Anne Dahl

Young Second Language Learners

The acquisition of English in Norwegian first-grade classrooms

Thesis for the degree of Philosophiae Doctor

Trondheim, November 2014

Norwegian University of Science and Technology
Faculty of Humanities
Department of Language and Literature

NTNU – Trondheim
Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the degree of Philosophiae Doctor

Faculty of Humanitites
Department of Language and Literature

ISBN 978-82-326-0530-9 (printed ver.)
ISBN 978-82-326-0531-6 (electronic ver.)
ISSN 1503-8181

Doctoral theses at NTNU, 2014:309

Printed by NTNU-trykk

Abstract

This thesis deals with the acquisition of English as an L2 in young Norwegian children aged approximately six. The main focus is on learning in a traditional foreign-language setting, but also an immersion setting in an international school is investigated. The thesis contains four research papers, each addressing a separate topic related to the overall issue of early SLA.

The first paper studies vocabulary acquisition in two Norwegian classrooms where English is taught as a foreign language, which differ in the volume of target-language input to which the students are exposed. A significant and substantial effect is found of increased input. The second paper evaluates the use of a sentence repetition test as a measure of early second language competence, by administering it to four groups of children with different competence levels in English. This paper finds that sentence repetition is indeed a valid measure of early L2 competence, which assesses a range of competences such as phoneme perception, lexical knowledge, and morphology. The third paper studies the two foreign-language classes described for the first paper, and evaluates their performance on receptive vocabulary, sentence comprehension, and sentence repetition after one year of school, finding a clear advantage for the group with the highest English exposure on all measures, but on sentence comprehension in particular. The final paper focuses on the group of students immersed in English in an international school, and finds rapid development but great individual variation in English competence after one year of school.

The general discussion of the four papers focuses on what they can tell us about foreign-language learning as a form of SLA, about early L2 competence, and about the role of input in acquisition. In particular, it is argued that foreign language learning is indeed to be viewed as a form of SLA, and that the reported lack of benefits of a young age of onset in foreign-language settings may be a result of teaching styles being more suited to older learners, and to a lack of target-language input in particular. Results in the present thesis indicate that in early FL learning, input is in particular beneficial for processes such as phoneme perception, speech segmentation, and lexical access, facilitating sentence comprehension. Results also show that L2 lexical learning can be rapid in a school context, both for immersion and foreign language learners, at least as far as receptive vocabulary is concerned.

Acknowledgements

First and foremost I would like to thank my supervisor, Professor Mila Dimitrova Vulchanova. I am grateful to Mila for giving me the opportunity to start this project in the first place, for her help and support in designing and carrying out the study, and for insightful comments on all my work.

I am also grateful to a number of people for their help on various aspects of this project. I am especially indebted to Professor Joel Talcott at Aston University who has provided invaluable statistics advice, thorough comments on part of the thesis, and help and support in a number of situations. I am also grateful to Professor Rosa Alonso at the University of Vigo as well as two anonymous reviewers for invaluable comments on Paper 3, as well as to associate professor Stanka A. Fitneva, Queen's University, Canada, and Professor Teresa Cadierno at the University of Southern Denmark for reviewing Paper 1 for publication.

I would also like to thank the principals, teachers, parents and students at the schools included in this project, without whom I could not have completed this thesis, but who must unfortunately remain anonymous.

Thank you to Bjørn Grønnesby for producing the sentence comprehension test, to Rhonna Robbins-Sponaas, Emma Birkett, Angie Johnson, and Gillian Warner-Søderholm for lending their voices to various versions of this test, and to Jason Fox and Maria Boer Johannessen for serving as raters for the repetition test.

Also thank you to Terje Lohndal and Merete Anderssen for kindly agreeing to read and comment on various parts of my work, and to the Language and Cognition group at NTNU for many fruitful discussions, and for useful input and advice. I am also grateful to Anja Angelsen and Helene Hauge for proof reading parts of the thesis.

The Department of Language and Literature at NTNU is a great place to work, and I am very grateful to Arne Halvorsen for his help in giving me the chance to be here. Thank you to my fantastic colleagues, who are also great friends - both those of you who know the difference between R2-D2 and Obi-Wan Kenobi, and those of you who really, really don't. Your friendship, advice, encouragement and support are greatly appreciated.

The world's biggest thank you to my PhD buddies and excellent sputniks Sindre Aasheim Norås and Kjersti Faldet Listhaug for being someone with whom I can share ups and downs and triumphs and defeats. Extra special thanks to Kjersti for many, many things, including reading most of my work before anyone else, for always being ready to discuss and advise on any dilemma, and for holding my hand whenever needed. I don't think I could have done this without you.

Finally, thank you to all my friends and family for moral support and for distraction from my work. Special thanks to Hilde and to my wonderful parents for always being there for me.

Note on Paper 1

Paper 1 in this thesis, entitled "Naturalistic acquisition in an early language classroom," is joint work with Professor Mila Vulchanova, who has supervised this thesis, and was published in *Frontiers in Psychology. Special issue. Learning a non-native language in a naturalistic environment: Insights from behavioural and neuroimaging research,* Volume 5, 2014.

Anne Dahl had main responsibility for the study reported in the paper, including research design, data collection, analysis and interpretation, and drafting and revising the paper. Mila Vulchanova contributed substantially to the conception and design of the research, and to critical revision of the paper for important intellectual content. Both authors have final approvement of the version to be published and agree to be accountable for all aspects of the work.

The version of the paper submitted in the thesis is the final manuscript, corrected for final changes at the proofs stage, and differs from the published version in layout and formatting only.

List of abbreviations

ANOVA:	Analysis of variance
AoA:	Age of acquisition
BCa:	Bias corrected accelerated
CG:	Control group
CI:	Confidence intervals
CLIL:	Content and language integrated learning
EAL:	English as an additional language
EFL:	English as a foreign language
FL:	Foreign language
GSV:	Growth scale value
HG:	High-input group
ICC:	Intraclass correlation coefficient
IG:	Immersion group
K-Bit 2:	Kaufmann Brief Intelligence Test, 2nd version
L1:	First language
L2:	Second language
LG:	Low-input group
MANOVA:	Multivariate analysis of variance
OE:	Old English
ON:	Old Norse
PPVT-IV:	Peabody Picture Vocabulary Test, Fourth Edition
SD:	Standard deviation
SES:	Socioeconomic status
SLA:	Second language acquisition
YLL:	Young language learner
2L1:	Two first languages

Contents

Introductory chapter:

The context of the four papers – introduction, background, main findings, and methods

1 Introduction

The present thesis deals with second language acquisition (SLA) of English in three groups of Norwegian children in their first year of school. The research questions deal with the development of linguistic competence in child SLA, and specifically with the role of input for acquisition. Both immersion and child foreign-language (FL) settings are studied, and the assumption is that both contexts are essentially instances of processes through which children become bilingual. The articles in the thesis are thus relevant to SLA classrooms and to language education, but also to the study of multilingualism in general. Studying English in Norway is particularly interesting. It is a foreign language with no official role, and Norwegian children generally do not speak the language when they start school. Yet, English has a strong presence in the Norwegian society, is taught from grade 1, and most Norwegians achieve relatively high competence in the language. This situation is particularly suitable for a study based on the assumption that FL learning can lead to functional bilingualism, with the overall hypothesis being that all language acquisition depends first and foremost on input. A further aim of the study was to develop test materials suitable for such early language learners.

There are a total of 82 participants in the studies included in the project, including an English-speaking control group. All participants were approximately six years old, and attended their first year of school in Norway[1]. All participants were recruited through their schools, where teachers contacted the parents. All parents provided informed written consent for their child to participate; children whose parents did not respond to the request were not included. Testing took place at the children's school, during school hours or in the after-school program. Test times were decided in agreement with the teachers to minimize disruption of teaching. Even if parents had provided written consent, children were free to refuse to participate. All participants were assigned a code, under which their results were recorded, and the key connecting their names and codes was stored on a separate hard disk to ensure anonymity. When results are reported in this

[1] Some children in the CG were in their second year because of differences in starting ages between schools.

thesis, care is taken to protect this anonymity; for example, no school names are mentioned. The project was registered with and approved by the Norwegian Social Science Data Services (NSD).

The thesis is organized as follows: Section 2 provides a summary of the four papers which report on the studies in the project and which thus form its core. Section 3 provides the overall theoretical background for the project. Section 3.1 addresses ideas about bilingualism and second language acquisition, and discusses child SLA and early FL learning in particular. Section 3.2 provides a brief comparison of Norwegian and English, describes the role of English in Norway, discusses English teaching in Norwegian schools, and provides an overview of previous research on the acquisition of English in Norway. Section 3.3 sums up the assumptions of the project based on the theoretical background. Section 4 discusses the main findings from the four studies, and their implications for our knowledge about early second language (L2) competence in children and about child classroom SLA. Section 5 discusses in detail the methodology used, including considerations, main approaches, test administration, methods of analysis, and the limitations of the study. This section is followed by four appendices providing details of the tests used in the project, and one appendix providing original and translated versions of all parental information. The final section is the list of references. Then follow the four papers, in the order in which they are summarized in section 2, including their relevant appendices and lists of references.

2 Summary of the papers

2.1 Paper 1

The first paper is entitled "Naturalistic acquisition in an early language classroom," is joint work with Professor Mila Vulchanova, who has supervised this thesis, and was published in *Frontiers in Psychology. Special issue. Learning a non-native language in a naturalistic environment: Insights from behavioural and neuroimaging research* (Dahl & Vulchanova, 2014).

The paper investigates vocabulary development in two groups of children, all L1 speakers of Norwegian, in their first year of primary education in Norwegian state schools, mean age 6;1[2] at the start of the project, n=60. In one group, n= 29, labelled the native-language based group, English was taught as mandated by the curriculum (Utdanningsdirektoratet, 2006). In this group English classes took place for approximately 30 minutes per week. In addition, a few minutes were spent on English in morning meetings every day, where individual English words and expressions for e.g., the weather and days of the week were discussed. Total time spent on English per week was thus approximately 45 minutes. The medium of instruction was mainly Norwegian, which is the community first language (L1). Classroom activities were largely based on a workbook intended for Norwegian first- and second grade English classrooms, which contains activities such as coloring and finding pairs of written words and pictures. This book also contains recommendations to the teacher for formulaic instructions and communication in English. Furthermore, songs and rhymes played a central part in this group's English classes.

In the other group, n=31, referred to in this thesis as the bilingually-based group, time for formal English classes was about 30-40 minutes per week, which is well within the normal range based on the Norwegian curriculum for English in grade 1. However, in this group, activities focused on providing the students with input, in the form of the teacher speaking or reading aloud. Thus, target-language input was increased in that more communication during English classes took place in the L2 rather than the L1, but input was also more naturalistic and varied than in the native-language based group since it was used for real communication. In addition, time was spent on English during morning meetings. Unlike the L1-based communication accompanied by limited and formulaic English in the native-language based group, the teacher in

[2] All ages in this thesis are given as [year; month].

the bilingually-based group spoke as much English as possible during these meetings. However, there was no pressure on the children to produce more English than would be expected in a normal English classroom. Total time spent on English in this class was approximately 70 minutes per week.

The development of receptive vocabulary in both groups was studied in a test-retest design using the Peabody Picture Vocabulary Test, Fourth version (PPVT-IV) form B (Dunn & Dunn, 2007a), with a pre-test session as early as possible in the school year (approximately in September) and a post-test session as late as possible (approximately in May). Mean time between test sessions in both groups was eight months. Data were analyzed using non-parametric tests; Mann-Whitney U for differences between the two groups, and Wilcoxon rank sum test for the repeated measures analysis.

The main findings were that while the two groups' initial receptive vocabulary was not significantly different, their scores in the post-test were. Furthermore, the native language-based group did not show significant development from the beginning to the end of the year on the PPVT-IV. The bilingually-based group, on the other hand, did. Their development was compared to age equivalents as normed in the test manual (Dunn & Dunn, 2007b), and found to be greater than can be expected for language-matched (younger) native speaking peers, as the average increase in age equivalents in the bilingually-based group was 10 months in a chronological time span of only eight months. The conclusion is that early-start FL programs do not guarantee acquisition, but that substantial vocabulary development is possible in such classrooms through naturalistic exposure, and within a normal curriculum, i.e., without immersion.

2.2 Paper 2

Paper 2 is entitled "Sentence repetition as a measure of early L2 competence" and discusses the use of a sentence repetition test to measure early SLA. Sentence repetition has been quite widely used to measure language and language-related aspects such as aptitude and working memory. The research questions in this paper are whether such a test is a valid and reliable test of language competence in a second language for different competence levels, what the effect of sentence length is, and which specific language competences it can be used to test.

The test is evaluated based on data from four different groups of children; the two groups described in Paper 1, a group of children learning English in an immersion setting, and a control group of children who were native speakers of English. The two groups described for Paper 1,

i.e., the native language-based group and the bilingually-based group, both completed this test during the same test session as the post-test for vocabulary, i.e., at the end of their first year of school, when they had attended English classes for approximately eight months. At this time, the mean ages were 6;10 for the native-language based group and 6;9 for the bilingually-based group[3]. In this paper, the native-language based group is referred to as the low-input group (LG), while the bilingually-based group is referred to as the high-input group (HG). The third group discussed in this paper consisted of seven children, mean age 6;8 at the time of testing, who attended grade 1 (or equivalent[4]) in international schools in Norway, where the language of instruction was English. This group is referred to as the immersion group (IG). All children had no or very limited knowledge of English at the start of the school year, and were tested at the end of the year (approximately in May), i.e., after approximately eight months of immersion. The fourth and final group was a control group (CG) of 15 bilingual or monolingual native speakers of English, mean age 6;6, attending international schools in Norway. All of these children either had English as a home language, or had grown up in and only recently left an English-speaking country where they had attended child care in and been exposed to English from a very young age.

The test consisted of 17 sentences of various length and complexity, read by the researcher and repeated by the participant. The session was recorded, and two independent raters scored each recording based on the number of correctly repeated morphemes, including zero morphemes (see Appendix 3). Pronunciation was disregarded as long as the morpheme was comprehensible.

Performance on the test in each of the four participant groups, who are argued to represent four different competence levels in English, was compared. The first finding was that performance varied across groups as would be expected if sentence repetition is a valid test of language competence. Secondly, the results showed that even though all L2 groups scored significantly better on short than on long sentences, even with very short sentences, the test does not simply entail the mimicking of an acoustic image. Evidence for this is the finding that the

[3] The fact that the mean age was the same for both groups in the pre-test (6;1) and different in the post-test (6;10 and 6;9, respectively) does not imply a significantly longer span between tests in one group but is simply a result of each child's age being rounded to the nearest month.

[4] Some international schools have grade systems different from the standard Norwegian system. However, all children in this group were of the same age, i.e., turned six in the calendar year when they had started their current grade, and this was their first year of school.

participant groups with the lowest English competence scored significantly below the groups with higher competence also on the short sentences. This finding is explained by the fact that even in very short sentences, the participants must accurately perceive the words to successfully repeat them. Thus, thirdly, it is argued that sentence repetition may test phoneme perception, at least at very early stages. It is also argued that the test measures lexical access and grammatical processing, based on the finding that uninflected lexical words are significantly more successfully repeated than uninflected lexical words and functional words, at least in the bilingually-based group.

The overall conclusion of this paper is that sentence repetition is in fact a valid measure of language competence in SLA, and that such competence at very low levels includes both phoneme perception and production, as well as lexical knowledge and grammatical processing.

2.3 Paper 3

Paper 3 is entitled "Input and language competence in early-start foreign language classrooms" and is to appear in a volume in the series *Applications of Cognitive Linguistics* by Mouton, edited by Teresa Cadierno and Søren W. Eskildsen, on the basis of selected papers from the conference *Thinking, Doing, Learning: Usage Based Perspectives on Second Language Learning*, organized by the Second Language Research Center at the University of Southern Denmark on April 24 – 26, 2013, where this work was presented. This paper investigates group differences between the bilingually-based and the native-language based group in the post-test session on three measures of English competence, in order to investigate what benefits in particular can be expected from the increased L2 exposure in the bilingually-based group. The first measure is the PPVT-IV, but unlike in Paper 1, both forms A and B of the test are used. Each participant's reported score is the mean of both forms, presenting us with a more solid measure of receptive vocabulary. The second test is a sentence comprehension test of the same format as the PPVT-IV, i.e., where the participant was exposed to linguistic stimuli and for each stimulus was expected to select the matching picture from a set of four. While in the PPVT-IV the linguistic stimuli are vocabulary items, in the sentence comprehension test they were simple sentences with basic vocabulary. The final test used in this paper was the sentence repetition test discussed in Paper 2.

