribot, & Welsh, 2014; Hoff, Welsh, et al., 2014). Although such studies are important, the gap between bilingual and monolingual learners takes time to narrow (Hammer et al., 2014). Studies of young bilingual learners therefore bring limited clarity to the long-term language and literacy outcomes of early bilingual children and how they affect children's early school performance (Hammer et al., 2014).

For studies examining the prediction of early bilingual preadolescents' reading skills, this is even less studied than early bilingual learners' long-term language outcomes. The heterogeneity of samples in previous studies, coupled with few studies, might result in differences in language levels and predictive patterns for reading comprehension across different subgroups of bilingual learners.

Although it is well accepted that bilingualism, socioeconomic backgrounds (SES) and language and literacy outcomes are intertwined, prior studies of bilingual and monolingual comparisons do not always report or control for SES in an acceptable manner (Hammer et al., 2014). This study goes beyond exclusively reporting SES background by additionally exploring how SES is related to reading comprehension, controlling for oral language skills. Furthermore, whereas most previous studies of young bilingual children have examined bilingual children of low to middle socioeconomic backgrounds, the present study explores the language and literacy outcomes of monolingual children and a subgroup of early bilingual children mainly originating from mid- to high-SES backgrounds (Hammer et al., 2014).

### **Levels of skills for reading comprehension and their underlying factors in bilingual children**

The simple view of reading is the most common theoretical framework to understanding the development of reading comprehension (Hoover & Gough, 1990). 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 transform a string of letters into words (Hoover & Gough, 1990). The construct language comprehension is not as clear, and Hoover and Gough also used this interchangeably with listening comprehension. Importantly, the language comprehension component of the simple view comprises the meaning-based aspect of language, and typically, in addition to listening comprehension, comprises vocabulary, morphology and an understanding of narratives. The simple view has received support in a vast number of studies in both consistent and inconsistent languages and in monolingual and bilingual children (Farnia & Geva, 2013; Lervåg, Hulme, & Melby-Lervåg, 2018; Lesaux, Rupp, & Siegel, 2007; Verhoeven & van Leeuwe, 2012; Verhoeven et al., 2019; Hjetland, Brinchmann, Scherer, Hulme & Melby-Lervåg, 2020).

With respect to skill levels, a meta-analysis of the underlying components of reading comprehension in bilingual children found decoding skills in bilingual children to be on a similar level as those of monolingual children (Melby-Lervåg & Lervåg, 2014). However, if we take a closer look at language comprehension and its components, the differences in favour of monolingual children are worrying. In most linguistic aspects, numerous studies find that the differences are medium to large, even after 4-7 years of L2 exposure (Bialystok, Luk, Peets, & Yang, 2010; Droop & Verhoeven, 2003; Farnia & Geva, 2013; Hutchinson et al., 2003). For vocabulary, there are large differences in skills in favour of monolingual peers, and this gap is found for different ages and, in some studies, across different SES levels (Kieffer & Lesaux, 2012; Leider, Proctor, Silverman, & Haring, 2013; O'Connor et al., 2019; Oller, 2005; Silverman et al., 2015; Verhoeven, 2000; Verhoeven, Voeten & Vermeer, 2019). Notably, the gap in vocabulary is not restricted to low frequency academic words but also persists for high frequency words (Schwartz & Katzir, 2012; Verhallen & Schoonen, 1993).

Furthermore, it is not only vocabulary skills in the instructional language where bilingual children demonstrate poor performance. In fact, bilingual children have lower levels of vocabulary in both languages than their monolingual peers do (Monsrud, Rydland, Geva, Thurmann-Moe, & Lyster, 2019; Oller, 2005). This is called the distributed language profile and is believed to be caused by bilingual children's engagement in different learning experiences in L1 and L2; hence, part of their vocabulary is in L2 and other parts are in L1 (Oller, 2005). Additionally, vocabulary is closely linked to other linguistic aspects (Hjetland et al., 2018; Lervåg, Hulme, & Melby-Lervåg, 2018); hence, the profile effect most likely influences linguistic aspects beyond vocabulary.

Moreover, text cohesion vocabulary is measured by a task in which the child typically fills in the missing word in a sentence. The word represents an additive, temporal or causal or adversative relationship necessary to understand and draw logical relations of how ideas in one clause are related to those in an adjacent clause (Crosson, Lesauc, & Martiniello, 2008; Geva, 2007). Additionally, in this task, minority-language learners in 4<sup>th</sup> and 5<sup>th</sup> grade lag behind their monolingual peers, and the group difference in favour of monolingual learners is medium to large (Droop & Verhoeven, 2003). This is also the case for listening comprehension; minority language learners perform at a lower level than their peers do, ranging from small to large effect sizes in favour of monolingual learners (Geva & Farnia, 2012; Hutchinson, Whiteley, Smith, & Connors, 2003; O'Connor et al., 2019). Thus, bilingual children, although at a level comparable to that of their peers in decoding, tend to struggle with tasks that tap meaning-based language skills.

Finally, morphology refers to knowledge about the smallest meaning-bearing units of language, and this is suggested to draw mainly on meaning-based skills (Nagy, Carlisle, & Goodwin, 2013). For morphology, the results are more ambiguous: Some studies of young bilingual learners find bilingual learners to have a relative advantage in

morphological skills (Barac & Bialystok, 2012; Friesen & Bialystok, 2012). However, the results from studies of older bilingual children are mixed, ranging from no differences at all (Lipka & Siegel, 2012) to an even larger gap in favour of monolingual children in morphological knowledge than in vocabulary skills (Droop & Verhoeven, 2003).

### **Predictive patterns for reading comprehension and their underlying factors in bilingual children**

For differences in predictive patterns of reading comprehension across language groups, the results are somewhat inconsistent. There are studies of minority-language learners where the predictive strength of L2 skills is similar across language groups (Babayiğit, 2015; Verhoeven et al., 2019). For instance, the study by Babayiğit (2015) showed that the direct path coefficients from oral language (latent variable of sentence repetition, verbal working memory and vocabulary) and decoding were comparable across language groups. On the other hand, a review of Proctor and Louick (2018) concludes with preliminary support that language skills are a stronger predictor of reading comprehension for bilingual than monolingual children. Thus, if so, this implies that it is more crucial to strengthen bilingual than monolingual learners' language skills in instructional language to facilitate proficient reading comprehension.

Therefore, if we look at the different components of language comprehension, do some of them stand out as particularly important for reading comprehension in bilingual children? A meta-analysis synthesized studies of correlations between different L2 aspects and L2 reading comprehension in minority-language learners (Jeon & Yamashita, 2014). The results showed that the relationship between L2 vocabulary and L2 reading comprehension was similar to L2 listening comprehension and L2 reading comprehension ( $r = 0.79$  and  $r = .77$ , respectively) and L2 morphological skills and L2 reading comprehension ( $r = .61$ ). These constructs do, however, share some underlying common language abilities. As for decoding the relationship

with reading comprehension seems somewhat weaker ( $r = .56$ ) (Jeon & Yamashita, 2014). However, it is important to note that the correlation with reading comprehension for the predictors in the simple view of reading varies over time. In young readers, decoding skills are more important; in older children, language comprehension gradually takes over and explains most of the variation, since children then master decoding (Lervåg, Hulme & Melby-Lervåg, 2018; Hjetland et al 2018).

As for the relative strength between the different language comprehension predictors, most studies of the examined group of minority-language learners who were 3<sup>rd</sup>-7<sup>th</sup> graders find L2 vocabulary to uniquely predict L2 reading comprehension over and above other linguistic constructs and to appear to be independent of assessment tests (Burgoyne, Whiteley, & Hutchinson, 2011; Hutchinson et al., 2003; Kieffer, 2012; Proctor, Montecillo, Silverman, & Haring, 2012). There is also support for listening comprehension uniquely predicting L2 reading comprehension over other language comprehension constructs in minority-language learners in this age group (Burgoyne et al., 2011; Hutchinson et al., 2003; Kieffer, 2012; Proctor et al., 2012; Geva & Farnia, 2012; Proctor, Carlo, August, & Snow, 2005). However, this contribution sometimes overlaps with vocabulary skills or is not even present at all when controlled for other L2 linguistic skills (Burgoyne et al., 2011; Hutchinson et al., 2003; Kieffer, 2012; Kieffer, Biancarosa, & Mancilla-Martinez, 2013). Furthermore, text cohesion vocabulary sometimes explains the largest proportion in L2 reading comprehension but is often redundant when controlling for other vocabulary skills when reading comprehension is assessed by other texts (Rydland, Aukrust, & Fulland, 2012).

### **Levels of literacy skills and predictive pattern for reading comprehension in early bilingual learners**

If we take a closer look at studies of language levels in bilingual children with an early AoA, they are few in number and inconsistent. Most studies targeting early bilingual learners, often “bilingual first” children

(bilingual children introduced to two languages at birth), investigate their language skills in toddlerhood. In one synthesis of studies, it is claimed that bilingual first children reach milestones at the same pace as their monolingual peers and that differences between bilingual first children and monolingual children are not larger than what can be expected between monolingual children (De Houwer, 2009a, 2009b, 2012). In contrast, Bialystok and Feng (2011) examined the vocabulary skills of a large sample ( $N = 963$ ) of bilingual first children from five to 9 years old and found a large effect size difference in vocabulary breadth in favour of monolingual children.

Unfortunately, studies that examine the language skills of preadolescent bilingual first children and their relations to reading comprehension are limited and have collapsed samples of bilingual first learners and bilingual learners with early AoA (often referred to as early bilingual learners). These studies of early bilingual children show conflicting results for reading comprehension, ranging from large differences between the two language groups to no difference at all. For instance, Kovelman, Baker and Petitto (2008) found a large and significant gap in favour of monolingual children ( $d = 1.47$ ); Grant, Gottardo and Geva (2011) and Bonifacci and Tobia (2016) found a medium and significant gap in favour of monolingual children ( $d = 0.57$ , and  $d = 0.69$ ). In contrast, Wagner (2004) found no significant differences between groups.

This divergence in results could partly be caused by a focus on different subgroups of bilingual learners. The two studies with the largest effect sizes are of bilingual children with a mix of AoA from birth until 3 or 4 years of age (Bonifacci & Tobia, 2016; Kovelman, Baker, & Petitto, 2008). However, both Wagner (2004) and Grant et al (2011) investigated bilingual learners from birth and found highly different results (n.s. difference to  $d = 0.57$ ). Notably, the studies of Wagner (2004) and Grant et al. (2011) differ regarding sample characteristics and measures of reading comprehension. Wagner studied a large sample of bilingual children with one native majority-speaking parent and one



minority-speaking parent. Native language input is found to predict growth in bilingual children's language skills above the amount of exposure (Place & Hoff, 2011, 2016). Grant et al. (2011) do not provide information on bilingual parents' nationality and native language. Another difference is that Wagner used the PISA reading comprehension measures of 2003, and Grant et al. (2011) used the NARA reading comprehension test, a test that draws heavily on linguistic skills (Nation, 2006). The study by Wagner was the only study of 5<sup>th</sup> graders, and all the other studies of early bilingual children included somewhat younger children.

Regarding vocabulary in early bilingual learners in 2<sup>nd</sup> – 4<sup>th</sup> grades, studies also show conflicting findings. Some studies find a large gap in favour of monolingual learners (Grant et al., 2011; Vernice & Pagliarini, 2018;  $d = .97$  and  $d = .139$ , respectively). Others, however, find no differences in vocabulary skills between monolingual and bilingual children systematically exposed to instructional language from at least the age of 2 (Hsu, Ip, Arredondo, Tardif, & Kovelman, 2019). Interestingly, the sample in Grant et al. (2011) was exposed to the instructional language from birth; one should therefore perhaps assume that this subgroup would outperform the sample of Hsu et al. (2019) and of Vernice and Pagliarini (2018).

*Morphology* skills in early bilingual children are less studied than other aspects of language. Vernice and Pagliarini (2018) found a large and significant gap ( $d = 1.14$ ) in favour of monolingual children, yet not as large as the gap in vocabulary. This could indicate that even though there is a gap in morphological skills between monolingual and bilingual children, morphology is a relative strength compared to vocabulary. However, this study examines a small sample of 2<sup>nd</sup> and 4<sup>th</sup> graders ( $N = 24$ ). The study of Hsu et al. (2019) has larger sample sizes. They investigated morphological skills across language groups in 3<sup>rd</sup> graders, and in line with their findings concerning vocabulary skills, there were no significant differences between monolingual and bilingual children with AoA before the age of 3.

Bonifacci and Tobia (2016) studied 1<sup>st</sup> to 5<sup>th</sup> graders with AoA before the age of 4 and found a medium effect size difference in *listening comprehension* in favour of monolingual learners. However, it is debatable whether an AoA from birth and up until the age of 4 can be considered indicative of an early bilingual child. It should perhaps be expected that bilingual learners with lower AoA and a mean age of 5<sup>th</sup> graders would develop better linguistic skills than the sample of Bonifacci and Tobia (2016) did, since they then have had more time to develop their second language.

There are few studies that examine predictive patterns from linguistic skills to reading comprehension in early bilingual children. The study by Grant et al. (2011) is the only one that investigated the predictive pattern from linguistic skills to reading comprehension in preadolescent children. They examined the prediction of 2<sup>nd</sup> and 3<sup>rd</sup> graders' vocabulary and decoding skills on reading comprehension and found that the pattern differed across language groups. Both decoding and vocabulary predicted reading comprehension for bilingual first children, yet only decoding significantly predicted reading comprehension in English-speaking monolingual children.

### **The present study**

The aim of this study is twofold: first, to determine to what extent mid- to high-SES early bilingual learners richly exposed to the instructional language by at least the age of 2 catch up to language levels similar to those of their monolingual peers in vocabulary reading comprehension, listening comprehension, morphological knowledge and text cohesion vocabulary at 10 years old. In addition, we will examine how language skills and SES are related to reading comprehension.

This research will add to the literature because there are few studies of early bilingual learners, particularly of long-term language outcomes and the relations between these skills and reading comprehension. Furthermore, while summing up existing research on early bilingual learners, Hammer et al. (2014) found it difficult to

conclude which factors would promote language and literary trajectories for early bilingual learners due to a lack of sample characteristics reported in primary studies. Thus, the present study addresses this need identified in the review of Hammer et al. (2014) by providing information on the sample's SES, the amount of instructional language input (at least 5 days a week for a minimum of 8 years), and the AoA, all of which are factors suggested to play a key role in early bilingual developmental language trajectories.

The study also addresses another critical need within the field of bilingualism. There are few studies of early bilingual learners with middle to high SES, a subgroup where the gap between monolingual and bilingual learners is narrower than that for children of lower SES levels (Oller, Jarmulowicz, Pearson, & Cobo-Lewis, 2011). Furthermore, compared with other studies in this area, we use a more robust methodology suited for comparing groups in regard to both level and predictive patterns. This is because we examine whether the measures hold across the two groups and are invariant. We also examine whether the items on each measure load on the overall construct that is targeted. Theoretically, the study adds to the understanding of early bilingual long-term language and literacy development. It could shed light on whether the distributed language profile also likely affects the language and literacy skills of early bilingual learners. In addition, the study can have implications for educational practice because more knowledge about the typical development of early bilingual learners could reduce the over- and under identification of early bilingual learners in need of special need education.

The research questions are as follows:

- 3) To what extent do bilingual 5<sup>th</sup> graders with an AoA of the instructional language from birth to 2 years old have levels of language and reading comprehension skills similar to those of

their monolingual peer across different aspects of language and reading?

- 4) Are the patterns in which aspects of language comprehension, decoding skills and SES that predict reading comprehension the same for bilingual and monolingual children?

## Method

### *Participants*

One hundred ninety-six monolingual and 91 bilingual children were recruited to participate in this study. Ethical approval was obtained from the Norwegian Social Science Data Service, and informed parental consent was collected. The majority of the sample are of average mean SES of the children was middle to high SES (above 3 years of college), and the SES was at the same level in both groups. Sixty of the bilingual children are bilingual first children, with one native Norwegian-speaking parent, whilst 31 of the children had two minority-speaking parents with an AoA of at least age two. The largest language groups were English ( $N=22$ ) and German ( $N= 14$ ), but 31 different languages were represented.

The participants are a subsample of the children enrolled in the longitudinal Stavanger Project: The Learning Child, a collaboration between the National Centre for Reading Education and Research at the University of Stavanger and the municipality of Stavanger (<https://lesesenteret.uis.no/our-research/research-projects/the-stavanger-project/>). In the recruitment of this subsample, invitations to participate were sent out to parents of bilingual children in The Stavanger Project born in 2006 and 2007. Monolingual participants were only recruited from the 2007 birth class. Of this cohort, only the monolingual participants who attended schools where there was a bilingual child born in the 2007 cohort in The Stavanger Project were recruited as control children.

### *Measures*

The children were tested with a wide range of language tasks as well as reading comprehension and decoding skills.

*Reading comprehension* was examined by a Norwegian adaptation of the NARA (Neale, 1997). The tests consist of six texts, with increasing length and complexity. If children decoded a word erroneously, the correct word was presented to them. The children were asked questions about the texts immediately after they read them. The test procedures were followed to ensure that the child's decoding skills were at a sufficient level for the child to be able to answer the questions correctly. The test was stopped, and all following questions were scored 0 if the decoding skills required the testing to stop or if the child could not provide an acceptable answer to five ascending questions related to the same text. The alpha reliability was .820.

*Listening comprehension* was tested 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 the text. If a child could not correctly answer four ascending questions related to one of the texts, the test was stopped, and all following items were scored 0. The alpha reliability was .936.

*Text cohesion vocabulary* was tested with a Norwegian translation of the test developed by Droop and Verhoeven (2003). The test involves reading two cloze texts and filling in the connectives, such as *in spite and in contrast to*, using a multiple-choice format. To choose the correct alternative, the child must understand both the meaning of the sentences and the relationship between the sentences. The test was group-based, administered in a pen and paper format and accompanied with a verbal presentation of the sentences and the response alternatives. The alpha reliability was .631.

*Vocabulary* was measured 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, rewarding the child's description with

0, 1 or 2 points depending on the quality and accuracy of the description. The alpha reliability was .729.

*Morphological knowledge* is a version of the test used in the study of Brinchmann, Hjetland, and Lyster (2016), supplemented with additional items with greater difficulty to provide normally distributed data for 5<sup>th</sup> graders. This is also a group-based cloze test in pen and paper format, with additional verbal support to prevent influence of the child's decoding skills on the test results. The child is presented with a sentence that includes a nonword and asked to describe the meaning of this nonword within a multiple-choice format. The nonword comprises two meaningful morphological items, a derivational morpheme in combination with a prefix or suffix and is interpretable if the core meaning of the morphological items is combined and understood.

*Decoding* skills were tested with a word chain test (Høyen & Tønnesen, 2008), which is a Norwegian version resembling the Test of Silent Word Reading Fluency (TOSWRF) (Mather, Hammil, Allan, & Roberts, 2014). The test is in pen and paper format and is administered on a group level with time limitations (4 minutes). The tests consist of 60 chains of high-frequency words, where four words are presented together in a continuous string of letters. The child is asked to mark where one word ends and the next begins. Each word chain where all marks were correctly placed was awarded 1 point.

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

Information on *parents' educational length (SES)* was obtained through a questionnaire comprising 4 categories (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, the two lower SES categories were collapsed. Thus, three categories of SES were used in the further analysis. Note that only a

marginal proportion of the participants' parents had vocational education or lower (N= 24).

### ***Procedure***

The children were tested on vocabulary and the two NARA tests individually by trained test leaders in a quiet room at their school. The test order was fixed, starting with NARA listening comprehension and followed by NARA reading comprehension; the test ended with the vocabulary test. The testing took 1 hour on average. The participants' teachers attended a course on how to administer the two group tests and were encouraged to perform the testing on separate days to prevent test fatigue. 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 before 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. This worked as a guideline for the teachers to adjust the test pace in order for all pupils to finish answering one question before the teacher presented the next.

### **Data analysis**

#### ***Measurement invariance analysis***

Invariance was tested to ensure that differences across groups represent true differences across language groups rather than a comparison of skills across different constructs. Given the categorical-ordinal nature of the items (i.e., right/wrong on the tests), multigroup CFA based on polychoric correlations was used to evaluate measurement invariance across monolingual and bilingual learners' language and reading skills. Model estimation was performed using the WLSMV estimator in Mplus version 8.4 (Muthén & Muthén, 1998). In total, 5 linguistic factors were explored to ensure that all items loaded on an invariant latent variable. Thus, for instance, for the text cohesion vocabulary test, we examined whether all items that aimed to represent text cohesion loaded the same

latent variable and whether these items were invariant across language groups. A similar procedure was followed for all measures. The invariant latent variables, along with a manifest variable of decoding, were then used to examine the predictive patterns of linguistic variables and decoding on reading comprehension. In line with T. A. Brown (2015), preliminary tests were conducted to identify items not recommended for inclusion in the latent variable before formal invariance tests were run. First, items that did not have significant factor loadings on the overall latent variable were removed. Additionally, items that were not significantly related to the latent variable across the different language groups and items with negative factor loadings were removed. Before formal invariance tests were conducted, the overall model fit as well as the model fit for the two language groups separately were investigated to ensure acceptable model fit. In some cases, adjustments of the model were necessary 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 the inferences in the CFAs (M. B. Brown & Benedetti, 1977; Olsson, 1979). Therefore, items with limited variance were removed.