The research questions in this study were 1) whether the bilingually-based group would demonstrate an advantage over the native-language based group on all three measures, 2) which

measure would show the greatest difference, and 3) what this can tell us about the nature of early L2 competence.

Significant differences were found for all measures, i.e., receptive vocabulary, sentence comprehension, and sentence repetition. Sentence comprehension was the measure where the advantage of the bilingually-based group was the greatest. This is argued to reflect benefits of increased input in English in particular on processes such as "hearing words" (S. E. Carroll, 2004) and "good enough for now" processing (VanPatten, 2012).

To follow up on this finding, the question was asked of how big the impact of the increased exposure on sentence comprehension was, compared to other potentially influential factors. Multiple regression showed that the effect of group membership, i.e., volume of exposure, was more important than, for example, English comprehension at the start of the school year eight months earlier, and L1 Norwegian scores. It is argued that this is evidence that the effect of naturalistic exposure was substantial and thus that input is very important for comprehension in such young learners. It is finally argued that even though the effect of exposure on sentence repetition, expected to require more detailed processing, is smaller, the improved comprehension in the bilingually-based group may also be expected to benefit further acquisition.

2.4 Paper 4

Paper 4 is entitled "Individual differences and different outcomes in young children starting school in a new language," and focuses on the small immersion group of seven children described in Paper 2, who started international schools where English was the language of instruction, with little or no knowledge of the language. Their competence in English after their first year of school is studied, measured by the same three tests as those in used Paper 3 (i.e., receptive vocabulary, sentence comprehension, and sentence repetition). Results of the PPVT-IV forms A and B are reported individually. Form B was administered both in the pre-test session at the beginning of the school year (approximately in September) and in the post-test session (approximately in May), allowing us to draw conclusions about receptive vocabulary development. Results show rapid receptive vocabulary development for all seven children, and this development is faster for all participants than what is to be expected for younger, vocabulary-matched native speaking children as measured by age equivalents in the PPVT-IV manual (Dunn & Dunn, 2007b). However, most participants predictably still score well below the test's age-matched native-speaker reference group on the PPVT-IV, and below the means of the native-

speaker control group included in the project (see summary of Paper 2 above) on sentence comprehension and repetition. Furthermore, both receptive vocabulary development and results on all measures administered in the post-test session reveal that there is great intra-individual and inter-individual variation.

Results are discussed with reference to the debate about language competence in bilingual children and consequences for education. While linguists agree that bilingualism is natural and not problematic for children's development, there are concerns among parents, educators, and politicians about children who start school with low competence in the language of instruction. Problems are often attributed to bilingualism in itself, e.g., to claims that bilingual lexicons may on average be smaller than those of monolinguals, and that bilingual children may thus risk falling behind in class.

However, the main conclusion from this study is that bilingual children in school cannot be treated as one homogeneous group but must be discussed as individuals. Differences in age of acquisition[5] (AoA) and length of exposure in particular can be expected to cause differences in competence levels. On the other hand, there is no evidence that the number of languages a child knows has an impact on the acquisition of a new language in school. However, even for children with similar backgrounds, a range of factors may influence competence and development in a new language.

[5] The term *age of acquisition* in this thesis refers to the age at which acquisition begins, and is thus synonymous with other commonly used terms such as age of onset (AO) or age of onset of acquisition (AOA) (cf. Birdsong, 2006; Meisel, 2009).

3 Background

3.1 Forms of multilingualism

In the present thesis, the acquisition of a new language at any point in life is seen as a process of becoming bilingual or multilingual. However, the terms bilingualism and multilingualism are not unambiguous. The first term may refer to the use of two languages and the latter to the use of more than two, but in reality they are often used interchangeably. This is also the case in the present thesis. Although there are certainly differences between individuals who speak two languages and those who speak more than two (cf. Aronin & Singleton, 2012 ch. 2; Butler, 2012, pp. 110-111; Wei, 2000, p. 6), such differences will not be addressed here.

Furthermore, the terms multilingualism and bilingualism are ambiguous in terms of age and context of acquisition, and of competence in each language (Cook & Bassetti, 2011; Grosjean, 2010; Meisel, 2004; Wei, 2000, pp. 3-4). This is related to the traditionally assumed division between two types of language competence which a speaker may possess, namely L1, or native, competence, and L2 competence. The acquisition of the former is assumed to be a generally successful process for all normally developing children, while the acquisition of the latter is sometimes successful, often unsuccessful, and anything but uniform (e.g., Birdsong 1999; Bley-Vroman, 2009). The assumed fundamental difference between the two processes is, for instance, reflected in the fact that L1 and L2 acquisition are different research fields, largely with different journals and different conferences (cf. Ortega, 2009).

3.1.1 Multiple first languages

Some definitions of bilingualism emphasize native language competence in both languages (e.g., Bloomfield, 1933; Thiery, 1978). Acquiring more than one language during early childhood is absolutely normal (Cenoz & Genesee, 2001; Romaine, 1999; Tucker, 1998). However, in most of the western world, monolingualism has traditionally been the norm, and research on child bilingualism was scarce for most of the 20th century, although some studies existed, notably that of Ronjat (1913) and Leopold (1939, 1947, 1949a, 1949b). Many early studies focused on observed phenomena of cross-linguistic influence, often discussed in terms of language mixing, and concluded that such mixing reflected confusion on the part of the child. Volterra and Taeschner (1978), for example, suggested that the bilingual child goes through a

stage of acquisition where both languages are represented by one unitary linguistic system. Worries about linguistic confusion in bilinguals also led to recommendations such as the *one parent-one language* or *one environment-one language* strategy (see Romaine, 1999). However, such recommendations were often influenced by the situation for western families, typically of relatively high socioeconomic status, where bilingualism may be a choice. This is not the typical bilingual situation, and the reality for many children is that they grow up to be multilingual because a they are naturally exposed to a number of languages in their family and community (Nicolay & Poncelet, 2013). Furthermore, concerns that cross-linguistic influence in bilingual children is a sign of confusion have been debunked. We know, for example, that code-switching is entirely normal also in adult multilinguals, and that it is not a sign of low language proficiency (Poplack, 1980). Finally, evidence has made it clear that multilingual infants are able to discriminate between languages in early speech perception and that children growing up with more than one L1 differentiate their languages from the start (e.g., Bosch & Sebastian-Galles, 1997; Bosch & Sebastián-Gallés, 2001; Meisel, 2001, 2004; Serratrice, 2012; J. Werker, 2012; J. F. Werker & Byers-Heinlein, 2008). This is important not only because it means that parents should not worry about "language confusion" in their bilingual child, but also because this apparent ease of language separation, i.e., of building more than one linguistic system from an early age, is evidence that the human mind is indeed able to handle multilingualism (Meisel, 2001).

Other early studies of bilingualism found that growing up with two languages is detrimental to intelligence (e.g., Saer, 1923) or to linguistic competence (see e.g., Romaine, 1999 for an overview). However, beginning with Peal and Lambert (1962), study after study has concluded that there are no cognitive disadvantages to bilingualism. It is also clear that bilingualism does not cause substantial delays in children's language development, nor lead to generally impoverished linguistic competence (Bosch & Sebastián-Gallés, 2001; Meisel, 2004). During the past three or four decades, researchers have come to the realization that growing up bilingually with two L1s is quite common (Cenoz & Genesee, 2001), and during this period, studies have convincingly shown not only that bilingualism is not detrimental, but also that there are no fundamental differences between monolingual and bilingual L1 acquisition (Meisel, 2001, 2004; Romaine, 1999). There may even be cognitive advantages to being bilingual (Bialystok, 2009; Bialystok, Craik, & Freedman, 2007).

On the other hand, bilingual speakers are not "two monolinguals in one"; their linguistic competence in each of their languages is not necessarily directly comparable to that of a monolingual speaker of each language (Grosjean, 1989). It is possible that bilingual L1 acquisition may tend on average to be slightly later than monolingual acquisition, but not outside the normal range (cf. Meisel, 2004). On the other hand, perfectly balanced bilingualism is hardly the norm (e.g., Haugen, 1969). In order to fairly study the language competence of bilinguals, it is important to properly investigate language competence in each language, to establish whether or not the language in question is dominant, and to check factors such as AoA and length of exposure (cf. De Houwer, Bornstein, & Putnick, 2013; Grosjean, 2008; D. K. Oller & Pearson, 2002; D. K. Oller, Pearson, & Cobo-Lewis, 2007; Treffers-Daller, 2011).

3.1.2 Second language acquisition and use

Many definitions of bilingualism focus on the regular use of more than one language, rather than on AoA or native-like competence, and thus include L2 use (e.g., Grosjean, 2010; Mackey, 1962; Weinreich, 1953/1968). Until the second half of the 20th century, the learning of languages additional to your native language(s) was mainly regarded as a pedagogical issue (cf. Gass, Fleck, Leder, & Svetics, 1998; Mitchell & Myles, 1998 ch. 2). SLA developed as a research discipline in its own right in the 1960s and 1970s, with seminal work such as Corder (1967) and Selinker (1972) establishing that SLA, like L1 acquisition, entails the development of a linguistic system, which can be studied in its own right. The L2 grammar, or *interlanguage*, is different from a native language grammar in several respects, notably in being less stable. As the prefix *inter-* indicates, the idea was that the interlanguage is a developing system somewhere in between the learner's initial state and the ideal final state, the target-like native grammar. However, Selinker argued, more often than not, L2 learners in fact do not reach this native-like grammar; they tend to *fossilize* in forms short of the target grammar in one or more language areas. Also views on the role of errors changed. While traditionally, not least influenced by the habit-formation ideas of behaviorism and audiolingualism (e.g., Skinner, 1957), errors had been seen as problems in acquisition and elements to be avoided, both Corder and Selinker emphasized how errors are necessary for learners' hypothesis testing about language; they are in fact what drives acquisition forward.

Another influential theoretical advance for SLA was the Monitor Model (e.g., Krashen, 1981b; Krashen, 1985), particularly two of its hypotheses, the Input Hypothesis and the

Acquisition - Learning Hypothesis. The former, in broad terms, claims that input at exactly the right level (termed the i+1 level, or one step above the current interlanguage) is the necessary and sufficient condition for all language acquisition, including SLA. The latter hypothesis draws a sharp distinction between acquisition, as the natural development of the ability to use real language for communication, on the one hand, and learning, i.e., the explicit learning of rules about the language, on the other. In Krashen's view, such learned knowledge can never turn into acquired, subconscious and automatic, knowledge, and as such has very limited value in L2 acquisition.

These ideas, in conjunction with the observation that L1 acquisition seems to be both more effortless and more successful than SLA, as well as findings that SLA in naturalistic contexts is generally more successful than in instructed settings, led to a strong focus on communicative approaches in the language classroom in the 1980s. This was a not entirely unwelcome focus in an era when behaviorist ideas, or even older traditions of grammar and translation, still strongly influenced language classrooms in most countries (Lightbown, 2000). However, such attempts to mimic the L1 acquisition process in SLA classrooms did not lead to comparable acquisition levels (see, e.g., Lightbown, 2000; Lightbown & Spada, 1990). This led to the hypothesis that there is a fundamental difference between L1 and L2 acquisition (Bley-Vroman, 1988, 1989, 1990), and that children are able to acquire language implicitly, while adults depend on conscious problem-solving mechanisms. Thus, the distinction between what we call a first and a second language is not based on the number of languages which an individual knows[6] but on context of acquisition, language competence, and above all, on age of acquisition.

The role of age in SLA is one of the most studied and debated issues in the field. Among lay people, it is a common assumption that children acquire languages a lot more easily than do adults. Evidence suggests that the picture is more complicated, and that older learners may in fact start out faster than younger ones, but there is a relatively general agreement among linguists that the earlier one starts to acquire a language before adulthood, the better the end result will normally be (cf. DeKeyser, 2000; DeKeyser & Larson-Hall, 2005; Hyltenstam, 1992; Hyltenstam & Abrahamsson, 2003; Johnson & Newport, 1989; Krashen, Long, & Scarcella, 1979; Muñoz,

[6] This is not to say that the number of languages which an individual knows is not relevant for some purposes. We know that how many languages you know before learning a new one may make a difference and that previously learned L2s may influence new languages, and the term L3 may thus be useful in some contexts (e.g., Bohnacker, 2006; Lindqvist, 2010; S. Murphy, 2003; Tremblay, 2006). This, however, is not the focus of the present project.

2008a; Muñoz & Singleton, 2011; Singleton & Ryan, 2004 ch. 7; Snow & Hoefnagel-Höhle, 1977).

Exactly how children and adults are different in their language acquisition, and at what age this difference appears, is not clear. It has been suggested that there exists a critical period for language acquisition, within which human beings are naturally equipped to acquire language, and after which they are not. According to Lenneberg's (1967) Critical Period Hypothesis, native-like language competence cannot be achieved from mere exposure after puberty. Other researchers have taken the idea of the critical period to mean that language cannot be acquired to a native-like level at all after a certain age, regardless of circumstances. Different ages for the end of the critical period have been suggested, and it has also been argued that the time in a person's life when language can be acquired to native-like levels does not have the sharp on- and offset required by a critical period, so that the term *sensitive* period is more appropriate. Finally, a suggestion is that there may be not one but multiple critical or sensitive periods constraining different aspects of language, and some research has even suggested that the capacity to acquire language to a completely native-like level starts declining already from birth, so that exposure to the language from the very beginning is necessary for completely native competence (De Houwer et al., 2013; Johnson & Newport, 1989; Locke, 1997; Long, 1990; Meisel, 2009; Newport, 1988, 1990; Ruben, 1997; Seliger, 1978). Hyltenstam and Abrahamsson (2003) propose that this is in fact the case, and that examples of native-like language acquirers with AoAs later than soon after birth are results of underanalyzed data, i.e., that closer scrutiny might have revealed non-nativelike characteristics. This, however, does not mean that age does not matter for acquisition; although they see the decline in language acquisition abilities as linear rather than abrupt, studies show that an earlier AoA generally matters at least for whether interlocutors *perceive* speakers as native-like (Abrahamsson & Hyltenstam, 2006).

Explanations for age effects in language acquisition include theories that lateralization or loss of plasticity in the brain causes the decline in language acquisition abilities, e.g., because of myelination of the neurons in the brain (cf. Lenneberg, 1967; Penfield & Roberts, 1959; Pulvermuller & Schumann, 1994), or because of a decrease in synaptic density (cf. Chechik, Meilijson, & Ruppin, 1997; de Bot, 2006; Roe, Pallas, Hahm, & Sur, 1990). Other explanations argue that children are inherently more motivated or that they receive more and more appropriate input than adults (e.g., Bialystok & Hakuta, 1999; Piske & Young-Scholten, 2009). Cognitive

explanations typically argue that the cognitive problem-solving mechanisms that adults necessarily use in the learning process somehow inhibit acquisition (e.g., Bley-Vroman, 1989; Felix, 1985; Newport, 1990). It has also been suggested that affective factors may hinder acquisition in adolescents and adults (Krashen, Long, & Scarcella, 1982; Rosansky, 1975). N.C. Ellis (2006b) argues that the L1-L2 difference is not in age per se, but that L2 acquisition is hindered by factors shaped by the L1. As an example, he mentions MacWhinney's (2001) report of a highly advanced L2 speaker of English who still depended on cue strength for syntactic role assignment from his native German, i.e., favoring agreement and animacy, rather than from English, where word order is a stronger cue, when they were in conflict. Ellis furthermore argues that low-salience cues which are regularly experienced together with a more salient cue predicting the same outcome may be overshadowed by the stronger cue to the extent that it is not perceived, and thus eventually blocked from being seen as a cue. He argues that this may be the case for ,for example, for tense and number inflection in L2 English. Factors such as cue competition, overshadowing, and blocking may thus prevent L2 input from becoming intake (N. C. Ellis, 2006b). Since the L1 system is "entrenched" with use, such an account also generates age effects.