Performing a multi-group analysis in Mplus with categorical data requires an equal number of categories across groups. Problematic items were dropped in cases in which the item contained an unequal number of categories for the two groups. Model fit statistics were then applied to ensure that the model to be tested for measurement invariance had acceptable fit. Acceptable or good model fit is typically set to RMSEA values below 0.8 or 0.6, respectively, and CFI and TLI are above 0.90 and 0.95, respectively (Hooper, Coughlan, & Mullen, 2008; Hu & Bentler, 1999). Formal measurement invariance was thereafter investigated.

Measurement invariance involves testing a sequence of nested models and assessing in each step whether the imposed constraints are in



line with the data. Four models, referred to as the configural, metric, scalar and strict model, were tested using the theta parameterization (Millsap, 2012). Modification indices were used to identify variant factor loadings, thresholds and residual variances. Variant items were dropped to create comparable constructs. We followed the recommendations of Sass, Schmitt and Marsh (2014) and used chi-square difference tests to formally detect invariance across nested models when analysed with the estimator WLSM, yet we report RMSEA and CFI for transparency.

***Latent means and regression patterns across groups***

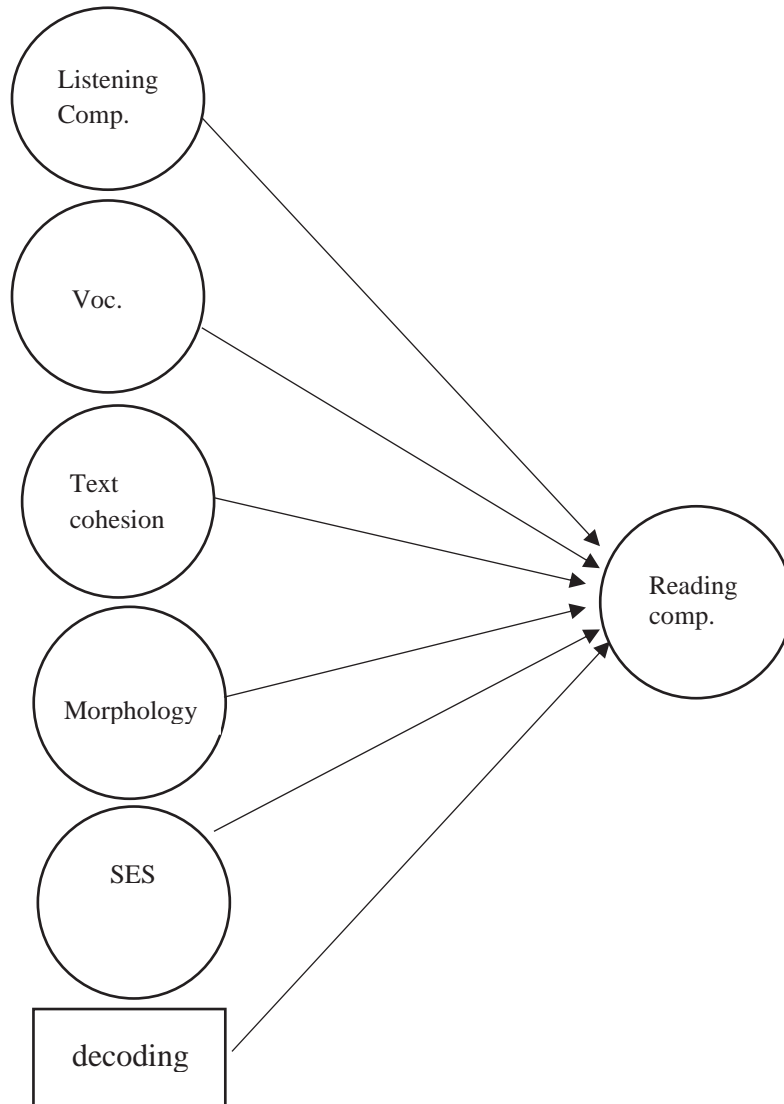
Differences in latent variable means were studied by comparing the means across groups in the strict model. In this model, factor loadings and thresholds were constrained to be equal between the two groups, and residuals were fixed to 1 in the bilingual group whilst they were freely estimated in the monolingual group. The invariance testing makes such a comparison meaningful, since differences in group levels can then be ascribed to true differences in performance, not to comparisons of skills across different constructs. Whether differences in factor variance were significantly different in the two groups was tested by comparing the scalar model (where factor variance is allowed to differ across groups) with a model where factor variances for both groups were set to be equal.

Given the complexity of our measurement models, combined with the categorical-ordinal nature of the items and relatively small sample sizes, model estimation is complicated by issues with 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 comprised of the sum scores of two or more manifest variables. Parcels can be used in SEM modelling when the underlying nature and dimensions of such items are known (Little, Cunningham, Shahar, & Widaman, 2002). Parceling significantly reduces the number of parameters to be estimated while maintaining the content validity of our latent variables.

Parceling was done systematically by replacing at least five items with a single parcel item containing their mean score. Only items that were invariant were included in the parcel items, which ensures that the parcel items are also invariant. The resulting parcels had at least six categories each. Since the parcels had symmetrical distributions, we decided to treat them as continuous variables. This facilitates model estimation, since we then do not need polychoric correlations but rather use the observed Pearson correlations. Only the three-level indicators of the SES of fathers and mothers were treated as truly categorical by employing polyserial and polychoric correlations.

To examine regression patterns across groups, the model in Figure 1 was tested but without the morphological knowledge variable. The model examining the prediction of reading comprehension is more complex than its individual measurement models used for latent mean and variance testing. To handle the problem of missing data, we decided to employ multiple imputation. The results of model estimation in each imputed dataset were combined using the R package *semTools* (Team, 2019). Notably, we also conducted the above analysis using a different statistical approach, where the item factor score regressions were combined with multiple imputation. This procedure gave the same conclusions as with the SEM approach described above.

Figure 1. Preliminary regression model to be tested by multigroup SEM for monolingual and bilingual children.



## Results

### *Descriptive statistics*

Descriptive statistics were performed in SPSS, version 25 (Pallant, 2020). Table 1 reports the mean, SD, minimum, maximum, skewness and kurtosis values for all manifest variables. Correlations between the variables are reported in Table 2. Table 1 shows that all variables, with the exception of morphology, are normally distributed. As shown in Table 2, all language variables and reading comprehension were correlated. Decoding skills were only correlated with text cohesion vocabulary, vocabulary and reading comprehension but were uncorrelated with listening comprehension and morphology.

Table 1

*Minimum, maximum, mean, SD and missing values for the decoding, reading comprehension, listening comprehension, vocabulary, text cohesion vocabulary and morphology variables*

Variable	Min	Max	Mean	SD	Skewness	Kurtosis
Decoding	5	59	31.05	9.62	.131	-.120
Reading comprehension	0	28	11.69	5.96	.161	.287
Listening comprehension	0	33	15.11	8.56	.150	-1.154
Vocabulary	0	21	9.66	4.14	.475	.022
Text cohesion vocabulary	1	11	7.14	2.08	-.202	-.280
Morphology	2	4	3.93	0.27	-3.918	15.718

*Note.* The different variables consist of sum scores based exclusively on the invariant items of these variables.

Table 2

*Correlations with confidence intervals between decoding, reading comprehension (RC), listening comprehension (LC), vocabulary, text cohesion vocabulary (TCV) and morphology*

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

*Note.* Correlations between sum scores of different variables created by merging invariant items only. \* indicates  $p < .05$  \*\* indicates  $p < .01$ .

#### ***Confirmatory factor analysis—overall and for each group***

Reading comprehension was first fitted as a one-factor model, with correlated items on the text level. Non-significant items were then removed from the model. No correlation between items was assumed in the other CFA models. Table 3 presents model fit statistics for both groups for the final CFA models.

As shown in Table 3, fit indices for the CFAs of reading comprehension and listening comprehension indicated that the model had good fit in both language groups. Model fit indices for the CFA of text cohesion vocabulary differed slightly between the language groups, with somewhat better fit for the bilingual group. This indicates that there might be some difference in CFI structure between the two groups. However, since model fit indices are within acceptable range for both groups, we proceeded with invariance testing.

Table 3

*Separate model fit statistics for CFA models of the variables listening comprehension, reading comprehension, vocabulary and text cohesion vocabulary for the monolingual and bilingual groups*

Model fit statistics for	Monolingual group				Bilingual group			
	$\chi^2$	p.	RMS EA	CFI	$\chi^2$	p.	RMS EA	CFI
Listening comprehension	646.86	0.00	0.03 4	0.98 3	540.42	0.33	0.01 7	0.996
Reading comprehension	254.78	0.10	0.02 5	0.98 6	269.70	0.03	0.04 5	0.933
Vocabulary	57.22	0.36	0.01 7	0.99 3	67.20	0.11	0.05 2	0.895
Text cohesion vocabulary	66.82	0.13	0.03 6	0.90 5	59.55	0.28	0.03 4	0.950

*Note.* \*\*  $p < .05$ , \*\*\*  $p < 0.001$

The primary analysis of the variable vocabulary was more challenging. After all items that could reduce model fit were removed, the overall model fit was good ( $\chi^2 (104) = 124.93$ ,  $p = 0.079$ , RMSEA = 0.026, CFI = 0.977), accompanied by an acceptable model fit of the monolingual group ( $\chi^2 (104) = 107.31$ ,  $p = 0.392$ , RMSEA = 0.013, CFI = 0.995). The model fit for the bilingual group did, however, indicate a mismatch between the data and the model ( $\chi^2 (104) = 143.43$ ,  $p = 0.006$ , RMSEA = 0.065, CFI = 0.853). 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 was a hindrance for configural invariance, the most difficult item was excluded. In line with suggestions in the modification indices, several more items were excluded 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 an acceptable fit (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  $\chi^2$ -difference test that constrained factor loadings, thresholds and residuals to be equal between the groups was not significant; therefore, we concluded that there was adequate model fit for both groups.