The fact that it is difficult to pinpoint an age at which it is no longer possible to acquire a new language to native-like levels (Birdsong, 2006; Hyltenstam & Abrahamsson, 2003) challenges the idea of a fundamental difference between first and second language acquisition, since it means that AoA simply is not an absolute yardstick for competence. Furthermore, as already mentioned, we know that even individuals who have two L1s need not be equally proficient in both at all times, and that even bilinguals whose languages are both native differ from monolingual speakers (cf. Grosjean, 1989; Haugen, 1969; Treffers-Daller, 2011). This means that when we are comparing the competence of an L2 user to that of a monolingual L1 speaker, we may be exaggerating the effect of the L2's non-native status - it might be the person's multilingualism, not her late start with the L2, which causes the difference (cf. Birdsong & Gertken, 2013; Ortega, 2014). Finally, cognitive benefits have been found not only in early onset (2L1) bilinguals, but also in late onset (L2 user) bilinguals (cf. Bialystok & Barac, 2012; Bialystok et al., 2007; Bialystok, Peets, & Moreno, 2014; Nicolay & Poncelet, 2013). Structural changes in the brain due to bilingualism may be correlated with AoA and proficiency, but such changes have still been found also in late bilinguals (Mechelli et al., 2004). This suggests that

there need not be a fundamental qualitative difference between various forms of bilingualism, i.e., whether both languages are best described as L1s or whether one language is an L2.

Finally, while a massive literature exists on how SLA differs from L1 acquisition, the question has arisen of whether it is really fair to judge L2 learners on the basis of the competence of (monolingual) native speakers (Bley-Vroman, 1983; Cook, 1991, 1992, 1999; Klein, 1998; Ortega, 2014). After all, as mentioned above, findings from research on bilingual children shows that competence in each language is not identical to that of monolinguals, even when both languages are acquired from birth. Over the past few decades, focus has thus tended to shift from the deficiencies of the L2 user as compared to the native speaker, to valuing the multilingual competence of L2 users in its own right (e.g., Cook, 1991; Cook, 1992, 1999; Kay, 2014; Ortega, 2014). From such a perspective, native-like competence is not an issue.

3.1.2.1 Child SLA

In spite of the general agreement among linguists that bilingualism is a positive phenomenon which is not detrimental to a child's development, there are still concerns about aspects of children's multilingual competence in society at large. The ambiguous use of the terms *bilingualism* and *multilingualism*, and the fact that there is no absolute boundary between the acquisition of an L1 and child SLA, may be one reason for this situation. This is particularly the case when it comes to school children with home languages other than the language of instruction, e.g., minority language children in majority language schools. The concern is that they will not be able to follow normal teaching because of low competence in the language of instruction, and typically the issue is vocabulary size, which has been argued to possibly be smaller for bilinguals than for monolinguals, and which may especially be important for literacy (e.g., Bialystok, 2009; Bialystok, Luk, Peets, & Yang, 2010; Cummins, 1979; Lee, 2011; Lervåg & Aukrust, 2010; Melby-Lervåg & Lervåg, 2013).

However, a problem with many such studies is exactly the fact that the terminology is ambiguous and not used consistently in all reports. An example is a large-scale analysis of receptive vocabulary in bilingual children by Bialystok et al. (2010). This analysis found a lower average score for receptive vocabulary in bilingual than monolingual children. The criterion for inclusion in the bilingual group in this study was that the child's home language was not that of the school (English), that she was fluent in both languages according to parental reports, and that she used both languages on a daily basis. It is explicitly stated that children learning English as an

L2 were not included, but it is not entirely clear what this entails, i.e., what the definition of an L2 learner is. No ages of first exposure are reported, nor clear cut-off criteria for when a child is considered bilingual as opposed to being a learner. The explicit criterion of a different home language, however, at the very least entails that these children had not been brought up bilingually from birth, and thus that their AoAs were likely to be higher on average than those of the monolinguals to whom they were compared. In turn, this would mean that their exposure to English had been shorter than that of the age-matched monolinguals who had been acquiring the language since birth (cf. De Houwer et al., 2013). Thus, when vocabulary is found to be smaller for the bilinguals in Bialystok et al.'s study, it might be tempting to suggest that these children are in fact still learners, in the process of learning English vocabulary which their monolingual classmates already know, having been exposed to the language from birth. Conversely, a meta-study by Melby-Lervåg and Lervåg (2013) found problems in reading comprehension in L2 learners of the language of instruction compared to L1 speakers. This is not necessarily an interesting finding unless AoA is discussed since generally lower proficiency, including lower reading comprehension, is to be expected for children whose exposure to the target language has been short. This problem of different criteria for bilingualism in different studies is also pointed out by Bialystok, McBride-Chang, and Luk (2005); see also Vulchanova, Vulchanov, Sarzhanova, and Eshuis (2012) for a similar argument. In a sense, one might argue that this discussion of bilingual children in schools still reflects earlier prejudices about bilingualism since the relative strength and status (L1 or L2) of the children's languages are not properly assessed. Thus, we are looking at a very old phenomenon, pointed out ,for example, by Garcia (1974) in the US.

In the present thesis, child SLA in an immersion setting is studied, in a group of children who start international schools in Norway where English is the language of instruction, with little or no knowledge of English. This context is different from that of minority language children starting school in a majority language, in that these children are in fact speakers of the community L1 (Norwegian), in which education would have been available. Thus, their situation may be said to resemble that of other immersion contexts chosen by the parents for speakers of majority languages, e.g., French immersion for Anglophone children in Canada or foreign-language immersion projects in Europe (cf. Cummins, 1998; Genesee, 2004; Jalkanen, 2009; Lightbown, 2012; Wesche, 2002; Wode, 1996, 1997, 1999). However, there are some

fundamental differences between the international school setting and such immersion classrooms. In the case of the children in the present project, most of their classmates already knew English, and did not necessarily speak Norwegian. Thus, the class did not have a common language to fall back on if communication failed. Similarly, these children could not expect to be able to use their L1 extensively during breaks, and their social relationships with classmates would predominantly take place in the new language. Importantly, they would have access to high-quality input not only from teachers, but also from their peers. Finally, unlike immersion classrooms such as those described above, which are often "bilingual" to varying degrees in that teaching in the immersion language ranges from 50 to 100%, English was the sole language of instruction for this group of children, with the exception of mother-tongue classes. Thus, their situation can be assumed to be more similar to that of minority language children than to that of other immersion settings. Their situation is not unique, and international schools exist all over the world. However, this group of students do represent a specific type of child language learners, and the aim of this study was to shed further light on such children's language development, in order to contribute to our knowledge of the complex phenomenon of child bilingualism.

3.1.2.2 Foreign language learning

While definitions of bilinguals normally focus on individuals who use two (or more) languages in their everyday life (cf. Grosjean, 2010), large numbers of the world's population acquire languages in addition to their L1s through foreign language (FL) learning. This form of language learning takes place in a classroom situation in a country where the target language is not in general use; learning is often aimed at later use in other countries, and access to target language input is usually limited (cf. Cook, 2008; Wei, 2000). The distinction between FL and other L2 learning is not universally recognized, and often not explicitly addressed in SLA studies (cf. Crystal, 2003; Ortega, 2009). However, for certain purposes, especially for questions about input, it may be useful to treat FL learning as a specific subtype of SLA.

In FL acquisition, the classroom constitutes the learning environment. Since classrooms can be manipulated to a much greater extent than what is the case in naturalistic acquisition, and since for an FL there is little or no naturalistic acquisition outside the classroom, questions about the best way to acquire language become central. For example, the role of explicit instruction is particularly pertinent to such classrooms. The central question is whether input is in fact the necessary and sufficient condition for acquisition also in language classrooms (Krashen, 1981b,

1985), or whether explicit instruction may also be beneficial. An early study to find no benefit of instruction was that of Felix (1981), who found no effect of formal instruction on developmental sequences. However, Pienemann (1984, 1989) found that instruction may speed up the progression. By now, a large literature exists dealing with questions of explicit instruction, generally concluding that it does have a role in classroom SLA (e.g., De Graaff, 1997; R. Ellis, 2006; Hulstijn & De Graaff, 1994; Laufer, 2005; Lightbown, 1983, 2000; Lightbown & Spada, 1990; Long, 1991; Nassaji & Fotos, 2011; Norris & Ortega, 2000; Odlin, 1994; Robinson, 1997; Schwartz, 1993).

However, the effect of instruction depends on the understanding of what the term refers to (Doughty, 2003). For example, it is now common to distinguish between the isolated explanation of linguistic elements as a set of accumulated knowledge, often referred to as *focus on formS*, on the one hand, and *focus on form* as the practice of drawing students' attention to linguistic forms in the course of real language use on the other (Doughty & Williams, 1998; Long, 1991). It is likely that explicit instruction may play a role in helping students notice linguistic forms, and that such consciousness-raising may speed up acquisition (cf. N. C. Ellis, 1995). Another suggested role for explicit instruction is processing instruction, i.e., the explicit instruction of input processing strategies (VanPatten, 1996, 2012; VanPatten & Cadierno, 1993a, 1993b). Clearly, however, any type of explicit learning has to operate on target-language input and not in isolation.

Another question in the FL classroom is the role of students' production, i.e., of the extent to which "output" has an effect on acquisition. For Krashen (e.g., 1981b), output has a limited role in adult SLA in that it enables the "Monitor" to compare output with explicit knowledge, which may lead to alternations in the output from the acquired system. Others have argued for a more central role for output, in particular of "pushed output", which may lead to acquisition when learners encounter gaps in their linguistic competence; like explicit instruction, then, production may facilitate "noticing" (e.g., Swain, 1985, 1995; Swain & Lapkin, 1995).

Given what we know about age and language acquisition, lowering starting ages for FL in school seems like a promising initiative to achieve increased competence with little extra effort. It is becoming increasingly common to begin learning languages in school at a young age; European children are now typically between 6 and 9 when they start learning an FL, and English is by far most commonly taught (Eurydice & Eurostat, 2012). However, a number of studies

indicate that an early start in an FL is not necessarily beneficial for the students' eventual competence. Early studies reporting no benefit of an early start were Oller and Nagato (1974), finding no long-term advantage for early starters on the acquisition of English in Japan, Burstall's (1975) review of the effect of French instruction in British primary schools, and the EPÅL project in Sweden (Holmstrand, 1982) which indicated that in grade 6, there was no difference in various skills in English between a group of students who had started learning the language in grade 1 and a group who had started in grade 4. Also a number of recent studies indicate that later starters perform as well as or even better than younger starters in instructed settings (Álvarez, 2006; Cenoz, 2002, 2003; García Lecumberri & Gallardo, 2003; García Mayo, 2003; Lasagabaster & Doiz, 2003; Miralpeix, 2006; Mora, 2006; Muñoz, 2003, 2006; Navés, Torras, & Celaya, 2003; Torras, Navés, Celaya, & Pérez-Vidal, 2006).

Judging by the results of such studies, one might argue that foreign languages could be kept out of classrooms during the earliest years, since the time spent in earlier grades seems to make no difference; later starters seem to catch up anyway. In fact, studies of age effects in SLA often distinguish between age of first exposure, e.g, as an FL, and AoA as the age of immersion, often in an immigrant context, assuming that FL learning does not equal *significant* exposure (cf. Birdsong, 2006; Muñoz, 2008b). Lightbown (2000, p. 449) argues that in FL, "[t]he reality is that perfect mastery of a target language is rarely attained, even when learners begin at an early age," and thus that the relevance of a critical period in such a context is questionable.

However, the question is exactly what an early start in an FL means. At least two effects are likely with a young AoA in an L2, both of which are relevant to the classroom setting: Firstly, it is likely that a young AoA can facilitate high ultimate attainment in the target language even when native-like competence is not realistic. Secondly, however, a young AoA probably means *different*, not just *better* language acquisition abilities. In particular, children may be superior at implicit learning but benefit less from explicit instruction, which means that target language input may be especially important for young language learners (cf. Muñoz, 2006, 2008a, 2009). As V. A. Murphy (2010) points out, contexts of child SLA vary, and she argues that high-input contexts are the successful ones. For example, in one of the few studies in an English as a foreign language (EFL) context where an advantage was found for younger learners (Larson-Hall, 2008), the young starters were exposed to significantly more input than is the case in the above reports. Djigunović and Vilke (2000) also found an advantage for early starters learning English in

Croatia, but this advantage depended on quality of exposure. Also Abello-Contesse et al. (2006:17) argue that lowering the starting age for FLs is pointless unless accompanied by extensions of L2 teaching time and in-service training programs focusing both on language/communication and on pedagogy. Burstall (1975) and Stern (1983), both cited in Lightbown (2000), suggest that in instructed FL acquisition, AoA is less important than intensity and continuation of exposure. As already mentioned, younger learners probably excel at implicit as opposed to explicit learning, meaning that input plays a particularly important role with a low AoA. For a low FL starting age to have an effect, exposure to the target language must thus be sustained and substantial (cf. Lightbown, 2000; Ruiz-González, 2006).

Although most European countries have requirements in place to ensure that teachers in early-start FL programs are qualified language teachers, very limited target language use in the classroom is a common problem in primary education in most European countries (Eurydice & Eurostat, 2012). It is possible that this is partly a result of the tradition of FL teaching being developed for older learners. In Norway, which is the country in focus in the present project, there is also no requirement that teachers of English in grade 1 are qualified language teachers. This is a fact to which we will return.

With the spread of early-start FL teaching, especially for English, research on so-called young language learners (YLLs), normally defined as learners at primary school level, has become more common and includes a range of questions beyond just effects of early starting ages (see e.g., Nikolov, 2009b; Nikolov & Djigunović, 2006; 2011 for overviews). There exist a number of reports describing early foreign language programs in various countries, published by institutions involved in or with an interest in language teaching, as well as edited volumes of research on foreign language learning at the primary level (e.g., Edelenbos, Johnstone, & Kubanek, 2006; Enever, Moon, & Raman, 2009; Eurydice & Eurostat, 2012; Gonzìez Davies & Taronna, 2013; Nikolov, 2009a, 2009c; Nikolov, Djigunović, Mattheoudakis, Lundberg, & Flanagan, 2007; Rhodes & Pufahl, 2008). However, research on the very earliest grades is still relatively scarce, and for the very youngest children tends to focus on aspects such as motivation, attitudes, and teacher perspectives, rather than on language acquisition per se. For example, in Nikolov (2009c), the vast majority of chapters deal with the upper primary level, and among the few chapters including students age seven and younger, we find two studies focusing on attitudes and motivation (Djigunović, 2009; Enever, 2009). This partly reflects the fact that goals of such

early language classrooms are often to foster favorable attitudes towards FL learning as much as acquisition in itself, but also problems in testing very young learners may be a reason, not least since oral skills are in focus in these very young learners (cf. Drew & Hasselgreen, 2008; Nikolov & Djigunović, 2006, 2011). Most work on language testing focuses on learners beyond the very first years of school (cf. Hasselgreen, 2005; McKay, 2006).

Among the studies that do exist of language acquisition in young FL learners, we find Alexiou's (2009) study of the relationship between cognitive skills and vocabulary learning in Greek learners, and Orosz' (2009) investigations of which English word types are most easily learned through explicit instruction for Polish seven-year-olds. In the Netherlands, Goorhuis-Brouwer and de Bot (2010) and Unsworth, Persson, Prins, and De Bot (2014) both found some, but limited, English learning in groups of children from age four. The latter study found an effect of volume of target language exposure, but an even stronger effect of the teacher's language proficiency. This effect, interestingly, depended on teacher proficiency but not on nativeness. These latter studies are relevant since the situation in the Netherlands is in some respects similar to that of Norway, in that general English proficiency in the population is high, that English is commonly encountered in the media, and that the majority L1 (Dutch) and English are relatively closely related. However, while Dutch children start school at a younger age than Norwegian children, English is not yet compulsory in the earliest grades. Thus, findings from schools which choose to offer early-start English programs may not be directly transferrable to Norway, where English is compulsory from grade 1 (cf. section 3.2.3).

In short, while research on primary-level foreign language learners has increased dramatically over the past ten years, such studies still most often focus on learners nine years or older. Furthermore, learning contexts and classroom cultures vary greatly from one country to another, as does language distance between the L1 and L2. Thus, not all findings from those studies that do exist can necessarily be applied to another context, although certain factors such as target language exposure and teacher competence may be universal success factors. Very little research relevant for very young English learners in Norway exists; this research will be reviewed with other research on L2 acquisition of English by Norwegian speakers in section 3.2.4.