Factor analysis of morphological knowledge indicated a multi-factorial structure, which was supported by the identification of a three-factor model in an exploratory factor analysis ( $\chi^2(187) = 202.626, p = 0.206, RMSEA = 0.018, CFI = 0.979$ ). Invariance testing of the factor containing most of the items for the morphology variable indicated that few factor loadings were significant in both groups, resulting in a reduction of 20 out of 25 test items. Since 73.5% of all participants scored within the two highest performance levels, the test items that were left were considered unsuitable for identifying possible differences across groups. Morphology was therefore excluded from further analysis.

### ***Invariance analysis***

Based on the previous analysis, we tested invariance for the variables reading comprehension, listening comprehension, vocabulary and text cohesion vocabulary. The results of this analysis are shown in Table 4.

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 the sample size, which has led to a recommendation to rely on other fit statistics often acknowledged as 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 text cohesion vocabulary (RMSEA = 0.035, CFI = 0.926) are all acceptable; hence, we conclude with configural invariance for all variables. Chi-square tests between nested models show no significant differences for configural,

metric, scalar and strict models for any of the variables. We therefore conclude with strict invariance across all tested variables.

Table 4

*Test of measurement invariance between the two language groups for reading comprehension, listening comprehension, vocabulary, and text cohesion vocabulary*

	$\chi^2$	df	diff $\chi^2$	diff df	RMSEA	CFI
Reading comprehension						
Overall sample MI across groups	273.01**	227			0.027	0.984
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 MI across groups	795.10***	561			0.038	0.979
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
Vocabular						
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
Text cohesion vocabulary						
Overall	52.211	54			0.000	1.000
sample						
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

Note. \*\*  $p < 0.05$ , \*\*\*  $p < 0.001$ , MI = measurement invariance.

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

SES data were collected at the formal end of The Stavanger Project, before parents were invited to take part in the present study; hence, missing data were treated as missing at random. We therefore imputed missing values for SES for both parents. The robust Mann-Whitney U did not lend support for any differences in SES between mono- and bilingual children ( $p = 0.26$  for the SES of mothers and  $p = 0.51$  for the SES of fathers).

Group differences in the manifest variable decoding were investigated by a t-test, and 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 have full scalar

invariance for the vocabulary, text cohesion vocabulary, listening comprehension and reading comprehension variables. There were significant differences in the means for reading comprehension, listening comprehension and vocabulary, while the means for text cohesion vocabulary were equal across groups. Differences in factor means and variances are displayed in Tables 5 and 6. Differences in means are standardized and can be interpreted as group differences measured by Cohens *d*.

Table 5

*Standardized differences in factor means between the two language groups for reading comprehension, listening comprehension, text cohesion vocabulary, vocabulary and morphology*

Construct of comparison	Standardized difference in factor means	SE	<i>p</i>
Reading comprehension	0.78	0.145	0.009
Listening comprehension	0.60	0.191	0.002
Text cohesion vocabulary	0.34	0.206	0.102
Vocabulary	0.74	0.210	<i>p</i> < 0.001

*Note.* The factor mean of the bilingual learners is set to zero and used as a reference group. A positive difference in means in the standardized model indicates a favour for monolingual learners.

Table 6

*Standardized differences in factor variance between the two language groups for reading comprehension, listening comprehension, text cohesion vocabulary, vocabulary and morphology*

Construct of comparison	Standardized difference in factor means	SE	<i>P</i>
Reading comprehension	1.055	0.254	n.s.
Listening comprehension	0.416	0.126	n.s.
Text cohesion vocabulary	0.671	0.218	n.s.
Vocabulary	0.812	0.274	n.s.

*Note.* The factor variance of the bilingual is set to 1 and used as a reference group. A variance less than 1 for the unstandardized model indicates larger factor variances of the bilingual group.

Chi-square difference tests were used to determine whether variance differed across groups. The  $\chi^2$  difference tests were not significant for any of the constructs, i.e., the factor variance for all variables was equal across groups [reading comprehension ( $\chi^2(1) = 0.049$ ,  $p = 0.824$ ), listening comprehension ( $\chi^2(1) = 2.607$ ,  $p = 0.106$ ), text cohesion vocabulary ( $\chi^2(1) = 1.409$ ,  $p = 0.235$ ), vocabulary ( $\chi^2(1) = 0.032$ ,  $p = 0.858$ ), and morphology ( $\chi^2(1) = 0.967$ ,  $p = 0.325$ )].

***Comparison of predictive patterns for reading comprehension between the two groups***

To analyse multiple datasets using WLSM as an estimator, a change of software from Mplus to R was necessary. Descriptive statistics of all variables used in the multi-group SEM model are listed in Table 7, and correlations between the linguistic variables are presented in Table 8.

Table 7

*Minimum, maximum, median, mean, SD and missing values for the variables listening comprehension parcel 1-5, reading comprehension parcel 1-3, vocabulary parcel 1-2, text cohesion vocabulary parcel 1-2, decoding, SES and language group*

Variable	Min	Max	Median	Mean	SD	Missing
Listening comprehension parcel 1	0	1	0.6	0.6	0.3	0.7%
Listening comprehension parcel 2	0	1	0.6	0.6	0.3	0.7%
Listening comprehension parcel 3	0	1	0.4	0.4	0.3	0.7%
Listening comprehension parcel 4	0	1	0.4	0.5	0.3	0.7%
Listening comprehension parcel 5	0	1	0.4	0.5	0.3	0.7%
Reading comprehension parcel 1	0	1	0.6	0.6	0.2	0%
Reading comprehension parcel 2	0	1	0.3	0.3	0.2	0%
Reading comprehension parcel 3	0	1	0.6	0.5	0.2	0%
Vocabulary parcel 1	0	2	0.8	0.9	0.4	0%
Vocabulary parcel 2	0	2	0.8	0.8	0.5	0%

*Article III*

Text cohesion vocabulary parcel 1	0	1	0.7	0.6	0.2	4.88%
Text cohesion vocabulary parcel 2	0	1	0.7	0.7	0.2	4.88%
Decoding	5	59	31	31	9.6	0.7%
SES father	1	3	3	2.3	0.8	41.46%
SES mother	1	3	3	2.5	0.7	41.46%
Language group	1	2		1.7		0%

As shown in Table 8, the intercorrelations among parcels of the latent variables of vocabulary and listening comprehension are large. Correlations between parcels across these constructs, such as listening comprehension parcel 1 and vocabulary parcel 2, are, however, only moderate. This is also the case for parcels representing text cohesion vocabulary. Although they are only moderately correlated, this correlation is higher than the correlation between these parcels and parcels representing other latent language constructs.

The regression pattern for listening comprehension, vocabulary text cohesion vocabulary, decoding skills and SES on reading was first tested on the overall sample. This provided an excellent model fit ( $N = 287$ ,  $\chi^2 [80.0] = 53.996$ ,  $p = 0.989$ ,  $RMSEA = 0.000$ ,  $CFI = 1.000$ ,  $TLI = 0.937$ ). Listening comprehension was the variable that had the strongest relationship with reading comprehension ( $\beta = 0.52$ ,  $SE = 0.07$ ,  $p = 0.000$ ), followed by text cohesion vocabulary ( $\beta = 0.428$ ,  $SE = 0.19$ ,  $p = 0.03$ ). When controlling for text cohesion vocabulary and listening comprehension, vocabulary and SES were not significant. Decoding was only marginally related to children's reading comprehension skills ( $\beta = 0.003$ ,  $SE = 0.001$ ,  $p = 0.02$ ).

Table 8

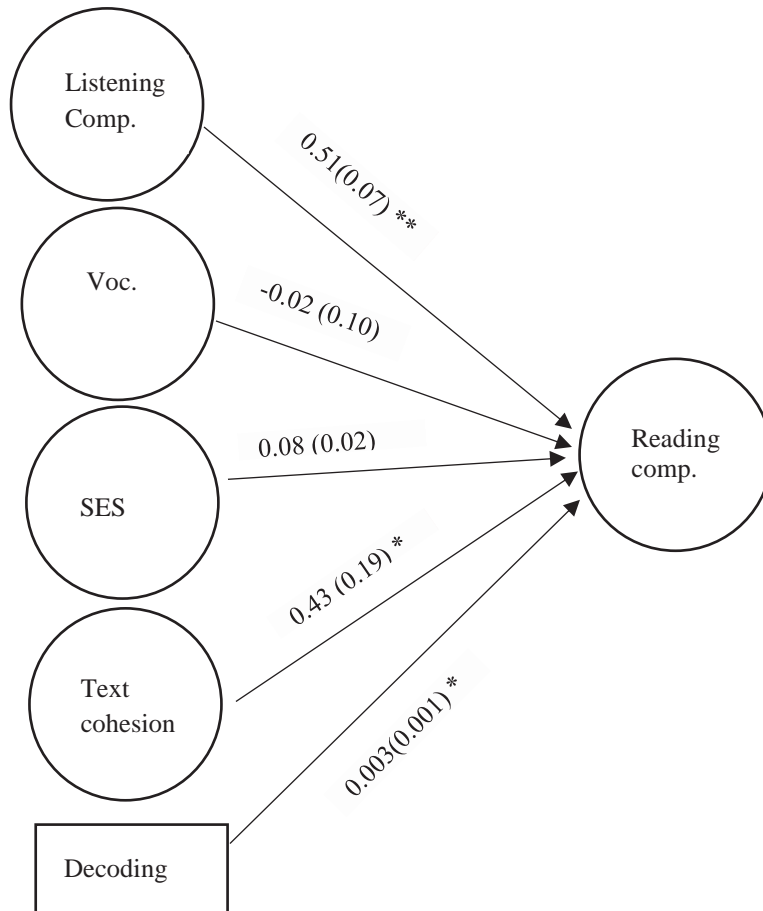
*Correlations between listening comprehension parcel 1-5, reading comprehension parcel 1-3, vocabulary parcel 1-2 and text cohesion vocabulary parcel 1-2*

Variable	1	2	3	4	5	6	7	8	9	10	11
1. lparcel1											
2. lparcel2	.78** [.73, .82]										
3. lparcel3	.66** [.58, .72]	.73** [.67, .78]									
4. lparcel4	.67** [.61, .73]	.72** [.66, .77]	.71** [.65, .76]								
5. lparcel5	.76** [.71, .81]	.77** [.71, .81]	.71** [.64, .76]	.73** [.67, .78]							
6. readparcel1	.57** [.48, .64]	.55** [.47, .63]	.47** [.37, .55]	.50** [.40, .58]	.51** [.42, .59]						

7.	readparcel2	.58**	.55**	.51**	.54**	.60**
		[.50, .66]	[.46, .63]	[.42, .60]	[.45, .62]	[.52, .67]
8.	readparcel3	.41**	.43**	.37**	.38**	.59**
		[.30, .50]	[.33, .52]	[.27, .47]	[.27, .47]	[.51, .66]
9.	vocparcel1	.37**	.37**	.39**	.32**	.43**
		[.27, .47]	[.27, .47]	[.29, .49]	[.21, .42]	[.33, .52]
					[.21, .42]	[.16, .37]
10.	vocparcel2	.39**	.42**	.38**	.39**	.44**
		[.28, .48]	[.32, .51]	[.28, .47]	[.29, .48]	[.34, .53]
					[.29, .48]	[.18, .39]
					[.24, .44]	[.51, .66]
11.	tcvparcel1	.25**	.22**	.22**	.22**	.32**
		[.14, .36]	[.11, .33]	[.10, .33]	[.11, .33]	[.24, .45]
					[.20, .42]	[.29, .49]
					[.20, .42]	[.33, .52]
12.	tcvparcel2	.21**	.23**	.26**	.25**	.30**
		[.09, .32]	[.11, .34]	[.15, .37]	[.14, .36]	[.18, .40]
					[.16, .38]	[.23, .44]
					[.16, .38]	[.25, .46]
					[.34**	.47**
					.36**	[.37, .56]

*Note.* Values in square brackets indicate the 95% confidence interval for each correlation. The confidence interval is a plausible range of population correlations that could have caused the sample correlation (Cumming, 2014). \* indicates  $p < .05$ . \*\* indicates  $p < .01$ .