3.1.2.3 Individual differences in language acquisition

Success in SLA is subject to individual differences, for example, in cognitive factors. Such differences have often been described in terms of language aptitude, which can be defined as "a largely innate, relatively fixed talent for learning languages" (Abrahamsson & Hyltenstam, 2008, p. 485). Exactly what this talent entails is not entirely clear, but there is agreement that it is not *one* general ability, but rather that it consists of more or less independent, specialized abilities. Suggested components of aptitude include e.g., phonetic coding, grammatical sensitivity, analytic, associative and inductive learning, and memory (J. B. Carroll, 1962, 1993; Skehan, 1986, 1989, 1998, 2002, 2012).

Aptitude was traditionally discussed in classroom SLA or foreign language contexts with older learners, and the extent to which the concept is also relevant for young language acquirers such as those in the present project is not clear. Krashen (1981a) argued that aptitude is only relevant to explicit learning and not to implicit acquisition, while recent studies (De Graaff, 1997; DeKeyser, 2000; Harley & Hart, 2002; Robinson, 1997) indicate that aptitude plays a role also for naturalistic acquisition in adults. Abrahamsson and Hyltenstam (2008) found that aptitude effects are most prominent for older L2 acquirers, but that there are small but significant such effects also for younger L2 acquirers.

However, measuring language aptitude in young children is difficult since their L1 abilities are still developing, and since most aptitude tests require literacy (cf. Milton & Alexiou, 2006; Sebastian-Galles & Diaz, 2012). As a result, aptitude in very young learners has rarely been discussed in the literature (cf. Nikolov & Djiguović, 2006). An exception is Alexiou (2009), who found a relationship between cognitive skills and L2 English vocabulary learning in young Greek learners aged 5-9. Even though aptitude is generally assumed to be relatively stable throughout an individual's lifetime (cf. Abrahamsson & Hyltenstam, 2008; J. B. Carroll, 1962), it is also possible that aptitude effects may be of a slightly different nature at very young ages than later in life. For example, while working memory has been found to influence many aspects of SLA (cf. Juffs & Harrington, 2011), memory may be more important with younger learners while analytic skills may become more important with age (Harley & Hart, 1997; Milton & Alexiou, 2006).

Although focus on factors influencing language learning has tended to be on L2 acquisition, research has also found individual differences in L1. For example, correlations have

been found between L1 verbal skills and genetic factors and with other cognitive abilities, in particular short-term memory, in children of various ages (e.g., Adams & Gathercole, 1995, 2000; Baddeley, 2003; Baddeley, Gathercole, & Papagno, 1998; Colledge et al., 2002; Foyn, Vulchanova, Nilsen, & Sigmundsson, under revision; S. E. Gathercole, Hitch, Service, & Martin, 1997; S. E. Gathercole, Tiffany, Briscoe, Thorn, & team, 2005; Hayiou-Thomas, Dale, & Plomin, 2012). Such L1 individual differences in turn seem to correlate with L2 acquisition (cf. Dale, Harlaar, Haworth, & Plomin, 2010; Sparks, Patton, Ganschow, & Humbach, 2009; Sparks, Patton, Ganschow, Humbach, & Javorsky, 2006).

3.2 Norwegian and English

3.2.1 A crosslinguistic comparison of English and Norwegian

Since the topic of the present thesis is the acquisition of English by native speakers of Norwegian, a comparison of the two languages is called for. However, since no specific grammatical phenomena are addressed, but rather receptive vocabulary, overall comprehension, and processing, the linguistic comparison will only provide a brief overview of general lexical and structural similarities and differences.

English and Norwegian are related languages, both belonging to the Germanic branch of the Indo-European language family. While English is a West Germanic language, Norwegian is North Germanic and thus closely related to the other Scandinavian languages. In particular, it is very similar to Danish and Swedish, which are both mutually intelligible with Norwegian.

In the lexicon, English is, for historical reasons, different from the other Germanic languages in a high number of lexical items from Romance languages, especially French and Latin. The lexicon of Norwegian, on the other hand, is largely Germanic in origin. The two languages still share a high number of cognates, which can generally be attributed to four different sources. First, basic lexical items in English are often of Anglo-Saxon origin and thus reflect the common ancestry with the other Germanic languages. Examples of words which share the same origin in English and Norwegian in this way are abundant, and are typically frequent words, such as *man* (Norwegian *mann*), *house* (Norwegian *hus*), and *apple* (Norwegian *eple*) (Hoad, 1996). A second source of lexical similarity is the influence of Old Norse (ON) on Old English (OE) during the 9th and 10th centuries, i.e., the Viking ages. Because of similarities between OE and ON, it is not always possible to ascertain whether a Modern English word is an

adaption of the OE or the ON word, but examples of words believed to be of ON origin are *cake* (Norwegian *kake*), and *guest* (Norwegian *gjest*) (Hoad, 1996). Third, common loan words from other languages is another source of similarities, in a wide range of topics such as sciences (e.g., *psychology/psykologi*), technology (e.g., *mechanic/mekaniker*), and words for concepts of non-European origin (e.g., *kangaroo/kenguru*) (Berulfsen & Gundersen, 2000; Hoad, 1996). Fourth, Norwegian contains a high number of loan words from English, often related to new concepts which have been new to Norway in the 20th and 21st centuries, such as *shorts*, *muffin*, and *layout* (cf. Berulfsen & Gundersen, 2000; Johansson & Graedler, 2002). Thus, for a Norwegian, some English vocabulary will be easily understood because it is similar to Norwegian, while other words are not transparent.

A further complication for spoken comprehension between Norwegian and English is phonology. No attempt will be made here to give an exhaustive account of the phonological systems of either language, especially since this system differs greatly between varieties of the two languages. In very broad terms, however, the consonant systems of the two languages are very similar, although each language contains consonants not found in the other language (cf. Helland, 2008; Helland & Kaasa, 2005; S. Lie, 1993). For example, Norwegian does not use the dental fricatives /θ/ and /ð/, while it has the palatal fricative /ç/. Also the phonetic realization of many consonant phonemes differs to various degrees. The biggest difference in phonology, however, is found in the vowel system since Norwegian has not seen a change similar to the English Great Vowel Shift. On the other hand, all Norwegian vowels come in pairs of long and short vowels. This means that cognates between the two languages very often differ in pronunciation precisely in the vowels. This is, for example, a difference between English *cake* /keɪk/ and Norwegian *kake* /kɑːkə/. Thus, spoken comprehension between the two languages is complicated even with etymologically related and orthographically very similar words.

Grammatically, Norwegian and English are typologically similar. Both are analytical languages, but the inflectional system of Norwegian is richer than that of English. Among the main differences is the fact that while English is an SVO language, Norwegian has V2 word order in declarative main clauses, i.e., with the finite verb always as the second element. This causes differences in a range of sentences including adverbial-initial main clauses and topicalizations, where there is subject-verb inversion in Norwegian (cf. Faarlund, Vannebo, & Lie, 1997 ch. 9). However, with subject-initial clauses, i.e., the majority of declarative sentences,

word order is generally the same in the two languages. Another difference in word order is that Norwegian lexical verbs display the same movement properties as auxiliaries, meaning that no mechanism similar to English *do*-support is needed. In morphosyntax, the basic system reflects the close relationship between the languages. For example, the system of verb tenses, voice, and mode is very similar, as is the case system, which is only overt for the genitive in nouns, and for subject versus object forms in pronouns. There are differences in the noun system, where Norwegian has more regular declensions, three genders, and adjective-noun agreement. The definite article in Norwegian is a morphological suffix, not a free morpheme as in English (cf. Faarlund et al., 1997 chs. 3 & 5). In the verb system, Norwegian has more classes of regularly inflected verbs, and displays no subject-verb agreement nor a progressive aspect (cf. Faarlund et al., 1997 ch. 7).

In sum, the two languages are similar enough that for Norwegian learners of English, we can expect positive transfer from the L1 in a number of respects. The relatedness and similarity of the two languages is a relevant point in that linguistic distance, particularly similarity in commonly used words, has been found to be an important predictor for acquisition in young FL learners (Lindgren & Muñoz, 2012). However, successful comprehension of L2 English for Norwegians must still entail acquisition of both lexical, phonological, and morphosyntactic knowledge.

3.2.2 English in Norway

During the past 60 years or so, English has become by far the most widely spoken language in the world. Non-native speakers of English now vastly outnumber native speakers (cf. Lewis, Simons, & Fennig, 2013). Reasons are historical, political, and cultural. British colonization from the 17th to the 20th century brought the language to all continents. Since World War II English-speaking countries, in particular the US, have been leading culturally and politically, and English-language media such as movies, television shows, and music have spread globally. English competence is increasingly important across the globe. In Europe, English is by far the most commonly taught FL, and starting ages in schools have been lowered over the past two decades (Eurydice & Eurostat, 2012). This development reflects not only the global role of English-speaking countries, notably the US. The continuously increasing focus on English is more than anything a result of the fact that English seems to have, at least for the time being, won the contest to become the world's leading international language.

The context of the four papers

A common way to describe the global roles of English is Kachru's (1985, 1992) model of three concentric circles. *The inner circle* in this model consists of countries where English is the primary language, such as in the US and the UK. In *the outer circle*, English has official status but is used alongside other, local languages; these countries are typically former British colonies. In countries in *the expanding circle*, English has no official status but is a widespread FL. By traditional definitions, Norway falls into the latter category, as do very many of Norway's political allies, business partners, tourist markets, and holiday destinations. In the European Union, for example, English has become the dominant lingua franca of the union's business, and French, which was traditionally considered the language of the union, has lost ground. Among the population of the Union, 34% report that they speak English to some degree, compared to only 12% and 11% for German and French respectively (Eurobarometer, 2005).

However, the situation of English in Norway is more complicated than that of a traditional FL. On the one hand, the vast majority of Norwegian children come to school speaking Norwegian, sometimes in addition to a minority language, as their L1, but with no or very little knowledge of English. English has no official intra-national role. However, since World War II, English has been the first and most important FL learned in Norwegian schools, and its position is becoming constantly stronger. It is taught from year one in primary school, and is currently the only compulsory FL in the obligatory school system. For a majority of Norwegians, it remains the only FL which can be used with any degree of fluency, and the current curriculum does not even officially define it as a foreign language (Utdanningsdirektoratet, 2006, 2013a). In fact, Robert Phillipson already in 1992 described a shift for English from foreign to second language status in the Nordic countries (Phillipson, 1992).

English is ubiquitous both in private and professional contexts in Norway. Anglophone TV shows and films are very common, and these usually come with Norwegian subtitles, but are not dubbed (MCG Media Consulting Group, 2009; Seip & Arnt, 2004). With new platforms of media consumption, such as streaming and downloading, it is increasingly common, especially for young Norwegians, to watch such media even without Norwegian subtitles. English is present in advertising, such as in posters in stores and television ads (Johansson & Graedler, 2002). Magazines and books in English are readily available, and while literature is still largely translated into Norwegian, new non-Norwegian literature is normally available in English before in Norwegian. Thus many Norwegians read the English versions rather than wait for the

Norwegian translation; this is not least the case with Anglophone bestsellers (cf. Nordbø, 2013). English-language music is also ubiquitous, both by foreign and Norwegian artists (cf. T. Lie, 2004). Finally, computer games aimed at adolescents and adults generally contain English dialogue only, and it is not common for menus and instructions to be available in Norwegian (cf. Boas, 2013). With the vast majority of internet pages in English, and a very small percentage in Norwegian, this is another source of English input for Norwegians. Norwegian itself is clearly under heavy influence from English, and Norwegians regularly use English words and expressions when they speak Norwegian to each other (Graedler, 2002; Norås, 2007).

Also professionally, many Norwegians use English regularly. The use of English in academia is increasing, and virtually all higher education makes extensive use of compulsory readings in English. Also more and more English-medium programs are introduced (cf. Broch-Utne, 2001; Ljosland, 2007, 2011). In business, there is concern about the widespread use of English, a concern which, for instance, led the Confederation of Norwegian Enterprise (NHO) and the Language Council of Norway (Språkrådet) to cooperate on developing specific advice to promote the use of Norwegian instead of English for intra-national purposes. This entails promoting Norwegian for advertising and job announcements for a Norwegian target audience, and for internal communication in Norwegian businesses (Språkrådet & NHO, 2012).

All these facts make the Norwegian situation particularly interesting to study in the context of multilingualism. Particularly, while the facts above illustrate the significance of English in Norway, and while Norwegians are generally proud of their high English proficiency level, studies (e.g., Hellekjær, 2005, 2007, 2010, 2012) show that many Norwegians have problems reading literature or following lectures in higher education in English, and that Norwegian companies experience problems in international business because of a lack of English skills. Such findings make it natural to ask what can be done to further improve the English competence of Norwegians, and whether it is possible to take better advantage of the young starting age for English in Norwegian schools.

3.2.3 English teaching in Norway

English was introduced from grade 1 in Norwegian primary schools in 1997, and with extended teaching hours from 2006 (KUF, 1997; Utdanningsdirektoratet, 2006). However, not much time is allotted to the language; for grades 1-4, the total number of hours of English teaching is 138, i.e., only 34.5 hours per year on average (Utdanningsdirektoratet, 2013a). With a

provision that teaching should take place for at least 38 weeks per year, this corresponds to less than one hour per week (Utdanningsdirektoratet, 2007). For the entire primary level, the total number of English teaching hours is 366, which out of a total number of teaching hours of 5234 (Utdanningsdirektoratet, 2013c) amounts to slightly under 7% of the total teaching time. This is a normal percentage compared to other European countries (Eurydice & Eurostat, 2012). Some projects to increase exposure to English, for example, Content and Language Integrated Learning (CLIL) have been implemented in Norway over the last few years and may provide increased exposure to English also in other subjects in some schools (European Commission, 2009; Eurydice, 2006; Pérez-Cañado, 2012; Svenhard, 2010, 2012). However, these are typically introduced at secondary level, and certainly not as early as grade 1.

In addition to the low number of teaching hours, another concern about the teaching of English in Norwegian primary schools is the fact that the subject is usually taught by a generalist teacher, typically the homeroom teacher. English is not a compulsory subject in Norwegian general teacher training, and there are also no requirements for language competence or formal English qualifications at this level. Thus, many teachers teach English without formal training (see e.g., Drew, 2004; Drew, Oostdam, & Toorenburg, 2007; Drew & Vigrestad, 2008; Eikrem, 2012; Eurydice & Eurostat, 2012; Kunnskapsdepartementet, 2006, 2008; Lagerstrøm, 2007). Since many teachers are not trained second/foreign language teachers, not only language competence but also knowledge of SLA may be limited. Lack of fluency, insecurity, and lack of knowledge of the importance of target language exposure may lead many teachers to avoid the use of English in the classroom; Drew et al. (2007), for example, found that most Norwegian primary teachers only partially agree that their communicative competence and language skills are sufficient for use in the English classroom.

Target language use in English in grade 1 has not, to my knowledge, been systematically studied in Norway. However, the hypothesis that low formal English competence among teachers will lead to extensive L1 use is supported by reports (Drew, 2004; Drew et al., 2007; Eikrem, 2012; Flemmen, 2006) that a relatively large percentage of primary school teachers use mainly Norwegian in their English classes. Also Waara's (2003) report indicating that grammar teaching largely takes place in Norwegian even in upper secondary English classrooms, and Nicolaisen's (2011) master thesis which found target language use in lower secondary school to vary and depend partly on teachers' formal education in English, indicate the same. Teachers in grade 1,

who lack formal training more often than those in higher grades, can be assumed to be among those who use English the least in the classroom.

The limited role of target language communication and exposure in early Norwegian English classrooms is reflected in the curriculum. The current national curriculum was originally introduced in 2006 (Utdanningsdirektoratet, 2006). It does not specify teaching methods, only competence aims at various stages. The first competence aims for English are after grade 2, i.e., after two years of English instruction. In the original version of the curriculum, which was valid during the test period in the present project, relevant aims after year 2 were that students should be able to "understand and use some common English words and phrases", " use the most basic English phonology [...]", "greet people, ask questions and answer simple oral questions", "understand simple instructions given in English", "recognise some words, expressions and simple sentences in spoken and written texts", and "use numbers in communication" [7]. Although there are elements of oral comprehension in some of these aims, they lend themselves well to a style of teaching based mainly on the L1, where L2 words and expressions are treated as objects to be learned rather than as parts of natural communication. Unlike the previous curriculum (KUF, 1997), this curriculum does not give recommendations for activities and processes, only aims.

Teaching materials for grades 1 and 2 published at the introduction of this curriculum, such as students workbooks, CDs and their associated web pages (e.g., Bruskeland & Ranke, 2005, 2006; Håkenstad & Vestgård, 2006; Lillevold & Moseng, 2006; Lillevold & Whittaker, 2006; Munden, Elind, Engvall, & Oscarsson, 2006) contain very little incentive for the use of language in context, but instead focus on activities which require simple routine instructions and largely treat words in isolation. Typical activities entail matching words and pictures in various forms. There is also focus on rhymes and songs. It is likely that these activities are mainly intended to provide motivation, in that they are the kinds of language activities which children think are fun.

The curriculum has recently been revised (Utdanningsdirektoratet, 2013a), and the new learning aims are slightly more ambitious. Objectives for oral communication after year 2 are that students should be able to "listen for and use English phonemes through practical-aesthetic forms

[7] This version of the curriculum is not in effect, and thus it is no longer available in English. However, quoted here are what were the official translations of the Norwegian objectives which can be retrieved from http://www.udir.no/kl06/ENG1-01/Kompetansemaal/?arst=372029328&kmsn=382571589.

of expression", "listen to and understand basic instructions in English", "listen to and understand words and expressions in English nursery rhymes, word games, songs, fairy tales and stories", "understand and use some English words, expressions and sentence patterns related to local surroundings and own interests", "greet people, ask and respond to simple questions and use some polite expressions", "participate in simple rehearsed dialogues and spontaneous conversations related to local surroundings and own experiences", and "use figures in conversation about local surroundings and own experiences" (Utdanningsdirektoratet, 2013a, p. 7). The curriculum states in the introduction to the English subject that "[l]anguage learning occurs while encountering a diversity of texts, where the concept of text is used in the broadest sense of the word" (Utdanningsdirektoratet, 2013a, p. 2).

With the new curriculum, new learning materials have been published, some of which may pay more attention to the importance of real communication and input. The Explore series (Edwards, Flognfeldt, Moen, Nilsen, & Møkleby, 2013), for example, uses pictures with accompanying CD texts where children listen to a text and look for familiar words in the picture. The Quest series (Hansen, Mørner, Næss, & Pritchard, 2012) specifically gives the teacher suggestions for instructions and classroom management which can be used outside of formal English class. However, the suggested volume of input is still very low, and a substantial increase in target language exposure for the students thus minimally depends on teachers emphasizing such activities.

Learning materials in the form of books and associated CDs and web pages of course are not the only resource available to teachers. However, combined with a lack of formal training in English, it is very likely that what seems to be prescribed in the available materials is also what is taught (cf. Drew et al., 2007; Flemmen, 2006). Many schools publish teaching plans on their web pages, including what they are working with in each subject in different weeks. This allows us to gain an (admittedly superficial) impression of the topics and the level of English expected in many schools at once. Common to such plans is that if they publish goals for English learning, they do so by exemplifying what vocabulary items they are working with at various times. In one randomly sampled school for the school year of 2013-14, students in grade 1 are working with the numbers 1-10 in the last week of October, i.e., after more than two months of school. Two weeks later, in mid-November, they report that the children are to learn the colors red, blue, green, and yellow. In mid-January, the goal is to learn some words for body parts in English. This

is representative of plans found on the web pages of different schools, all of which reflect clearly limited goals for lexical learning connected to specific topics. There is little doubt that part of the background for the early start in English in Norwegian schools is knowledge of the benefits of a young AoA for language acquisition. However, it is not clear that the way the language is introduced at this point is based on the knowledge we have about the effect of age on acquisition beyond the simplistic "younger is better".

As discussed above, English is ubiquitous in Norway. A common assumption is that this is also the case in the lives of young children and that they may be picking up the language from outside of school. This, however, is not necessarily true, especially before they are fluent readers in Norwegian. While Anglophone programs and films for adolescents and adults are generally in English with Norwegian subtitles, dubbing is more common for young children. This is the case for all the major children's TV channels and for films shown in Norwegian movie theaters aimed at young children (cf. Bjørkeng, 2012; Ursin, Aursland, & Ripegutu, 2012). Also computer games for very young children still mainly come in Norwegian, and while popular music in Norway is often in English, songs aimed specifically at young children are almost exclusively in Norwegian.

3.2.4 Previous research on the acquisition of English in Norway

The widespread use of English in Norway as well as its central place as a school subject means that the acquisition, teaching, and use of English in Norway have been studied from various angles. However, as mentioned in section 3.2.3, research relevant to the first years of primary school is still relatively scarce. Most research on English teaching and learning in Norway has been carried out for secondary level only. In 2007, a network was organized for researchers working with students at primary level, so-called Young Language Learners (Drew & Hasselgreen, 2008), and in 2012, an edited volume focusing specifically on young learners in Norway was published (Hasselgreen, Drew, & Sørheim, 2012). However, this volume included the secondary level, and none of the contributions focused specifically on the earliest grades. The individual studies are reported in the following, along with other previous research relevant to the present project. One study which will be discussed used the PPVT-IV at the upper primary level; except for this, to my knowledge, no previous study has employed the same methodology as this project, nor investigated the same aspects of language competence in a comparable manner, even for older learners.

The context of the four papers

Research on English learning and use in Norway can roughly be divided into two types; first, studies dealing with Norwegian speakers' acquisition of and competence in English, and second, studies dealing with the structure and content of English teaching in schools, including curricula, assessment practices, and methodologies.

Among the first type we find consistent reports of high levels of English proficiency in Norway compared to other countries (cf. Bonnet, 2004; Breivik & Hasselgren, 2002; EF Education First, 2013), and along with them, discussions of the strong position which English has gained in many areas of Norwegian society, e.g., academic life (Broch-Utne, 2001; Ljosland, 2007, 2010, 2011). Gooskens (2006) found that Norwegian teenagers understand spoken English better than Danish, a language very similar to Norwegian; although the focus of this study was on intercomprehension in the Scandinavian languages, it does provide an indication of high listening comprehension in English for Norwegian teenagers. However, although Ibsen's (2004) report on the English competence of Norwegian grade 10 students found high proficiency levels compared to other European countries, there was great variation and a strong correlation with family background and media use, indicating that much English learning probably takes place outside the classroom. Hellekjær (2005, 2008, 2009, 2010), as mentioned in section 3.2.2, found that a significant percentage of Norwegian students experience problems in reading academic materials and attending lectures in English, a finding similar to Lehmann's (1999a, 1999b) conclusions for students' writing. Another study already mentioned found that Norwegian companies experience problems as a result of insufficient English proficiency (Hellekjær, 2012).

In the school context, Olsen (1999) found widespread L1 influence in the texts of lower secondary students, while Drew (2003) found less syntactic complexity in English than in Norwegian in the writing of a group of grade 7 students. They both attributed their findings to a lack of target language exposure and low teacher competence. A highly relevant study of Norwegian students' general English competence is that of Abildgaard and Helland (2011), who compared the English proficiency of two groups of primary level students (grades 6 and 7) on a series of measures. One group was taught according to the first version of the new curriculum (Utdanningsdirektoratet, 2006), and the other was taught according to the preceding curriculum (KUF, 1997). Data were collected from both groups as part of the validation process for an English dyslexia test, described in a number of studies (cf. Helland, 2008; Helland & Kaasa, 2005). Abildgaard and Helland's study included only participants without dyslexia or any other

language-related diagnosis. The group who had been taught according to the new curriculum had on average had more hours of English instruction than the group taught according to the old curriculum, and the hypothesis was thus that their performance would in general be better. However, their scores were similar to or even lower than the group taught according to the old curriculum on every measure except quantity of oral production. Although this study has limitations, e.g., in that both groups were in fact attending school in transitional periods between curricula, the results indicate that closer scrutiny of the effects of the new curriculum is needed.

Among studies of the acquisition of specific aspects of English are Westergaard's (2002, 2003) studies of the acquisition of word order in young classroom learners, finding persistent transfer of V2, but also frequency effects and effects of lack of relevant input, specifically for sentences with *do*-support. These studies initially included children at all levels of primary school; however, all first-graders had to be excluded from the analysis because they had no relevant knowledge of English. A few other studies of specific aspects of English acquisition in primary school exist, but for grades 4 and up, e.g., the development of fluency and grammatical and lexical complexity in written English (Drew, 2010), or phonological competence in production (Rugesæter, 2012).

Langeland (2012) is, to my knowledge, the only study not related to the present project to employ the PPVT-IV to study English receptive vocabulary in Norwegian learners. She investigated vocabulary development from grade 5 (age 9/10) to grade 7 (age 12/13) in 40 students, and found a vocabulary increase of 26 months (from 4;6 to 6;8) on average in the course of two years of school. As in the bilingually-based group in the present study, then, learners showed faster development than would be expected in younger, vocabulary-matched native speakers. The group in Langeland's study had not had a specific focus on input similar to the bilingually-based group in the present project, but were reported to have read relatively extensively in English during the test period. This may have contributed to their rapid acquisition compared to the younger children at the same stage of vocabulary development in the PPVT-IV reference group. The fact that they were old enough to be literate also means that they are likely to have been exposed to English through media outside of the classroom (cf. section 3.2.2). Langeland did, however, find great individual variation in development.

Other studies have focused on acquisition and competence at the secondary level, e.g., on processing strategies related specifically to voicing contrasts, on complexity and accuracy in

writing, on the use of specific words or constructions, or on pragmatic competence (Brubæk, 2012; Flognfeldt, 2011; Hasselgreen, 2002, 2004; Helness, 2012; Raaen, 2009, 2010; Raaen & Guldal, 2012; Torgersen, 2002, 2003; Waara, 2004). Finally, a number of studies exist of the acquisition and competence of a range of specific linguistic features in university students or adult L2 users (Behrens, 2008; Dahl, 2004; Hasselgreen, 1994; Hasselgård, 2009a, 2009b, 2012; Iverson & Evans, 2007; Krave, 2012; Nacey & Graedler, 2013; Rugesæter, 1995; Rørvik, 2013; Rørvik & Egan, 2013; van Dommelen & Hazan, 2010; Wilson & Mollergard, 1981).

In the second category of studies, i.e., those focusing on English teaching, there are a few studies of the system of English teaching in Norway per se, but these are mostly too old to be relevant to the present system (e.g., Merkies & Tuin, 1987; Schei, 1977; Van Buskirk, 1979). Among more recent work is Simensen's (2010) review of curricula, which found that fluency only explicitly received a place in the English curriculum after 2006. Drew, Oostdam, and Toorenburg's (2007) comparison of the systems in Norway and the Netherlands found relatively low teacher competence and high reliance on text books in Norwegian primary classrooms, but found that the transition from primary to secondary level is smoother in Norway than in the Netherlands. Norwegian data in this comparison were based on a survey among Norwegian teachers (Drew, 2004). Reliance on the text book was also found by Ibsen and Hellekjær (2003) in a survey of grade 10 English teachers. In other studies of teacher competence, Drew and Vigrestad (2008) found that most teachers at lower primary levels did not have formal qualifications to teach English, while data in Lagerstrøm's (2007) survey of teacher competence in all Norwegian school subjects indicated that the percentages of teachers who taught English without formal qualifications was rising. Mellegård and Pettersen (2012) interviewed a group of teachers about the new (2006) curriculum and found that it gives teachers a freedom of interpretation for which they do not always feel competent. There also exist some discussions of Norwegian teacher training in relation to English as a subject, e.g., R. E. Lund's (2014) discussion of writing requirements in English in teacher training courses in Norway. Simensen (2011) discusses the influence of a number of academic "parent disciplines" on English teaching in Norway, underlining that this is a complex subject comprising more learning objectives than just L2 proficiency[8].

[8] The English subject in Norwegian schools includes not only language proficiency but also areas such as cultural knowledge and literature. Studies dealing with the teaching of such content are not discussed here.

Among studies of teacher practices, one report directly relevant to the present project is that of Flemmen (2006), who observed English classrooms in grades 1-4. She found great variation in teachers' approaches, but observed many of the patterns outlined in section 3.2.3, i.e., predominant use of Norwegian rather than English, discussion of vocabulary items in isolation, formulaic language, and coloring activities which have little to do with language learning. Eikrem (2012) investigated English teaching and students' experiences in grades 3, 6, and 9, and found, among other things, that students spent a large portion of time in English classes listening to the teacher, whose use of the target language during English classes ranged from 19% to 37%. In other words, students spend a significant portion of these classes listening to the teacher speaking Norwegian. Teacher-centered communication was, predictably, especially dominant for the youngest learners, who were also found to be the group who enjoyed English classes the least. Eikrem hypothesizes that this may be due to low teacher competence and a lack of target language use. Eikrem's (2006) PhD dissertation investigated a range of perceptions and attitudes in English teachers at various levels, including primary, at the time of the previous syllabus (KUF, 1997), and in particular found a need for pre-service and in-service training. Hestetræet (2012) studied teachers' approaches to vocabulary teaching in grade 7, and found that while teachers generally believe in learning vocabulary in context, there was a focus on output rather than on input, and also heavy reliance on the textbook. Activities facilitating implicit learning through input were among those least frequently used. Finally, Waara (2003) found, among other things, problems in integrating grammar in classes at upper secondary level.

A number of studies exist of the use of specific methodologies, e.g., for reading (Birketveit & Rimmereide, 2012; Charboneau, 2012; Drew, 2008, 2009; Drew & Pedersen, 2010, 2012), the use of ICT (A. Lund, 2006), the use of songs and music (Langeland, 2013; R. E. Lund, 2012), lesson organization and task complexity (Waara, 2009, 2011), and the use of the European Language Portfolio (ELP) (Larssen, 2010). However, none of these studies focus on grades 1 and 2; most deal with the secondary level. Also a number of studies of assessment and examinations exist (Blair, Moe, & Barsnes, 2011; Chvala, 2011; Chvala & Graedler, 2010; Hasselgreen, 2000, 2003, 2004, 2012; S. Lie, Hopfenbeck, Ibsen, & Turmo, 2005; Skeiseid, 1995; Ørevik, 2012). Again, however, focus in all cases is on students who are older than those in the present project.

The context of the four papers

English teaching in schools has been described for special contexts providing more input, such as CLIL (Brevik, 2012; Drew, 2013; Hellekjær, 1994; Svenhard, 2010, 2012; Svenhard, Servant, Hellekjær, & Bøhn, 2007); again, these studies focus on learners significantly older than those in the present project. Hestnes (2006) investigated the use of international student teachers in secondary school, and found positive effects and evaluations. Surprisingly, the younger students (grade 8) were the most positive to being taught in English in other subjects.

A number of master theses have been written over the years on various aspects of English acquisition and teaching in Norway. Again, focus tends to be on older learners, and very few deal with primary level students. Thus, only some particularly relevant master theses are discussed here. As mentioned in section 3.2.3, Nicolaisen (2011) studied teachers' target-language use in the lower secondary English classroom, and found it to be highly variable and dependent on a range of factors including formal competence in English. Sivertzen (2013), in a follow-up study to the present project, found that in grade 4, i.e., approximately three years after the studies reported here, the students in the bilingually-based group, whose extra exposure to English was maintained through grade 2, showed no advantage on the PPVT-IV compared to a group of students who had received standard English teaching from grade 1. However, Strand (2014) tested the same two groups for sentence comprehension eight months later, when they were in grade 5, using an expanded version of the sentence comprehension test developed for the present project, and found an advantage for the bilingually-based group. This is interesting in light of the results in Paper 3 in the present study, which find the greatest advantage of extra input precisely for sentence comprehension. Strand also found an advantage for reading comprehension in the bilingually-based group.

Providing an exhaustive overview of all research ever conducted on English acquisition and competence Norway is virtually impossible, given the long history of the language as a subject in school, the complexity of this subject, and not least the varied contexts in which the language is learned and used in Norwegian society. However, this section has reviewed the existing research to the extent possible, with particular emphasis on including all research which is recent and relevant to the present project. The conclusion is that there are a number of studies of English teaching and learning in Norway, but that in general, they tend to focus on secondary or at least higher primary levels. Furthermore, a number of reports and studies agree that general English proficiency levels are high in Norway compared to other countries, but with English

being ubiquitous in Norwegian society, factors outside of school may be more important than the quality of English teaching for this result. Finally, studies of English proficiency among Norwegian students do find that it is not necessarily sufficient for their purposes. These are indications that we need more research on English learning in Norwegian classrooms, and that we cannot equate the young starting age with high-quality learning. In this context, the present study represents important new knowledge, in being the first which specifically investigates English acquisition in Norwegian grade 1.