Figure 2. Regression model predicting reading comprehension for the overall sample from listening comprehension, vocabulary, text cohesion vocabulary, SES and decoding.



Model fit (N=287).  $\chi^2[80.0] = 53.996$ ,  $p = 0.989$ , RMSEA = 0.000, CFI = 1.000, TLI = 0.937, \*\* p level = 0.001, \* p level = 0.05



In the final step, all regressions were constrained to be equal between the two groups and were compared to a model in which the regressions were freely estimated. The  $\chi^2$ - difference test showed no significant differences between the models ( $F(5.0) = 0.555, p = .734$ ). Thus, the two groups had equal strength in predictions from decoding and linguistic skills to reading comprehension.

### **Discussion**

This study reveals several interesting findings about reading and language in bilingual children with an early age of acquisition. First, the study finds that early bilingual learners of (primarily) middle to high SES have levels of decoding skills and text cohesion vocabulary similar to those of their monolingual peers, but there are moderate to large differences in favour of monolingual learners in listening comprehension, reading comprehension and vocabulary. For the predictive patterns of linguistic skills to reading comprehension, the size of the predictive paths was equal across groups. This implies that linguistic skills do not play a more critical role for the early bilingual children than for the monolingual reader. Listening comprehension, text cohesion vocabulary and decoding were the only constructs related to reading comprehension, with listening comprehension explaining the largest proportion in reading comprehension.

#### ***Levels of language and reading comprehension skills compared to monolingual peers***

The results of this study support prior research findings that bilingual children have poorer language comprehension skills than monolingual children (Bialystok, 2009; Melby-Lervåg & Lervåg, 2014; Oller, 2005; Snow & Kim, 2007). The sample of bilingual children in the current study attended Norwegian ECEC institutions from at least 2 years of age. The Norwegian ECEC institutions are regulated by The Norwegian Framework Plan for content and tasks (Kunnskapsdepartementet, 2017).

This plan underlines the importance of ECEC staff to ensure that every child takes part in activities known to promote their communication and language skills. Since early bilingual learners are richly exposed to instructional language and are mainly of middle to high SES, one should expect the gap in instructional language performance to be less than that for minority-language learners. However, the gap identified in the present study is comparable to gaps found in some of the more optimistic studies of minority-language learners exposed to L2 in 5-7 years starting from just before school entry. These studies often find a group difference of approximately 1 SD unit in favour of monolingual children (Cummins, 1984, 2017; Hakuta, Butler, & Witt, 2000; Saunders & O'Brien, 2006). In the current study, the largest gap was found in reading comprehension (0.78 Cohens *d*), followed by vocabulary (0.74 Cohens *d*) and then listening comprehension (0.60 Cohens *d*). These results are similar to the gaps identified in the studies of early bilingual learners' reading comprehension and vocabulary skills (Bialystok & Feng, 2011; Bonifacci & Tobia, 2016; Grant et al., 2011). This indicates that even early bilingual learners with middle to high SES lag behind their monolingual peers in language and literacy skills, despite long and rich exposure to the instructional language. This suggests that the distributed profile effect might also affect the language and literacy skills of early bilingual learners.

The results from the current study are, however, at odds with the study of Bonifacci and Tobia (2016). They found no significant differences across language groups on listening comprehension. Both the present study and the one by Bonifacci et al. (2016) are studies of mixed samples of bilingual learners at birth and early bilingual learners. Interestingly, the early bilingual learners in Bonifacci et al. (2016) both have a higher AoA than the sample in the present study and were tested on listening comprehension at a younger age. Thus, early bilingual learners in Bonifacci et al. (2016) were less exposed to instructional language than were early bilingual learners in the present study. However, Bonifacci and Tobi (2016) controlled for SES on the

residential level only and compared levels across language groups on sum scores; hence, it might be factors other than AoA that cause the non-significant difference in listening comprehension.

Furthermore, the results of the Early Childhood Longitudinal Study, Kindergarten Class of 1998-1999 (ECLS-K), did, however, find bilingual children proficient in instructional language by kindergarten entry (6 years) to develop reading comprehension skills at a similar level as monolingual peers in 6<sup>th</sup> – 8<sup>th</sup> grade (Halle et al., 2012), which suggests that some bilingual groups have other language and literacy trajectories than the one in the present study. However, the ECLS-K study did not report AoA or other sample characteristics that could cast light on potential factors that influence reading comprehension in early bilingual children. It could be that the AoA and amount of exposure of the instructional language in the present sample were insufficient to produce proficiency by the age of 6. It could also be that these are artefactual findings due to the lack of invariance and that differences could be caused by comparisons across different constructs of reading comprehension. In fact, test descriptions show that the measures used to assess reading comprehension in the ECLS-K study were more dependent on decoding skills than the construct in the present study was (Nation, 2006; Pollack, Najarian, Rock, & Atkins-Burnett, 2005).

The study of Hwang, Lawrence, and Snow (2017), where early bilingual learners outperformed their monolingual peers in vocabulary levels, also contrasts with the results of the present study. Notably, even though the need to control for SES in studies of bilingual learners is widely accepted, the study of Hwang et al. (2017) included children enrolled in gifted and talented education programs. These children were not distributed equally in the two groups: fifty-one percent of bilingual learners were enrolled in gifted and talented education programmes compared to 35% of monolingual learners. Nevertheless, SES was only controlled for by using a dummy for the number of participants receiving free lunch, which is a rather crude measure. With all studies taken together, it seems more likely that when comparing groups of

monolingual and early bilingual children from similar SES levels on linguistic skills and reading comprehension assessed by tests less dependent on decoding skill, differences across language groups in favour of monolingual learners will surface. Hence, the profile effect also seems to affect early bilingual children.

However, not all language skills in the present study are equally affected. Early bilingual learners have text cohesion vocabulary levels that are equal to those of their monolingual peers ( $d = 0.34$ ,  $p = 0.12$ ). This finding is surprising because text cohesion typically correlates moderately with vocabulary in minority-language learners (Rydland, Aukrust, & Fulland, 2012), and the bilingual sample in the current study also has considerably lower vocabulary levels than their monolingual peers do. One possible explanation for the results of the present study could be that text cohesion vocabulary comprises a limited number of words in contrast to vocabulary as a whole; hence, text cohesion vocabulary might be less affected by the distribution profile effect. An early and rich exposure of instructional language could then be enough to lift the early bilingual learners' levels of text cohesion to match their monolingual peers. Here, the contrast to the level of text cohesion vocabulary of minority-language learners introduced to L2 at a later timepoint is large. Furthermore, prior studies find large gaps in text cohesion vocabulary in disfavour of minority-language learners with higher AoA and less exposure to the instructional language than the present sample (Droop & Verhoeven, 2003). This implies that the text cohesion aspect of linguistic skills is, as suggested by Crosson, Lesauc, and Martiniello (2008), especially important to target minority language learners. Notably, these conflicting findings suggest, not surprisingly, that minority-language learners and early bilingual learners have different needs for intervention.

For morphology, a large number of test items were invariant; hence, a comparison of morphological levels on latent means across language groups was invalid. Given the results of this study, one can question prior studies using sum scores in analysis and interpret the

results as an indication for morphonology to be a moderate strength in bilingual learners (Barac & Bialystok, 2012; Hsu et al., 2019; Lipka & Siegel, 2012). The same argument can also be used for studies examining morphology as a unique predictor of reading comprehension (Vernice & Pagliarini, 2018).

### **The predictive patterns for aspects of language comprehension and decoding skills to reading comprehension**

In the current study, the predictive pattern from language skills to reading comprehension was equal for monolingual and early bilingual learners. In contrast, in Grant et al. (2011), vocabulary predicted bilingual 3<sup>rd</sup> graders' reading skills, but only decoding predicted their monolingual peers' reading comprehension. The early bilingual learners in both the present study and in Grant et al. (2011) had comparably lower levels than their monolingual peers in linguistic skill. Thus, the deviating predictive pattern in these studies is unlikely to be because the language levels in bilingual learners differed between the two studies. Vocabulary is, however, sometimes found to predict reading comprehension at an earlier age for bilingual learners with lower levels of L2 skills than for their monolingual peers (Kieffer & Vukovic, 2013; Limbird, Maluch, Rjosk, Stanat, & Merkens, 2014). This is perhaps also the case for early bilingual learners. However, as soon as linguistic skills take over as the dominating predictor of reading comprehension in monolingual children, perhaps an equal predictive pattern of language skills to reading comprehension for the early bilingual and monolingual reader emerges.

Regarding the impact on early bilingual learners' reading comprehension, listening comprehension explained 26.01% of the variance in the present study, in contrast to vocabulary that did not contribute to reading comprehension after listening comprehension was controlled for. This is at odds with previous research on minority language learners. Comparisons across studies are, however, difficult due to differences in methodology. Prior studies have sometimes not included both vocabulary and listening comprehension as measures

(Grant et al., 2011; Hutchinson et al., 2003), used sum scores in the analysis allowing random noise and variant items to play a role in the analysis (Burgoyne et al., 2011), or investigated growth in reading comprehension of 1st graders (Kieffer, 2012) rather than the concurrent prediction of reading comprehension in 5<sup>th</sup> grade. It is also worth noting that even though vocabulary skill in the present study did not explain any proportion of the variance in reading comprehension, text cohesion vocabulary explained 18.49% of the variance in reading comprehension. This is in line with Rydland et al. (2012) and suggests that text cohesion vocabulary contributes to reading comprehension in some texts. Thus, the impact of text cohesion vocabulary on preadolescent children's reading comprehension seems to hold across minority-language learners and early bilingual and monolingual readers. Hence, listening comprehension and text cohesion vocabulary seem important to improve early bilingual preadolescents' reading comprehension.