3.3 Theoretical background – summary

As the background outlined so far indicates, it is the assumption in the present thesis that the human mind is prepared to acquire not only one, but multiple languages (cf. Meisel, 2004). In other words, it is assumed that there is nothing particularly natural or default about monolingualism. It is furthermore assumed that there is no fundamental difference between L1 and L2 acquisition, and that differences between monolingual and bilingual competence come in degrees where for individuals with two L1s acquired from early childhood, differences stem mainly from the coexistence of two language systems in the same mind, and in differences in input encountered and uses of the two languages (cf. Grosjean, 1989; Meisel, 2004). For L2 acquisition and use, more factors are involved; in particular age, e.g., brain plasticity and cognitive maturity (cf. Hyltenstam & Abrahamsson, 2003). However, also input factors and possibly the degree to which the L1 is already entrenched in the mind may cause differences making late bilingual competence different from monolingual competence to a greater degree than in the case of two L1s. The decrease in brain plasticity with age may contribute to making L2 acquisition a more laborious process, while cognitive maturity may be responsible for allowing for explicit learning in the L2. However, neither factor is assumed to be absolute in a sense that makes it meaningful to talk about an absolute critical period. We know that also the adult brain changes physically with experience (cf. Mechelli et al., 2004), and although explicit learning may facilitate L2 acquisition, it is still clearly best described as a process of *acquisition*, not *learning*, if we are to use Krashen's (Krashen, 1981b, 1985) distinction. This means that the crucial factor for all language acquisition is input.

Although no distinct critical period for language acquisition in the strict sense is assumed, AoA is assumed to be important, and a younger starting age is assumed to be beneficial for ultimate attainment. The age of approximately 6 or 7 may be important in this context, but

whether it marks a qualitative change in our acquisition abilities, i.e., in the sense of a sensitive period, or whether the decline in language acquisition abilities is linear and simply marked by a point at which competence can be expected to be below what is needed to be perceived as native-like (cf. Abrahamsson & Hyltenstam, 2006; Hyltenstam & Abrahamsson, 2003), is not clear, nor crucial for the present project. Finally, it should be pointed out that an obvious benefit of an early AoA in a new language is simply input; the earlier a person starts acquiring a language, the more years of input she can experience.

Crucially, it is hypothesized that there is no fundamental difference between FL learning and other types of SLA, and that language classrooms are essentially arenas for language acquisition. The idea is thus that differences between FL and other SLA come in degrees, and that FL learning depends on the same factors which are relevant to all SLA, i.e., the age of the learner, access to input, etc. Thus, English acquisition in Norway is in particular a FL situation with many shared characteristics with other types of SLA, because of the early AoA and the potential for accessible input given the position of the language in society.

4 Main findings and implications

This thesis addresses early child SLA in two main types of setting, namely an immersion setting and two different FL classrooms. We saw in Paper 1 that early-start L2 programs in school settings do not necessarily have an effect and that standard English classes in Norwegian grade 1 constitute a setting where acquisition is not guaranteed. This was evident from the lack of vocabulary development in the native-language based group. However, a modest increase in language exposure compared to the typical L1-dominated Norwegian classroom was found to have a significant effect on vocabulary development in the bilingually-based group, and receptive vocabulary in this context was found to develop slightly faster than in competence-matched native speakers of English at younger ages. This is an important finding since it may have consequences for the ambitions of early-start FL programs. It supports the hypothesis that FL learning is in essence a form of SLA, and that input plays a crucial role for acquisition. Input in such settings can always be expected to be limited. However, the present results indicate that this does not mean that we cannot achieve significant benefits from focusing on increased, better quality, and more naturalistic input, even within a traditional FL context.

A further overall finding of the studies in this thesis is that with very young, pre-literate learners, sentence comprehension and repetition tests can be good tests of early acquisition in addition to a vocabulary test such as the PPVT-IV. Papers 2 and 3 highlight how early L2 acquisition does not simply entail learning vocabulary items as translation equivalents from the L1, or as isolated items. Paper 2, which evaluates the sentence repetition test, suggests that phoneme perception and the ability to perceive lexical items in the speech stream have to be considered important L2 competencies in their own right, and Paper 3 shows how it is sentence comprehension, i.e., comprehension of words in context, not isolated lexical knowledge, which benefits the most from increased exposure. It is thus argued that processes such as phoneme discrimination, lexical retrieval, and building "good enough" representations for comprehension (cf. S. E. Carroll, 2004; Clahsen & Felser, 2006a; VanPatten, 2012) are important early steps in L2 development, and results in this study indicate that these skills develop early as a result of exposure.

For the seven children in the immersion group, we see that attending school for a year in a new language leads to rapid acquisition. In this group, also lexical acquisition and accuracy in

sentence repetition was generally closer to native-speaker means, indicating acquisition across the board. Although there was great variation between participants, we see again that exposure to an L2 at such a young age can lead to very fast development.

4.1 The nature of early L2 competence

Findings in the papers in the present study are relevant to the question of exactly what successful SLA is, i.e., what constitutes language competence in an L2. While pedagogical perspectives of language often focus on the four skills of listening, speaking, reading, and writing, which are useful as descriptions of multilingual competence (Cook & Bassetti, 2011), these are descriptions of overt language behavior, not of underlying competence. The findings in the present project are mainly relevant to listening, and emphasize that this in itself is a multifaceted skill.

From a psycholinguistic point of view, linguistic competence is normally described in terms of models of lexical and grammatical knowledge. Different models exist of how this knowledge is organized, and indeed of whether it makes sense to talk about two different systems. This debate is beyond the scope of this thesis, but for the discussion of the overall results of the present studies, a distinction between the two is useful, roughly assuming that the lexicon consists of structural knowledge which is at least partly subserved by declarative memory, while combinatorial rules including (regular) inflection is part of the mental grammar (cf. Clahsen, 1999; Pinker, 1999; Ullman, 2001). The exact nature of each component is not crucial, but as Ullman (2001) points out, mappings between (uninflected) words and their meanings are arbitrary, which must mean that they are to some extent memorized or "learned", unlike the meaning of larger linguistic units, which is derivable given lexical knowledge combined with grammatical rules. In Ullman's Declarative/Procedural model, the former system is an associative memory while the latter system, the mental grammar, subserves syntactic and morphological computations (Ullman, 2001). In this model, there is room for difference between L1 and L2 acquisition and use in the division of labor between the declarative and the procedural memory system, in that L2 acquisition may depend more on the former. However, this shift from procedural to declarative memory seems to depend on AoA and of L2 proficiency, again underlining the gradual difference between L1 and L2 competence. Differences may also depend on how the L2 was learned; Morgan-Short, Steinhauer, Sanz, and Ullman (2012) in fact found

native-like brain activation patterns in adults who had learned an artificial language through implicit training, but not in a group learning the same language explicitly.

Together, the four papers in the present thesis contribute to the picture of the language skills that make up early L2 proficiency in young learners, both in the lexicon and in the combinatorial system. Articles 1 and 4 focus especially on the speed of vocabulary acquisition displayed by young L2 learners in different contexts as compared to L1 acquisition. In explaining this speed of acquisition, it is important to keep in mind that the test employed, i.e., the PPVT-IV, only tests receptive vocabulary, and vocabulary breadth rather than depth. This is relevant to what the findings can tell us about what early L2 vocabulary acquisition entails. The speed with which English vocabulary is acquired by children both in the immersion context in Paper 4 and in the non-immersion bilingually-based group in Paper 1 is an indication that mechanisms are available which support this acquisition, aiding the development at least of receptive vocabulary. The mean vocabulary increase in the course of eight months in the bilingually-based group, as described in Paper 1, corresponds to 10 months in PPVT-IV age equivalents. In the immersion group described in Paper 4, mean scores increased from 47 to 83, which corresponds to an increase in age equivalents from 3;5 to 5;1, or 20 months in the course of eight months in chronological time. Thus, the findings are important for what we can expect of early lexical development both in FL teaching and for children starting school with low competence in the language of instruction.

There exist a number of models for the representation of the mental lexicon in the bilingual mind, which assume different relationships between the conceptual level and lexical items, and between the lexicons of the two languages. Early work includes Weinreich's (1953/1968) proposal that such models may either consist of shared conceptual representations in both languages, separate conceptual representations, or L2 access via the L1. The Revised Hierarchical Model (Kroll & Stewart, 1994), assumes that conceptual representations are shared between the two languages, while the lexicons are separate. There are connections between lexical items in the two languages, and between words in each language and their shared concepts, but connections between L1 and concepts are stronger than between L2 and concepts, and connections from L2 to L1 are stronger than vice versa. Especially at early stages, the L2 lexicon may be connected to L1 representations, while direct L2 access to the conceptual level

may develop with increased proficiency (Kroll & Stewart, 1994; Kroll & Tokowicz, 2001; Kroll, Van Hell, Tokowicz, & Green, 2010).

Another influential model of the bilingual mental lexicon is the Distributed Features Model, which sees word meanings as sets of features which may be more or less shared between the two languages (de Groot, 1992); this view of word meaning helps account for cross-linguistic differences. The model assumes no direct connections between words in the two languages in the bilingual mind, and their association thus depends on the shared features. These may depend on word type, for example, in more shared conceptual nodes for the two languages for concrete words and cognates than for abstract and non-cognate words (e.g., de Groot, 1995).

The results reported in this thesis cannot make any strong claims about the organization of the bilingual mental lexicon, nor of the nature of conceptual knowledge and its relationship to lexical encoding in two languages. However, the rapid vocabulary development of early L2 learners compared to native speaker norms in the present studies may be an indication that L2 lexical acquisition is supported by L1 knowledge, e.g., by being initially mediated through the L1 (cf. Sunderman & Kroll, 2006). Jiang (2000, 2002, 2004), for example, suggests that the entry for a new word in early adult SLA initially consists merely of information about form (phonology and morphology), and a pointer to the L1 translation equivalent for meaning and syntax. During the acquisition process, this pointer may first be replaced by transferred lemma information from the L1 word. At more advanced stages, the L2 lexical entry may gradually become fully specified, with independent strong links to the conceptual level, while connections to L1 translation equivalents are weakened. We certainly do not have evidence here to evaluate this proposal in full, and it was not proposed for young learners such as those in the present project. However, it is possible that what has been acquired by the bilingually-based group in this study is some kind of pointer to the L1 lexicon along the lines of Jiang's model, since only receptive vocabulary breadth, not depth, was investigated. This might explain the rapid increase in vocabulary scores compared to native-speaker norms. What happens at later stages of development, i.e., whether there is a transition to independent mappings for the L2, and whether word meanings are restructured to become more target-like and less influenced by the L1, is not really tested by the PPVT-IV[9].

[9] There is also no reason to assume that proficient L2 users are characterized by independence of the L2 and L1 lexicons, since some level of crosslinguistic lexical activation seems to be the norm for all bilinguals (cf. Kroll, Gullifer, & Rossi, 2013).

The declarative aspect of the mental lexicon clearly does not entail that lexical knowledge must be acquired explicitly. It is also important to note that while L1 lexical and conceptual knowledge may have supported lexical acquisition in all groups of learners in the project, this is not to be taken as an indication that the L1 was explicitly invoked. Learning in the bilingually-based and immersion groups did not rely on L1 translation but mainly on L2 input. Thus, it may be that the rapid lexical acquisition in these groups reflects not only L1 support but also the presence of mechanisms known to support L1 lexical acquisition in young children, e.g., fast mapping (Shintani, 2011). Together, the support of the already well developed L1 lexicon and general mechanisms facilitating vocabulary learning in young children may form a powerful combination for the acquisition of L2 receptive vocabulary in the age group in question, explaining their very rapid development on the PPVT-IV as compared to native speaker norms.

For online spoken language comprehension, the use of lexical knowledge minimally entails segmenting the speech stream, i.e., perceiving the words, as well as accessing the relevant items from the mental lexicon (S. E. Carroll, 2004). The finding in Papers 2 and 3 especially is that this processing in particular benefits from increased exposure to the target language. This highlights how L2 competence is not just a matter of structure, but also of processing (cf. Harrington, 2001). Harrington (2001, p. 94) outlines six steps in sentence comprehension: 1) segmentation of the sound stream into linguistic units, 2) accessing individual lexical items in the segmented string, 3) assigning a syntactic structure to the word string, 4) deriving a meaning for the words and syntactic structure as a unit, 5) establishing the real-life referent of the string, and 6) recovering the speaker's intention. According to Harrington, sentence processing research tends to focus on the middle steps (3 and 4), i.e., syntactic analysis and deriving meaning, with little focus on the previous steps in the process. Papers 2 and 3 in the present thesis show that the first two steps, i.e., segmentation and lexical access, may be important in order to arrive at a competence where the middle steps of syntactic and semantic analysis are possible, and thus that the very earliest stages of language development may be worthy of closer scrutiny.

We have seen in the present thesis that while measures of comprehension indicate that L2 exposure makes a significant difference even in the short run (in this case, in a matter of 8 months), tasks requiring more detailed grammatical processing, i.e., the repetition test in the present project, may not show the same difference. It is also possible that the smaller difference between the native-language based and the bilingually-based group on this measure is connected

to the productive aspects of the test. This, however, does not necessarily mean that exposure is "not enough" or will lead only to development of receptive skills; after all, we know from L1 and L2 acquisition alike that comprehension precedes production. It may be that explicit instruction or "pushed output" (cf. Swain, 1985, 1995; Swain & Lapkin, 1995) is necessary for the aspects of competence tested by the repetition test, but it may also be that more important is making sure that the input is maintained over time, so that exposure is sustained and substantial (Lightbown, 2000; Ruiz-González, 2006).

Based on Snow and Hoefnagel-Höhle's (1977) finding that young learners are initially slower, and need approximately one year to catch up with older learners in an immersion setting, Singleton and Ryan (2004, p. 200) estimate that normally instructed young learners even with relatively frequent language classes would need 18 years to achieve the same results, given the amount of input to which they can be expected to be exposed. However, this may be taking the quantification of input too far. It is absolutely plausible that input interacts with chronological time, i.e., that providing sustained input over time is more important than the absolute intensity of the input, as long as it is sufficiently substantial to have an impact at all. After all, we know that absolute volume of input is not crucial in L1 acquisition, and that e.g., bilingual children are not significantly delayed compared to monolingual children, even though time of exposure to each language can on average be expected to be less substantial than monolingual children's exposure to their one language. It is also worth remembering that "hours spent in an environment where the language might potentially be spoken" does not necessarily equal "hours of input". For example, we have already seen that "hours of teaching" in a FL need not mean "hours of input" since if the L1 is the language of instruction most of the time, learners may actually be exposed to the L2 in the classroom for only a fraction of the total of teaching hours. Thus, with more longitudinal study of the L2 development of a group of learners exposed to the target language in a manner similar to the situation of the bilingually-based group in the present project, beyond the first years of school, we might learn more about the volume of input needed for the development of different aspects of L2 proficiency.

4.2 The contribution of input to early L2 language acquisition

Taken together, the papers in this thesis shed some light on exactly what we can expect the role of input in SLA to be, especially at the very early stages. The debate on the role of input versus explicit instruction was outlined in section 3.1.2 and in section 3.1.2.2 in particular, but

this was in general terms, and not really addressing the question psycholinguistically in terms of processing (cf. VanPatten, 2012). The question thus remains of exactly how input contributes to acquisition. Any current theory of language acquisition will agree that acquisition happens through the interaction of input with specific or general mechanisms. However, what those exact mechanisms are, and how input works on different aspects of language competence, is debated. For example, it is common to assume a distinction between input and intake, where the latter depends on attention or "noticing" (e.g., Corder, 1967; Long, 1996; Robinson, 1995; Schmidt, 2001). However, studies have found that acquisition of various aspects of language can take place in the absence of attention to the input in both children and adults (Saffran, Newport, Aslin, Tunick, & Barrueco, 1997; Schachter, 1998; Truscott, 1998). S. E. Carroll (1999) points out that language learning is not the same as speech processing, and the term *input* is thus ambiguous in that input to language learning mechanisms is not the same as input to speech processing mechanisms. She views speech processing as a chain of representations beginning with the auditory signal and ending with a conceptual representation, where the output of each stage of processing serves as the input to the next stage. Within levels of processing, she argues, the conversion of input to intake may be a process to which it is not possible to pay attention. In her model, the role of learning mechanisms is specifically to resolve processing problems; (attempted) learning is automatically triggered when processing fails[10]. Input to the learning mechanisms thus consists of four elements: the partially analyzed parse representation, the unanalyzable item, the current set of parsing procedures, and whatever information external to the relevant processing level can be utilized to resolve the problem. What specific information this is depends on our theory of language; it may, for example, come from UG, from analogy or generalization mechanisms, or from inferences drawn from negative evidence or other mechanisms such as "negotiation of meaning" (S. E. Carroll, 1999, p. 365). Thus, in Carroll's view, attention cannot simply be seen as a function which selects information from the speech signal to feed to the learning mechanisms.