***Practical implications and limitations***

The main finding from the present study is that early bilingual learners still lagged behind their monolingual peers on most aspects of language and reading comprehension, even after 8-10 years of exposure in instructional language. Thus, despite years of enrolment in educational settings that facilitate the participation of children in activities that promote language skills, the gap to their monolingual peers has not been closed. This implies that medium to high SES background, early AoA, and rich exposure to the instructional language are not by themselves sufficient for early bilingual learners to develop language levels similar to those of their monolingual peers on all aspects of the instructional language. Thus, there is a need for the development of interventions to ensure the improvement of language and literacy trajectories for the early AoA subgroup of bilingual learners. The present study found comparable text cohesion vocabulary levels in early bilingual and monolingual children, implying that linguistic constructs other than text cohesion should be targeted in such an intervention. Note, however, that the

reliability of text cohesion in the present study was low. Thus, low reliability could perhaps have disguised a true difference in text cohesion levels.

As for the prediction of specific L2 skills on reading comprehension, even though the methodological approach of the present study is an improvement to that of most prior studies investigating the unique prediction of language skills on reading comprehension, very few of the cited studies (including the present one) have examined dimensionality in the 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. Therefore, the extent to which different linguistic skills are differently related to reading comprehension for early bilingual learners, minority-language learners and monolingual readers needs to be explored further. 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 uplifting results, which suggests that such interventions could change young bilingual children's learning trajectories (Rogde, Melby-Lervåg, & Lervåg, 2016).

## References

- Babayigit, 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.
- 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.
- Bialystok, E. (2009). Bilingualism: The good, the bad, and the indifferent. *Bilingualism: Language and Cognition, 12*(1), 3-11.
- Bialystok, E., & Feng, X. (2011). Language proficiency 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). New York: The Guildford Press.
- Bialystok, E., Luk, G., Peets, K., F., & Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism-Language and Cognition, 13*(4), 525-531.
- 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.
- 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*. New York: 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.
- Crosson, A. C., Lesauc, N. K., & Martiniello, M. (2008). Factors that influence comprehension of connectives among language minority children from Spanish-speaking backgrounds. *Applied Psycholinguistics*, 29, 603-625.
- Cruz-Ferreira, M. (2006). *Three is a crowd?: Acquiring Portuguese in a trilingual environment* (Vol. 6): Multilingual Matters.
- Cummins, J. (1984). *Bilingualism and special education. issues in assessment and pedagogy*. Clevedon: Multilingual Matters.
- Cummins, J. (2017). *Flerspråkiga elever: effektiv undervisning i en utmanande tid*: Natur & Kultur Akademisk.
- De Houwer, A. (2009a). *Bilingual first language acquisition*: Multilingual Matters.
- De Houwer, A. (2009b). *An introduction to bilingual development*: Multilingual Matters.
- De Houwer, A. (2012). Milestones in bilingual children's development In *Encyclopaedia of Language and Literacy Development* (pp. 1-8). London, Canada: Western Ontario University.
- De Houwer, A., Bornstein, M. H., & Putnick, D. L. (2014). A bilingual-monolingual comparison of young children's vocabulary size: Evidence from comprehension and production. *Applied Psycholinguistics*, 35(6), 1189-1211.
- Droop, M., & Verhoeven, L. (2003). Language proficiency and reading ability in first- and second-language learners. *Reading Research Quarterly*, 38(1), 78-103.

- 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.
- Friesen, D. C., & Bialystok, E. (2012). Metalinguistic Ability in Bilingual Children: The Role of Executive Control. *Rivista di psicolinguistica applicata, 12*(3), 47-56.
- Geva, E. (2007). *Conjunction use in School Children's Oral Language and Reading*. In R. Horowitz (Ed.), *Talking Texts. How Speech and Writing Interact in School Learning* (1st ed.).
- Geva, E., & Farnia, F. (2012). Developmental changes in the nature of language proficiency and reading fluency paint a more complex view of reading comprehension in ELL and EL1. *Reading and Writing, 25*(8), 1819-1845.
- 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.
- Hakuta, K., Butler, Y. G., & Witt, D. (2000). How Long Does It Take English Learners To Attain Proficiency?
- 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.
- Hammer, C. S., Hoff, E., Uchikoshi, Y., Gillanders, C., Castro, D. C., & Sandilos, L. E. (2014). The language and literacy development of young dual language learners: A critical review. *Early Childhood Research Quarterly, 29*(4), 715-733.
- 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., Lervåg, A., Lyster, S.-A. H., Hagtvet, B. E., Hulme, C., & Melby-Lervåg, M. (2018). Pathways to reading comprehension: A longitudinal study from 4 to 9 years of age. *Journal of Educational Psychology*.
- Hoff, E., Rumiche, R., Burrige, A., Ribot, K. M., & Welsh, S. N. (2014). Expressive vocabulary development in children from bilingual and monolingual homes: A longitudinal study from two to four years. *Early Childhood Research Quarterly*, 29(4), 433-444.
- Hoff, E., Welsh, S., Place, S., Ribot, K., Grüter, T., & Paradis, J. (2014). Properties of dual language input that shape bilingual development and properties of environments that shape dual language input. *Input and experience in bilingual development*, 13, 119-140.
- 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.
- 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.
- 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. doi: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. doi:10.1007/s11145-016-9703-3
- Høien, T., & Tønnesen, G. (2008). *Introduksjonshefte til Ordkjedeprøven*. Bryne: Logometrica AS.
- Jeon, E. H., & Yamashita, J. (2014). L2 reading comprehension and its correlates: A meta-analysis. *Language learning, 64*(1), 160-212.
- Kieffer, M. J. (2008). Catching up or falling behind? Initial English proficiency, concentrated poverty, and the reading growth of language minority learners in the United States. *Journal of Educational Psychology, 100*(4), 851-868.
- 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.
- Kieffer, M. J., Biancarosa, G., & Mancilla-Martinez, J. (2013). Roles of morphological awareness in the reading comprehension of Spanish-speaking language minority learners: Exploring partial mediation by vocabulary and reading fluency. *Applied Psycholinguistics, 34*(4), 697-725.
- Kieffer, M. J., & Lesaux, N. K. (2012). Knowledge of words, knowledge about words: Dimensions of vocabulary in first and second language learners in sixth grade. *Read. Writ., 25*(2), 347-373. doi:10.1007/s11145-010-9272-9
- Kieffer, M. J., & Vukovic, R. K. (2013). Growth in reading-related skills of language minority learners and their classmates: more

- evidence for early identification and intervention. *Reading and Writing*, 26(7), 1159-1194.
- 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
- Leider, C. M., Proctor, C. P., Silverman, R. D., & Harring, J. R. (2013). Examining the role of vocabulary depth, cross-linguistic transfer, and types of reading measures on the reading comprehension of Latino bilinguals in elementary school. *Reading and Writing*, 26(9), 1459-1485. doi:10.1007/s11145-013-9427-6
- 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.
- Lesaux, N. K., Rupp, A. A., & Siegel, L. S. (2007). Growth in reading skills of children from diverse linguistic backgrounds: Finding from a five-year longitudinal study. *Journal of Educational Psychology*, 99, 821-834.
- Limbird, C. K., Maluch, J. T., Rjosk, C., Stanat, P., & Merkens, H. (2014). Differential growth patterns in emerging reading skills of Turkish-German bilingual and German monolingual primary school students. *Reading and Writing*, 27(5), 945-968.
- Lipka, O., & Siegel, L. (2012). The development of reading comprehension skills in children learning English as a second language. *Reading and Writing*, 25(8), 1873-1898.
- 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.

- Mather, N., Hammil, E. E., Allan, E. A., & Roberts, R. (2014). *Test of silent reading fluency*. Austin: TX: PRO- ED.
- 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.
- Monsrud, M.-B., Rydland, V., Geva, E., Thurmann-Moe, A. C., & Lyster, S.-A. H. (2019). The advantages of jointly considering first and second language vocabulary skills among emergent bilingual children. *International Journal of Bilingual Education and Bilingualism, 1*-17.
- Muthén, L., & Muthén, B. (1998). Mplus (Version 8.4) [computer software]. Los Angeles, CA: 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. doi:10.1177/0022219413509967
- Nation, K. (2006). Assessing children's reading comprehension. In M. J. Snowling & J. Stackhouse (Eds.), *Dyslexia, Speech and Language. A Practitioner's Handbook*. (2. ed.).
- Neale, M. D. (1997). *Neale Analysis of Reading Ability- Revisted. Second revised British Edition*. London: 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. (2005). The distributed Characteristic in Bilingual Learning. *International Symposium on Bilingualism, 1744*-1749.
- 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 Durgunoglu & Goldberg (Eds.), *Language and literacy development in bilingual settings*. New York: The Guilford Press.
- Oller, D. K., Pearson, B. Z., & Cobo-Lewis, A. B. (2007). Profile effects in early bilingual language and literacy. *Applied Psycholinguistics*, 28(02), 191-230.
- Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika*, 44(4), 443-460.
- Pallant, J. (2020). *SPSS survival manual: A step by step guide to data analysis using IBM SPSS*: Routledge.
- Place, S., & Hoff, E. (2011). Properties of Dual Language Exposure That Influence 2-Year-Olds' Bilingual Proficiency. *Child Development*, 82(6), 1834-1849.
- Place, S., & Hoff, E. (2016). Effects and noneffects of input in bilingual environments on dual language skills in 2 1/2-years-olds. *Bilingualism: Language and Cognition*, 19(5), 1023-1041.
- Pollack, J. M., Najarian, M., Rock, D. A., & Atkins-Burnett, S. (2005). Early Childhood Longitudinal Study, Kindergarten Class of 1998? 99 (ECLS-K). Psychometric Report for the Fifth Grade. NCES 2006? 036. *National Center for Education Statistics*
- Proctor, C. P., Carlo, M., August, D., & Snow, C. (2005). Native Spanish-speaking children reading in English: Toward a model of comprehension. *Journal of Educational Psychology*, 97(2), 246.
- 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*. Boston, UNITED STATES: 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.
- 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.
- 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 Modeling: A Multidisciplinary Journal, 21*(2), 167-180.
- Saunders, W. M., & O'Brien, G. (2006). Oral language. *Educating English language learners: A synthesis of research evidence*, 14-63.
- Schwartz, M., & Katzir, T. (2012). Depth of lexical knowledge among bilingual children: the impact of schooling. *An Interdisciplinary Journal, 25*(8), 1947-1971.
- Silverman, R. D., Proctor, C. P., Harring, J. R., Hartranft, A. M., Doyle, B., & Zelinke, S. B. (2015). Language skills and reading comprehension in English monolingual and Spanish-English bilingual children in grades 2-5. *Reading and Writing, 28*(9), 1381-1405.
- Snow, C. E., & Kim, Y.-S. (2007). Large problem spaces: The challenge of vocabulary for English language learners.