The children both in the bilingually-based group discussed in papers 1-3 and in the immersion group discussed in papers 2 and 4 had all acquired both receptive vocabulary and sentence processing and comprehension skills from naturalistic input, and without any effort on

[10] The qualification that such learning is "attempted" means that there is no guarantee that the restructuring resulting from the application of these learning mechanisms, allowing a successful parse, is "target-like" from the perspective of the L2.

the part of the teachers to draw their attention to form in the input. Input may have been adapted to the learners in being simplified, with emphasized intonation, and speech being particularly slow. Yet, there is nothing unnaturalistic about such adapted input; it resembles child-directed speech in L1 acquisition. Just like it is possible that such speech modification may be beneficial to young children acquiring their native language (Yurovsky, Yu, & Smith, 2012), it may facilitate SLA.

There is little evidence in the present studies to answer the question of whether also more detailed grammar competence can be achieved based on such naturalistic input, and without focus on form. However, the results of the immersion group in the repetition test reported in Paper 4 may be an indication. Four out of seven children score within the range of the native speaker control group on the repetition test, an indication that the necessary skills can indeed be acquired from input. The study reported in this paper found intra-subject differences between the three measures of English (vocabulary, sentence comprehension, and sentence repetition) after eight months, and there is thus no straightforward relationship between e.g., vocabulary size and accuracy in the repetition test. However, it must be assumed that the high score of most participants in this group has come about based on their substantial exposure to the target language. They were "pushed" to produce language in the sense that the language of communication in their classroom was English and they could not expect all classmates to understand any of their other languages, but there was no focus on grammatical accuracy in these classrooms, and naturally no relevant explicit instruction, given that most of the students in each class were already fluent English speakers. The exception is one participant who attended EAL classes for a time. However, the fact that also three other children scored within the native speaker range on this test may indicate that also grammatical accuracy can in principle be acquired from naturalistic input in young L2 learners.

The role of input in SLA is likely to be connected to frequency. Frequency effects have been found for the acquisition of all areas of language, including phonology, vocabulary, morphology and syntax in children and adults. We know that the frequency with which items are encountered influences their recognition and recall. With increased input, both type and token frequency increases, which may facilitate statistical learning. While there is disagreement about the degree to which attention to the input is necessary for such learning to take place, it is clear that the mechanisms operate on mere exposure, and that attention to the input in this context does

not equal attention to specific forms in the input (N. C. Ellis, 2002, 2006a; Lieven, 2010; Mirman, Magnuson, Estes, & Dixon, 2008; Pelucchi, Hay, & Saffran, 2009; Saffran, 2001; Saffran, Newport, & Aslin, 1996; Thompson & Newport, 2007; Toro, Sinnett, & Soto-Faraco, 2005; Vouloumanos, 2008; Yu & Ballard, 2007; Yurovsky et al., 2012).

4.3 The contribution of other factors

While the main focus in all the studies in the present thesis was the role of input for L2 acquisition, a number of other factors were checked for their influence on acquisition. In the individual papers they were used mainly to control for differences between the two state school groups (i.e., bilingually-based and native-language based), to make sure that these groups were as closely matched as possible on background factors. In Paper 3, the impact of pre-test English vocabulary and Norwegian vocabulary on post-test English sentence comprehension is also compared to the impact of increased input (i.e., group membership) through multiple regression. In Paper 4, correlation analysis is conducted to look for relationships between background factors and acquisition in the immersion group. In the following, a closer look will be taken at the relationship between background factors and acquisition, i.e., correlations between control measures and English measures.

We will look first at the the immersion group, where six out of seven children were multilingual prior to acquiring English. We saw in Paper 4 that their number of L1s did not negatively impact on the acquisition of English. It is difficult to say anything about how L1 skills influenced acquisition in this group since no measures of the child's strongest L1 were administered. Norwegian vocabulary was tested, and a correlation analysis (Spearman's ρ) [11] in fact found a correlation with post-test English vocabulary (i.e., the mean score of the PPVT-IV forms A and B), ρ=.692, p=.043. Also the verbal ability measure, i.e., the Riddles section of the Kaufmann Brief Intelligence Test, 2nd edition (K-Bit 2), was correlated with post-test vocabulary in the immersion group, Spearman's rho .963, p<.001. Interpreting these correlations is difficult since no tests of language dominance were administered, and we thus do not know anything about the children's proficiency in their other languages. However, we have an indication that there is a relationship between language skills in different languages in multilinguals.

In the two state schools, e.g., the native-language based and the bilingually-based group, all subjects were monolingual, and correlations with L1 vocabulary and verbal ability are easier

[11] Non-parametric tests were used for the immersion group because of its small size (n=7).

to interpret. With Pearson's *r*, such correlations were found in the native-language based group between Norwegian and English vocabulary (the latter score being the mean of the two forms of the PPVT-IV in the post-test), r=.525, p=.002, and between the verbal ability measure and post-test English vocabulary, r=.550, p=.001. In the bilingually-based group, however, no significant correlation was found between Norwegian and post-test English vocabulary, r=.248, p=.089, while a moderate correlation was found between KBIT-2 Riddles and post-test English vocabulary, r=.337, p=.032.[12]

We obviously do not know whether the correlations found are causal, i.e., whether it is the case that L1 skills benefit L2 acquisition, or whether the correlation is a result of the same underlying, e.g., genetic, factors (cf. Colledge et al., 2002; Dale et al., 2010; Hayiou-Thomas et al., 2012). The stronger correlations between L1 and L2 skills in the native-language based group may indicate that for these participants, L2 vocabulary depends more on the L1, e.g., in that a larger Norwegian vocabulary facilitates cognate comprehension (see Unsworth et al., 2014, p. for a similar argument for English comprehension in Dutch children who had not attended English classes). This might be an indication that the native-language based group in the present project is still highly monolingual with little independent lexical learning in English. The lack of correlation between L1 and L2 vocabulary in the bilingually-based group may thus reflect their lexical learning from the L2 input. It was hypothesized in section 4.1 that this learning may still largely entail information about L2 form associated with the L1 translation equivalent. However, this would be different from the suggested explanation for the relationship between L1 and L2 vocabulary in the native-language group, since such development in the bilingually-based group would constitute the early stages of building an L2 lexicon, not simply the recognition of L2 words based on L1 cognates.

In Paper 1 it is argued that the lack of significant vocabulary development in the native-language based group in the course of eight months is an indication that there is not enough English in the environment of young Norwegian children to facilitate acquisition. This may be a slightly surprising finding to many Norwegians, but as discussed above, English media exposure is much more limited for young children than for adolescents and adults in Norway. Another question, of course, is whether such media exposure can be expected to facilitate acquisition in

[12] Robust confidence intervals were obtained by applying bootstrapping based on 1000 samples for all parametric tests in this section. However, for the sake of simplicity, only correlations and p-values are reported.

young children at all. In the background information, parents indicated whether the child had been systematically exposed to English through e.g., computer games, films, TV, and music prior to starting school, and for how many hours per week. Since not all parents clearly separated between input from the different sources, they were collapsed into one variable. Correlation analysis was run for this variable and initial English vocabulary in the two state school groups (bilingually-based and native-language based group) combined, to investigate the effect of prior input before the different learning contexts in grade 1 had had an effect. No significant correlation was found between this combined prior English input variable and pre-test English vocabulary, r=.191, p=.072. However, when hours of exposure to music, which was clearly specified by all parents, was removed from the variable, a moderate correlation was found with initial English vocabulary in both groups combined, r=.258, p=.023. This may be an indication that input from media does have an impact, but that the characteristics of language in music makes it less useful as a source of L2 input.

Out of the other control measures administered, very few were found to correlate with any of the three post-test measures of English. No such correlations were found in the immersion group; this is not surprising given the small number of children and the variable backgrounds and competence levels. In the native-language based group, a Pearson correlation was found for weeks spent outside of Scandinavia with sentence comprehension, r=.475, p=.005, and with sentence repetition, r=.417, p=.012. Exposure to other languages, particularly English, which is the lingua franca for most Norwegians outside of Scandinavia, may have facilitated some aspects of language processing, e.g., phoneme perception. In the bilingually-based group, no such correlation was found; it is not surprising that we find this effect only in the native-language based group, whose exposure to English in the classroom had been very limited. In the bilingually-based group, non-verbal intelligence, as measured by the Matrices section of the K-Bit 2, is correlated both with English receptive vocabulary in the post-test, i.e., the mean score on the PPVT-IV forms A and B, r=.523, p=.001. It thus seems that some aspect of non-verbal ability has facilitated lexical acquisition in particular from the input in this group; while this is not a traditional measure of language aptitude (cf. 2.1.2.3), we at least have evidence of individual differences also in very young FL learners. There is no corresponding correlation in the native-language based group, which may again be a reflection of the very limited learning that has taken place, and of the connection between the L1 and the L2 in this group.

The context of the four papers

The main purpose of the background measures was to ensure that the groups in the study were as closely matched as possible on factors which may potentially influence L2 learning. Thus, the measures were not administered primarily with the aim of looking for correlations with L2 development, and we therefore cannot draw firm conclusions based on the results reported in this section. However, they do indicate that SLA depends on a complex interplay between a number of individual factors, and also that the relative weightings of such factors may depend on context of acquisition, e.g., on volume of exposure.

5 Methods

In the following, the methods employed in the project are outlined, along with methodological challenges and choices. First, the selection of participants is discussed, and then the approach to classroom research itself. Then follow discussions of the background measures and the language tests, and finally of methods for analysis.

5.1 Participants and conditions

There are four groups of participants in the present project. The group referred to as the native-language based group consists of 29 children, 15 boys and 14 girls, attending a normal state school. They were exposed to standard English teaching as mandated by the Norwegian curriculum, i.e., as described for Norwegian English teaching in section 3.2.3. The bilingually-based group consists of 31 children, 17 boys and 14 girls, who attended a normal Norwegian state school but who were exposed to more English input than is normally the case in grade 1; this input was also more naturalistic, and thus better quality, than that of the native-language based group (see Papers 1 and 3 for details). The immersion group consists of seven children, three boys and four girls, who started grade 1 (or equivalent) in international schools in Norway where English is the language of instruction, with little or no knowledge of English. In this group, one child was monolingual, while the rest were (at least) bilingual, but they were reported by their parents not to speak English at the start of the year. Norwegian was, however, reported to be one of their main languages. Details for each child are given in Paper 4. Finally, the control group consists of 15 native speakers of English, eight boys and seven girls. In this group, 10 out of 15 children were reported by their parents to be bilingual, while five were reported to be monolingual speakers of English. Since they all lived in Norway at the time of testing and attended a school where Norwegian was taught as an L2, even these five control group participants did, however, have some exposure to and knowledge of a second language. In the native-speaker control group, only the English tests for which no norming exists were administered, i.e., the sentence comprehension and repetition tests reported in papers 2 – 4, and a yes/no question test to which we will return. Since acquisition was not studied in this group, they participated only in one test session.

The context of the four papers

The first challenge of the project was recruiting participants. It was necessary to find schools that would allow researchers to come into the school twice during the school year to test children. It was also necessary to have a quiet room where testing could take place, and teachers had to assist in obtaining informed written consent from parents. In the bilingually-based group, teachers were also required to make alterations to their teaching in order to expose the students to more English than would normally be the case. This proved to be difficult, and a number of schools were approached but not willing to participate. The school which did volunteer did so because one of the grade 1 teachers was a native speaker of English, and several other teachers also felt confident using the language in the classroom. However, the fact that one teacher was a native speaker was not ideal for methodological purposes since it may be argued to be a confounding factor. This is discussed in Papers 1 and 3.

Recruiting the native-language based group's school was easier, and the first school which was approached agreed to participate. This school was selected because it matched the bilingually-based group's school well in terms of socio-economic factors in the neighborhood from which it recruits its students, as well as on previous scores on national tests of English (see Papers 1 and 3 for details).

Recruiting participants for the third group, i.e., the immersion group, turned out to be very problematic. Norway has a number of international schools, but they generally have admission criteria for year 1 which favor children who already know English, e.g., native speakers or children who have previously attended Kindergarten at the same school. This means that the number of children who do not know English when starting such schools is low nation-wide. Furthermore, it was necessary that the children had good knowledge of Norwegian for them to have a language in common with the researchers to use during test sessions, and this excluded some children with international backgrounds. Over three years of recruitment for this group, only seven children with appropriate backgrounds were found whose parents consented to let their child participate.

Finally, recruiting children for the control group of native speakers was easier since it only required that English be (one of) their native language(s). However, the number of children attending international schools in Norway is relatively low, which is why this group has only 15 participants.

5.2 The approach to the classroom

An important methodological concern was to decide on the extent to which I, the researcher, would try to control exactly what took place in the classrooms attended by the participants in the study. In classroom SLA research, a multitude of variables outside of the researcher's control can complicate possible conclusions when two groups of students are compared. Many studies of the effectiveness of teaching strategies attempt to avoid this problem, i.e., the "global" problem of SLA research, by carefully controlling what happens in the classroom, focusing on a very specific teaching method or "treatment" and its value for language acquisition (see e.g., R. Ellis, 2012). In theory, such a strategy avoids the problem of not knowing exactly what in the treatment made the difference because it keeps the difference between the treatment condition and the control condition to a minimum. However, it creates new problems (Schachter, 1998). For example, in order to be sure that the difference between the two conditions is really minimal and consists of the treatment only, studies cannot be longitudinal. If the researcher is to be able to thoroughly control and analyze the teaching strategy, it has to take place within a few lessons; otherwise, in a real classroom setting, control is impossible. This is, for example, the case in most of the studies reviewed by Norris and Ortega (2000). Yet, we know that real acquisition does not take a few hours or even a few weeks; what such studies might discover is whether a very specific aspect of language can be acquired in a very specific limited time period given a very specific treatment. The knowledge acquired in this manner may be the least stable kind of knowledge, and it is not always clear that such tests are valid measures of L2 development (cf. Doughty, 2003; Lightbown, 1983). If we use delayed post-tests, we might be able to decide whether the effect is lasting, but we cannot then control all factors that may have influenced competence since the treatment, nor investigate more complex aspects of language that would need a longer process to develop. Carefully controlling all teaching for two groups of students over a long period of time, e.g., a year, is impossible.

Furthermore, even short experiments under carefully controlled conditions cannot be perfectly controlled for all possible confounding factors. We might use the same teacher and the same school for both conditions, and we might control for background factors in the two groups, but a multitude of other factors may still be different. For example, it might matter whether classes took place in the morning or the afternoon, or early or late in the week, and it might matter when they took place relative to another subject. It might make a difference whether

teaching took place in the students' home room, or the teacher's home room, and whether it took place in different rooms for the two groups. Furthermore, if one teacher is to teach in two different manners, chances are that one of the styles feels more comfortable to him. He may also have more confidence in one style, especially since it would be difficult not to let him know what the hypothesis of the experiment is. Such factors mean that the "minimal treatment difference" may be a deceptive notion which covers up real differences between conditions. Finally, with too careful control, studies may not be generalizable to other contexts. Cultural differences may play an important role in what strategies work in classrooms, but also more local factors such as class climate, teacher style and personality, and individual student preferences may decide whether a teaching method is effective, and such factors are hard to measure. In other words, finding that a certain treatment is effective for one teacher in one school does not necessarily mean that it would be effective for another teacher in another, or even the same, school (cf. R. Ellis, 2012).