- Team, R. C. (2019). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>.
- Verhallen, M., & Schoonen, R. (1993). Lexical knowledge of monolingual and bilingual children. *Applied linguistics*, 14(4), 344-363.
- Verhoeven, L. (2000). Components in early second language reading and spelling. *Scientific Studies of Reading* (4), 313-330.
- Verhoeven, L., & van Leeuwe, J. (2012). The simple view of second language reading throughout the primary grades. *Reading and Writing*, 25(8), 1805-1818.
- 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*. Stavanger: Høgskolen i Stavanger.
- Wechsler, D. (2003). *WISC- IV: Wechsler Intelligence Scale for Children*: Pearson inc.

## Appendices

### **Appendix 1. Study characteristics of primary studies in reviews where the authors examine bilingual learners' language levels after 5-7 years of L2 exposure**

Table 1. Study characteristics of the targeted bilingual subgroup, information regarding what their language levels are compared to, and the bilingual learners' L2 outcomes in the primary studies included in the reviews examining bilingual learners' proficiency levels after 5-7 years

Name of review	Examined bilingual subgroup	Compared to	Outcome
Hakuta et al (2000)	Minority language learners.	Norms and different bilingual proficiency tests. Age gaps to monolingual language levels	Lower levels of norms, SD differences from study to study (e.g. -0.5 SD, -0.75 SD, 2 years behind norms of native speakers)
Collier (1989)	Minority language learners.	*	
Collier (1987)	Minority language learners	Norms	50 <sup>th</sup> percentile
Cummins (1984, 2017)	Minority language learners.	Norms, control groups, bilingual proficiency tests	Lower levels of norms. (e.g., often around -1 SD), 50 <sup>th</sup> percentile;
Saunders and O'Brain (2006)	Minority language learners and immersion students	Norms, bilingual proficiency tests, age gaps to monolingual language levels	lower levels of norms (often scores -1 SD, or ranging from 32 <sup>nd</sup> -58 <sup>th</sup> percentile); 2 years behind norms of native-speakers

*Note.* \* A review of several topics related to bilingualism. For language outcomes after 5-7 years of L2 exposure, little information on the primary studies is available.

Bilingual proficiency tests: different tests to measure levels of L2 proficiency, NB, not compared to norms of native language performance.

**Appendix 2. The primary studies included in De Houwer’s (2009a, 2009b, 2012) reviews**

The table below includes studies comparing the language levels of monolingual and *bilingual first learners*. Studies comparing *bilingual first learners’* language levels to theories of monolingual language development or studies purely comparing the connection between L1 and L2 development in bilingual samples are excluded from the list below. Note that even though many sources are referenced in more than one of the reviews, they are listed in the table only once. Note also that De Houwer does not claim that all these studies find equal language levels between bilingual and monolingual learners. Some of the primary studies examine the language levels of *early bilingual learners* mixed with *bilingual first learners*. Several of these studies find a gap in language levels between these *early bilingual learners* and their monolingual peers. According to De Houwer, this gap would not be found if the group of bilingual learners consisted of only *bilingual first learners*.

Table 2. Overview of the study characteristics describing participant age, sample size, longitudinal status and type of comparison study for the sources referenced in De Houwer’s reviews (2009a, 2009b, 2012)

An introduction to bilingual development (2009)				
Source	Participant age	Compared to	N	Longitudinal
Pearson et al. (1993)*	8-30 months	Norms	Bilingual: 20	No
Cruz-Ferreira (2006)	Followed to the age of 10		Case study of triplets	Yes.

*Appendices*

Deuchar & Quay (2000)	Followed from 1:7 to 2:3		Case study	Yes
Pearson & Fernandez (1994)	8-30 months *	Norms	Bilingual: 20	No
Vila (1984)	Unpublished manuscript in Spanish			
Wanner (1996)	Not identified through searches			
De Houwer (2005)	Source not found in reference list			
Milestones in bilingual children's development (2012)				
Source	Participant age	Compared to	N	Longitudinal
De Houwer, Bornstein & Putnic (2014)	From 13 to 20 months	Control group	Monolingual: 30 Bilingual: 31	Yes
Hoff, Core, Place, Rumchie, Senor & Parra (2012)	From 1:1 to 2:6	Control group	Monolingual: 57 Bilingual: 47	Yes
Nakamura, Quay (2012)	From 1:3 to 2:3		Case study	Yes
Patterson (1998)	21-27 months	Primary article states comparison to monolingual norms is not appropriate due to adaptation of assessment tool	Bilingual: 102	No
Bilingual first learners' acquisition (2009)				

*Appendices*

Chapter 6-7				
Source	Participant age	Compared to	N	Longitudinal
Pearson (1993)	8-16 months	Control group	Monolingual: 12 Bilingual: 10	No
Umbell et al (1992)	5:11-8:6 years	Norms	Bilingual: 105	No
Sundara et al (2006)	4-year-olds	Control group	Monolingual: 24 Bilingual: 12	No
Genesee 2001	3- and 4-year-olds	Norms	Bilingual 3-year-olds: 12 Bilingual 4-year-olds: 6	No
Perez- Pereira (2008)	30 months	Control group	Bilingual: 431 Monolingual: 275	No
Thordardottir et al (2006)	32 months	Control group	Bilingual: 8 Monolingual: 10	No
De Houwer (2006) poster presentation, later published in De Houwer, Bornstein & Putnic (2014) (personal communication, with De Houwer)	From 13 to 20 months	Control group	Monolingual: 30 Bilingual: 31	Yes
Barrena et al (2008)	16-30 months	Control group	Bilingual: 209 dominant,	No

*Appendices*

			150	not
			dominant	
			Monolingual:	
			588	
Junker, Stockman (2002)	24 months	Control group	Bilingual: 10	No
			Monolingual:	
			10	
Rimel & Eyal (1996)	18-30 months	Control group	Bilingual: 19	No
			Monolingual:	
			20	
Nicoladis et al (2007)	5-year- olds	Control group	Bilingual: 10	No
			Monolingual:	
			20	
Lanza (2001)	4:6, 6:10 and 8:8 year-olds		Case studies	No
Silva- Corvalan (2003) **	2.5-5:8		Case study.	Yes
			N= 2	
Serratrice (2007b)	8-year- olds	Control group	Bilingual learners: 12,	No
			monolingual learners: 24	

*Note.* \* Studies claim that the total vocabulary levels of bilingual learners fell within monolingual norms (defined as between the 10<sup>th</sup> and 90<sup>th</sup> percentiles). Re-analysis by Bialystok (2001a) finds that the total vocabulary of 8 of 18 bilingual learners in these studies was below the 10<sup>th</sup> percentile. Their language levels in their strongest language were constantly below the monolingual norms. \*\* The study compares data from two case studies with two studies of monolingual children. One of the papers is available only in Spanish, and the other traces two monolingual children longitudinally.

**Appendix 3. Summary of study characteristics of prediction studies of reading comprehension**

Table 3. Study characteristics of studies examining the predictive pattern of multiple linguistic skills for reading comprehension for bilingual children with a minimum of 4 years of exposure to their second language.

Study names	Investigated constructs	Analytical approach	Bilingual <i>N</i> , Monolingual <i>N</i>
Burgoyne et al. (2011)	Listening comprehension, vocabulary	Hierarchical multiple regression of sum-scores	Monolingual <i>N</i> =39, Bilingual <i>N</i> =39
Geva and Farnia (2012)	Vocabulary, listening comprehension	Hierarchical multiple regression of sum-scores	Monolingual <i>N</i> =149, Bilingual <i>N</i> = <i>N</i> = 539
Hutchinson et al. (2003)	Listening comprehension, morpho-syntax (TROG), vocabulary	Multiple regression of sum-scores	Monolingual <i>N</i> = 43, Bilingual <i>N</i> = 43
Leider et al. (2013)	Vocabulary, morphology,	Multiple regression of sum- scores	Bilingual <i>N</i> = 123
Rydland et al. (2012)	Vocabulary, text cohesion vocabulary	Multiple regression of sum-scores	Bilingual <i>N</i> = 67
Kieffer & Lesaux, 2008	Morphology, vocabulary	Multiple regression of sum-scores	Bilingual <i>N</i> = 87 <b>NB-immersion study</b>
Kieffer, Biancarosa, et al. (2013)	Morphology, vocabulary, listening comprehension	Growth model with latent factor of reading comprehension and linguistic predictors as sum-scores	Bilingual <i>N</i> = 101
Proctor et al. (2012)	Vocabulary, morphology	Growth model with latent factor of	Monolingual <i>N</i> =165,

*Appendices*

		reading comprehension and linguistic predictors as sum-scores	Bilingual $N = 129$
Kieffer, 2012	Vocabulary, listening comprehension	Latent growth model of reading comprehension. Examines impact of latent SES with linguistic predictors as sum scores	Bilingual $N = 295$
Silverman, Proctor, Harring, Hartranft, Doyle and Zelinke, 2015	Vocabulary, morphology	Growth model with latent factor of reading comprehension, CFA tested linguistic predictors, yet not investigated for invariance	Monolingual $N = 212$ , Bilingual $N = 174$
Droop and Verhoeven, 2003	Vocabulary, morphological knowledge, listening comprehension, text cohesion vocabulary	Prediction examined by latent means in a SEM model, invariance tested across timepoint, thereafter separately fitted for the two language groups.	Monolingual $N = 143$ , Bilingual $N = 102$

***Appendix 4. Additional analysis of language skills between different subgroups of bilingual toddlers and their monolingual peers***

Additional group comparison analysis of differences in language skills between different subgroups of *early bilingual* toddlers and their monolingual peers.



Article 2 found no difference in language comprehension between *toddlers who were exposed mainly to L2* by their parents (median = 7.13, SD = 2.52) and *toddlers with parents exposing them to both Norwegian and L1* (median 7.12, SD = 2.92). There were, however, differences in language levels between these two groups, and the subgroup that was *exposed mainly to L1* at home. The *bilingual learners exposed mainly to L1* had lower levels of language comprehension than the other two bilingual groups. Thus, the scores of the first two subgroups increased the mean average effect size difference identified between the monolingual and *early bilingual learners* examined in research question 1, article 2 ( $d = 0.54$ ). By merging the two groups with *some L2 exposure* into one group (*toddlers who were exposed mainly to L2* and *toddlers with parents exposing them to both Norwegian and L1*), it is possible to calculate an effect size difference between this new subgroup ( $M = 6.69$ ,  $SD = 2.98$ ,  $N = 112$ ) and its monolingual peers ( $M = 8.32$ ,  $SD = 3.01$ ,  $N = 902$ ). A t-test examining differences between monolingual and *bilingual learners with some L2 exposure* at home showed significant differences in levels of language comprehension ( $t(1012) = -3.977$ ,  $p = 0.000$ ). The effect size difference is small ( $d = 0.389$ ) and favours the monolingual learners. Due to large differences in sample sizes, the language levels of monolingual learners ( $N = 902$ ) and bilingual learners exposed mainly to L1 at home were not examined ( $N = 49$ ). However, Study 2 reports differences in language levels between *toddlers exposed mainly to Norwegian (L2) at home* and *toddlers exposed mainly to L1*, and the difference approaches a medium effect size ( $r = .27$ ,  $p = 0.14$ ). Using Psychometrica's effect size calculator, this equals a  $d$  of 0.56 ([https://www.psychometrica.de/effect\\_size.html#transform](https://www.psychometrica.de/effect_size.html#transform)). The difference in L2 skills was slightly less between *toddlers who are exposed mainly to L1* at home (median = 5.69, SD = 3.17) and *toddlers who are exposed to L1 and Norwegian at home* ( $r = .26$ ,  $d = 0.54$ ,  $p = 0.003$ ). Since the subgroup *exposed mainly to L1* at home has lower language comprehension levels than its bilingual peers, who again have lower language levels than their monolingual peers, all the bilingual

subgroups examined in article 2 have lower language levels than the monolingual learners.

***Appendix 5. Information about Study 3 sent to parents accompanied by a consent form for their children's participation in the study.***

***Information letter sent to the 2006 birth class***

To parents/guardians



**Information about Extension of the Stavanger Project**

We are very grateful for the opportunity we were given to follow Your child's development from the age of two and through 4<sup>th</sup> grade of primary school. The first stage of the project focussed on the child's development of linguistic, mathematical and social skills, in addition to the development of motor function. The second stage, which was prompted by the onset of formal schooling, focussed on the child's development of reading, writing and mathematical skills. We wish to extend the project in order to collect more data about multilingual children's oral Norwegian proficiency.

Little research has been conducted on multilingual children in a Norwegian context. We know that children's reading and oral proficiencies are intertwined and that multilingual children generally struggle more in developing adequate skills in the reading and speaking of Norwegian than do monolingual speakers of Norwegian. We also know that the gap between multilingual learners who succeed in school and multilingual learners who struggle in school is wider than in the monolingual population. One way to prevent children from falling behind in their linguistic development is by early intervention. However, in order to find out how best to assist multilingual children in their linguistic development, we need further research. We do not, for example, have sufficient knowledge about factors that support multilingual children in their formal education. In order to obtain more

knowledge about such factors, it could prove useful to study the linguistic development of multilingual children who have attended Norwegian kindergartens from the age of two. Finding out to what extent multilingual children's oral proficiencies influence their development of reading skills is another aspect of interest. Obtaining more knowledge about this group of children would be valuable in terms of accommodating for multilingual learners in the Norwegian educational system.

The need for more research in the field has prompted an extension of the Stavanger Project. The next stage of the research study will involve Your child undergoing a battery of three individual assessments (all within the same day) during the spring semester of 5<sup>th</sup> grade. The assessment aims at mapping out the child's listening and reading skills, in addition to its understanding of the depth of the Norwegian vocabulary. The assessment of the child's oral Norwegian proficiency will be audio recorded. The assessments will take place in school. Parents/guardians need not attend the assessment as experienced staff from the Reading Centre will guide the child through it. As part of the extension of the study, we also kindly request that the parents/guardians respond to a short survey about the family and the child's exposure to the Norwegian language.

The entering of participants into the study is based on informed consent from parents/guardians, and participation is voluntary. Information about the participants will be anonymised; hence preventing the identification of individual children. The audio recordings and the identification codes will be erased upon the completion of the project. The design of the study complies with the guidelines of the Norwegian Data Protection Official for Research. The findings will be used in preschool teacher training and teacher education programmes, and they will be published in research articles. If You wish to enter Your child into the extended Stavanger Project, please fill in the form of Informed Consent and send it to the Reading Centre in the enclosed envelope. Should you need more information, please do not hesitate to contact [hilde.gunnerud@uis.no](mailto:hilde.gunnerud@uis.no).

Sincerely,

Elin Reikerås  
Project Supervisor

Hilde Gunnerud  
PhD Student

**Information letter sent to the 2007 birth class.**

To parents/guardians



**Information about Extension of the Stavanger Project**

We are very grateful for the opportunity we were given to follow Your child's development from the age of two and through 4<sup>th</sup> grade of primary school. The first stage of the project focussed on the child's development of linguistic, mathematical and social skills, in addition to the development of motor function. The second stage, which was prompted by the onset of formal schooling, focussed on the child's development of reading, writing and mathematical skills. We wish to extend the time frame of the project in order to collect more data about multilingual children's oral Norwegian proficiency.

We know that reading is a key skill in order for children to succeed at school, and that oral proficiency in Norwegian is of important influence on children's reading comprehension. Pupils with delayed language development and multilingual pupils are particularly at risk, including the fact that multilingual children as a group often have greater difficulties than monolingual children when it comes to Norwegian reading comprehension and oral skills. The gap between multilingual learners who succeed in school and multilingual learners who struggle, is wider than in the monolingual population. However, we lack knowledge on the linguistic development of multilingual children who have attended Norwegian kindergartens from the age of two. Are their oral skills and their reading comprehension in Norwegian comparable to their monolingual peers by the time they reach Grade 5? Also, it is not clear whether some aspects of the oral language play a more important role in relation to reading comprehension than other aspects, and whether this relationship is the same for monolingual and multilingual pupils.

## *Appendices*

---

Obtaining this knowledge is valuable in order to accommodate for more pupils to succeed in school, and, among other things, to clarify whether multilingual pupils and monolingual pupils with weak oral skills in Norwegian can benefit from similar measures.

The need for more research in the field has prompted an extension of the Stavanger Project. The next stage of the research study will involve Your child undergoing a battery of three individual assessments (all within the same day) during the spring semester of 5<sup>th</sup> grade. The assessment aims at mapping out the child's listening and reading skills, in addition to its understanding of the depth of the Norwegian vocabulary. The assessment of the child's oral Norwegian proficiency will be audio recorded. The assessments will take place in school, and parents/guardians need not attend. The assessment will be undertaken by experienced research assistants who have received training from the Reading Centre. As part of the extension of the study, we also kindly request that the parents/guardians respond to a short survey about the family and the child's exposure to the Norwegian language.

The entering of participants into the study is based on informed consent from parents/guardians, and participation is voluntary. Information about the participants will be anonymised; hence preventing the identification of individual children. The audio recordings and the identification codes will be erased upon the completion of the project. The design of the study complies with the guidelines of the Norwegian Data Protection Official for Research. The findings will be used in preschool teacher training and teacher education programmes, and they will be published in research articles. If You wish to enter Your child into the extended Stavanger Project, please fill in the form of Informed Consent and send it to the Reading Centre in the enclosed envelope. Should you need more information, please do not hesitate to contact [hilde.gunnerud@uis.no](mailto:hilde.gunnerud@uis.no).

Sincerely,

Elin Reikerås  
Project Supervisor

Hilde Gunnerud  
Ph.D. Student

## Appendix 6. Parent questionnaire sent to participants in Study 3

### Spørreskjema til foreldre

#### Bakgrunnsinformasjon:

Mors alder nå: .....  Mor er født i Norge.  Mor er født i et annet land.

Hvis mor er født i et annet land, når kom mor til Norge? .....

Hva er mors morsmål? .....

Fars alder nå: .....  Far født i Norge  Far født i annet land

Hvis far er født i et annet land, når kom far til Norge? .....

Hva er fars morsmål? .....

#### Barnets erfaringer med det norske språket

1. Hvor gammelt var barnet når det begynte i (norsk) barnehage?  
 1 år  2 år
2. Hvor lenge har barnet gått på SFO?  
 Barnet har ikke gått på SFO  
 1 år  
 2 år  
 3 år  
 4 år
3. Hvor omgås barnet jevnaldrende barn som det snakker norsk sammen med?  
 På skolen  
 Nabolaget  
 Familie  
 Kirke/ moske o.l.  
 Fritidsaktiviteter
4. Hvor ofte deltar barnet på aktiviteter i fritiden sin der det snakkes norsk?  
 Sjelden  
 Ukentlig  
 Flere ganger i uken  
 Daglig
5. Hva leser barnet på fritiden sin på norsk? (besvares sammen med barnet)  
 Ukeblader  
 Tegneserier  
 Skjønnlitteratur  
 Faktabøker  
 Aviser

## Appendices

---

6. Hvor mye leser barnet for fornøylesens skyld på norsk? (besvares sammen med barnet)
- Barnet leser ikke for sin egen fornøylesens skyld.
  - 30 min eller mindre hver dag
  - Mellom 30 og 60 min hver dag
  - 1-2 timer hver dag
  - Mer enn 2 timer hver dag

7. Er du/ dere bekymret for barnets språk (på norsk)? .....

**8. Foreldrenes arbeidsstatus**

Mor		Far
<input type="checkbox"/>	Jobber heltid	<input type="checkbox"/>
<input type="checkbox"/>	Jobber deltid	<input type="checkbox"/>
<input type="checkbox"/>	Hjemmeværende	<input type="checkbox"/>
<input type="checkbox"/>	Student	<input type="checkbox"/>
<input type="checkbox"/>	Annet	<input type="checkbox"/>

**9. Kjenner du/ dere til at noen i barnets familie har språkvansker eller dysleksi?**

Mor	
Far	<input type="checkbox"/>
Onkler	<input type="checkbox"/>
Tanter	<input type="checkbox"/>
Besteforeldre	<input type="checkbox"/>
Søsken	<input type="checkbox"/>

Dette spørreskjemaet er besvart av:

Mor       Far      Andre (hvem)..... Takk for hjelpen