Carefully controlling classroom practices is also an ethical question, since very carefully controlled teaching methods for experimental purposes might disrupt the natural working of the classroom. Furthermore, even if we do find a positive effect of the treatment, we only know what the treatment condition group learned over the test period that the control group did not. We do not, however, know what the treatment group did *not* learn; there is no guarantee that the total language acquisition of the treatment group is better than of the controls. This problem can to some extent be countered by including in the study both measures that should theoretically favor the treatment group and measures that should theoretically favor the non-treatment group, such as in the study of processing instruction by VanPatten and Cadierno (1993a). However, to investigate total language competence, approaches are needed that are more "global" (cf. R. Ellis, 2012), but thus less specific. For example, some such studies test the effect of an early starting age on a number of measures (e.g., García Mayo & García Lecumberri, 2003). Such studies give a better picture of overall language competence in the long run, but we know very little about what has been going on with the individual students from AoA until test age, since both teaching methods and all background factors are very hard to control. Also, with longitudinal, global approach studies, learners normally cannot be tested until after many years of acquisition, creating a long delay between e.g., changes in starting ages and the first reports of the effect.

Thus, any study of classroom SLA has the same inherent problems; there will always be a tradeoff between how controlled the experiment is and how comprehensive and generalizable the

measured outcome is. The most sensible way to handle the problems of local vs. global studies is probably to use a combination of both, including intermediate studies of a neither completely local nor completely global nature, where the advantages and disadvantages of both types of study are balanced. It is important to realize that SLA is a research field that deals with real people in real and varied circumstances, and that one or a few studies can never uncover absolute truths. A combination of a variety of studies using different methodologies in different contexts is likely to be the best way to uncover the factors which make up successful classroom SLA, both universal and particular to specific contexts.

In the present project, a fairly global approach was taken for the two state-school FL groups (i.e., native-language based and bilingually based), in that the study was relatively longitudinal, that classrooms were not carefully controlled through e.g., lesson analysis, and that tests did not investigate one specific item or subskill but rather sought to test overall competence. Classroom practices including information about time spent on English, activities, and materials were reported by teachers in the native language-based and the bilingually-based groups at various points during the year. During the test periods from late August to early October, and from late April to early June, these reports were frequent and informal, as I spent extensive time in each school and talked to the teachers on a regular basis. This also allowed for informal observation of the classrooms during the test periods. The only formal report took place in early March, when the project's contact teachers at both schools were asked about time, activities, and materials in an email.

This approach was deliberate since the desire was for both schools to operate as naturally as possible, and not to alter their practices to try to please me. Since observations and reports at various points were consistent, they can be assumed to be reliable. However, since careful control of classroom practice was abandoned, it is not possible to say exactly what made the difference in the two state school classrooms. We know that the bilingually-based group received more input in English and that this input was also more varied and more naturalistic. In this group, more time was formally spent on English classes and activities (with approximately 70 minutes per week vs. 45 minutes in the native-language based group), but more importantly, the teachers in the bilingually-based group spent most of this time actually speaking English, as opposed to the situation in the native-language based classroom where Norwegian was predominant. However, we do not know in detail what took place in each group. This may have been a problem if the

goal were to advise teachers to try to exactly replicate the bilingually-based classroom. However, in the present project, focus is not on what teachers should do, but on what the learners can achieve; *language learning*, not *language teaching* is discussed (cf. R. Ellis, 2012, p. 1). This is based on the perspective that foreign language learning is essentially a process of language acquisition. For the immersion group, control was even less strict, and the study of these seven children may be said to have been more descriptive than experimental (cf. R. Ellis, 2012, p. 42), since no manipulation was introduced compared to normal teaching in their school. Information about these classrooms was collected through descriptions in school curricula followed up by informal interviews with the teachers, e.g., to ascertain that no child had attended EAL classes without my knowledge. The study of progress in this group is even less generalizable than that of the bilingually-based group, but provides a picture of what competence may be gained in the course of one year of school in an immersion setting.

Thus, while the studies in the present thesis found that we can expect successful acquisition in young children provided enough input, the specific input provided to the bilingually-based class is not a recipe which can simply be copied and applied to other classrooms. Teachers in different contexts will have to work out for themselves how to best increase the input for their classes, and more research, possibly including more local and controlled studies, is needed to find out exactly what kind of strategies for input can be expected to be successful.

5.3 Tests

5.3.1 Considerations

There exist a number of standardized tests for language development in young children that are appropriate both for monolingual and for bilingual children. However, very early FL learning, especially in preliterate learners, is a fairly new phenomenon, and thus, the testing of such learners not a common practice. The usefulness of methods commonly used for more advanced bilingual children is limited with these FL learners because of low competence. For example, testing receptive grammatical skills through standardized tests would be difficult; a commonly used test for such language competence, the Test for Reception of Grammar, Version 2 (TROG-2, Bishop, 2003) is normed for native speakers from age 4. Since we know that the

receptive vocabulary skills of most children in the present study were below this age equivalent at the end of the test period, this test would not have yielded reliable results.

For tests of productive skills, on the other hand, it may be that individual factors not necessarily connected to language competence will affect how well the child performs; affective factors may already be relevant in such young FL learners. For example, some sort of "silent period" (cf. Clahsen & Felser, 2006b) may characterize early child SLA; while in immersion settings it has been suggested that such a period typically lasts for six months, it is not unlikely that young FL learners with limited input may be reluctant to speak for even longer. Willingness to communicate may thus be an individual difference in L2 classroom learners (cf. R. Ellis, 2012). It is also difficult to elicit production when comprehension is limited; for example, in the Expressive Vocabulary Test, Second Edition (EVT-2 K. T. Williams, 2007), vocabulary items are elicited through a stimulus question. As we have seen, there is no guarantee that this question would be comprehended by most children in this study if posed in English. If the prompt question is posed in Norwegian to elicit an English word, a range of other methodological problems would arise. Thus, most tests aimed at more proficient children were unsuitable for the FL learners in the two state schools.[13] Some of the tests mentioned above may have been useful for the seven children in the immersion group, but in order to be able to compare acquisition in the state school FL groups to that of the immersion group, the choice was made to administer exactly the same tests in the three groups of learners.

While tests designed for more proficient children were not appropriate for the state school FL groups, several factors make the use of more traditional FL tests equally unsuitable since they are usually intended for older learners. Such factors involve (lack of) literacy, motivational factors, and cognitive and maturational considerations. The first consideration, the fact that such young language learners may not be literate, is an obvious limitation on what types of tests may be used, and it means that much of the methodology common in SLA research cannot be used. Not only may the children be preliterate in their L1; L2 literacy skills may lag behind other L2 proficiency, and so any test involving reading or writing entails a risk that it is not language competence per se which is being tested. In the Norwegian curriculum, basic L1 literacy skills are

[13] See also Westergaard (2003) for a discussion of problems in testing very young English learners in Norway.

competence aims during the first two years of school. At the same stage, the curriculum has less ambitious competence aims for English literacy (Utdanningsdirektoratet, 2006, 2013a, 2013b).

A second consideration is cognitive development. As McKay (2006) points out, the attention span of young children is short and not suited to longer stretches of formal activities, be they instruction or testing, and young children tend to drop out of a task if they find it difficult. This means that in order to elicit the child's optimal performance, activities in test sessions must be varied, they must be motivating, and the focus should not be on "right" and "wrong" answers. Furthermore, the task and the instructions must be simple to make sure we are not simply measuring test-taking skills, and tests must be chosen carefully to keep test time short to avoid fatigue. It is also difficult to test many children at once, e.g., in a classroom, both because such mass testing is very difficult with non-written tests, and because children are easily distracted and may need someone with them at all times to make sure that they are carrying out the task correctly. Testing children one-on-one may also be advantageous for motivation since young children typically enjoy spending time alone with an adult, and since the tester can continually provide positive feedback. Encouragement and providing an overall feeling of success is very important not only ethically, but also to help the child perform to her full potential, increasing the reliability of the results.

When great variability is expected between children taking the same test, these considerations are even more challenging, in that tests must be selected that allow the high performers to show their full potential while at the same time bringing out the knowledge in children with lower proficiency, without leaving them with a feeling of failure.

In the papers in this thesis, three types of tests are used, namely tests of receptive vocabulary, sentence comprehension, and sentence repetition. The PPVT-IV tests receptive vocabulary in isolation, the sentence comprehension tests required comprehension of words in context, while the sentence repetition test required more detailed grammatical processing. A fourth test of English competence, namely a yes/no question test, was administered but found not to be valid; this will be further discussed in section 5.3.3.4. Most tests of English were thus designed specifically for the project. The only prefabricated tests used were the PPVT-IV (Dunn & Dunn, 2007a) for receptive vocabulary and the sentence comprehension test used in the pre-test session (Elllo, 2013a; 2013b; see also Appendix 1). All the tests are reviewed in section 5.3.3.

None of the tests in the project involved translation to or from the L1. This was a conscious choice since translation is a specific skill, and it is not obvious that having high competence in an L2 entails being good at translation (Grosjean, 2010). However, for all groups except the English native speaker control group, test administration took place in Norwegian since this was the only language in which it could be assumed that the children were competent enough to understand instructions.

As a natural consequence of the low expected proficiency level of the HG (bilingually-based group) and, especially, the LG (native-language based group), whose L2 lexicons could be expected to be very small, there is some overlap between vocabulary items in the different tests of English. This was necessary to avoid different difficulty levels in the vocabulary included in the tests, and to avoid that all tests simply became tests of lexical knowledge. However, since the tests were administered in the same order in all groups, with the PPVT-IV administered first (except in the control group), followed by the repetition test (in the post-test session only) and the comprehension test, any effect of encountering the same vocabulary more than once should also be the same in all groups.

Because of the children's low competence at the start of the project and the need to limit the duration of each test session, not all measures that were to go into the final analysis as dependent variables could be controlled at the start of the project. Thus, English competence could not be measured in terms of a repeated-measures study for all tests; this will be discussed in the next section. Time considerations also meant that for background measures, priority had to be given to a few measures and other potentially interesting background factors could not be checked. Among the factors not controlled, the most important is probably phonological short term memory. This problem is discussed briefly in section 5.3.4.4, and in detail in Paper 2. In the following, each test is presented, and its validity and reliability is evaluated.

5.3.2 Test sessions

Testing took place as early as possible in the fall for the pre-test, and as late as possible in the spring for the post-test. For most children, this meant testing in September and May, although a few children were tested in late August and April, or in early October and June. It was not possible to entirely control for time between sessions for each individual child, but the mean time between tests was the same in all groups, and sessions were balanced in that roughly every other day was spent in the bilingually-based and the native-language based groups to avoid a bias

toward early testing in one or the other group. In the immersion group, where early pre-testing was important because of the massive target-language exposure, pre-tests took place in the first week of September for all participants, with post-tests in the first week of May. As already mentioned, in the control group of English native speakers, each child only participated in one test session; this took place between November and January for each child.

Before the first test period started in the native-language based and the bilingually-based group, I spent a day at each school, informally taking part in various classes in order to make the children feel that they knew me when tests started. As test periods progressed, both I and the assistant who sometimes participated talked to all children as we frequented classrooms and corridors, and got to know the children fairly well. In the international schools, participation in classes in advance was not possible, but together with the the class teacher, I had a chat with each child before each test session started to make sure the child was comfortable with me before testing. No particular problems in establishing rapport with the children were experienced; in fact, many children asked to be tested several times. Test sessions lasted for approximately one hour per child, although this varied with the child's performance, notably in that higher performers were presented with more items on the vocabulary tests, and with individual factors such as talkativeness and response speed. If needed, children were given short breaks, but most children were able to complete each test session without interruption. The order of tests was designed to provide variation, and was the same for all children.

In the pre-test, Norwegian vocabulary (PPVT-IV form A) was tested first, followed by English vocabulary (PPVT form B), and the session was concluded with an English sentence comprehension test[14]. In the post-test session, the two forms of the PPVT-IV were administered first, both in English, followed by the two sections of the K-Bit 2, i.e., Riddles for (Norwegian) verbal ability and Matrices for non-verbal intelligence. Then followed the measure of visio-spatial working memory, the English yes/no question test, the English sentence repetition test, and finally, the English sentence comprehension test. The sentence comprehension tests were deliberately saved for last in both test sessions since they could be perceived as computer games, and thus provided motivation for the child throughout the test session. For the 15 participants in the English the native-speaker control group, the yes/no question test was administered first,

[14] The pre-test sentence comprehension test was not the same as the post-test sentence comprehension test. The difference will be described in detail in section 5.3.3.2

followed by the repetition test, the pre-test sentence comprehension test, and the post-test sentence comprehension test.

The schedule for testing was decided in agreement with class teachers to minimize disruption to the classroom. Tests times for individual children were not set up in advance; on test days, teachers continuously aided in selecting the next child to be tested. If a child refused to participate when asked, she was encouraged but not forced. If she chose not to participate at this time, she was asked again later. All children except one agreed to participate given this option. One boy in the native-language based group, however, opted out of the post-test, and was thus not included in the study.

Testing took place in a quiet room. Only I, the child, and sometimes the above-mentioned assistant were present. The assistant was a male master student of English linguistics at NTNU, and a native speaker of Norwegian. He participated in both pre- and post-test sessions whenever he was available, and administered non-English measures as well as the computerized tests. However, he never administered a test which required him to speak English since it was important that stimuli for all participants were as uniform as possible. Thus, the PPVT-IV, the repetition test, and the yes/no question test were always administered by me (a Norwegian L1 speaker and proficient L2 speaker of American English). There was no indication that the presence or absence of the assistant made a difference to children's performance since, like me, he spent extensive time at the schools during test periods so that the children felt that they knew him, and since when one of us was not administering a test, we kept quietly in the background and did not participate in the interaction.

5.3.3 Measures of English

5.3.3.1 The Peabody Picture Vocabulary Test

The only normed and standardized test of English competence in the project was the Peabody Picture Vocabulary Test, Fourth Edition (PPVT-IV, Dunn & Dunn, 2007a). The PPVT-IV is a well-tested and normed test for English vocabulary development which is suitable for children from the age of 2;6. It has been shown to have high validity and reliability in the reference group of American English speakers (Dunn & Dunn, 2007b). However, it only tests breadth of receptive vocabulary, a fact discussed in section 4.1.

The context of the four papers

Because the test was used in the studies in this thesis to test L2 rather than native language, it was administered in a manner slightly different from the guidelines for L1 testing. The recommended start ages were disregarded, and testing began with set 1 for all children on both forms of the test. On form A, administered in English only in the post-test, all children performed well enough to be "testable" according to the guidelines of the PPVT-IV because they made fewer than 8 mistakes in the first set of 12 words. In set B, however, three children in the pre-test session and four children in the post-test session made 8 or more errors, which means that they were technically "untestable" and only achieved chance scores. Their scores have still been included in the total results since they accurately represent very low scores. Interestingly, the children achieving such low scores were not the same in the pre- and the post-test sessions; all the low scorers in the pre-test were in the bilingually-based group, while in the post-test, all were in the native-language based group. Unless we assume that the latter four children had actually forgotten some English in the course of their first year of school, this may be an indication that some participants also made it past the first set by chance in the pre-test. However, it is at least clear that according to the guidelines of the PPVT-IV itself, competence in native-language based group was sometimes too low for testing after one year of school. This is an interesting finding which underlines the very limited impact that English classes have had in this group during grade 1, and which highlights that we cannot assume that a low starting age automatically equals acquisition.

The higher performance in the pre-test than the post-test for certain children in the native-language based group mentioned above points to a potential challenge of the PPVT-IV discontinuation rule. The rule is to stop testing when the child has made 8 or more errors in a given set, based on careful norming in an American context, where the subject is unlikely to know words from more difficult sets once she has made 8 errors. However, the progression might not be the same for Norwegian children if there are more cognates with Norwegian in a later set than in an earlier set. It is possible that the discontinuation rule may thus sometimes yield unfair results. After all, whether a child is moved on from one set to the next might come down to one word, i.e., whether she scores 5 rather than 4 on a given set and is thus allowed to continue. If the later sets are then easier for this particular group of participants, a child who by chance obtained a score of 5 rather than 4 might achieve a disproportionately high score compared to a participant who by chance obtained only a score of 4 and was not moved on. Results may thus not be as

reliable as they are when the test is used as intended, i.e., for English-speaking Americans. Still, since no standardized tests of English for Norwegian children at this age exists, it was judged to be the best choice for a comprehensive survey of English development[15]. It does, however, mean that it is possible that the results are not always entirely representative of each individual child's exact level.

These concerns were the reason why both forms A and B of the test were administered in the post-test session, i.e., that once the participant was discontinued on form A, the procedure was repeated with form B. Administering both forms thus meant providing twice as many words at low levels to each child, which gives a clearer picture of individual competence. This procedure also worked as a counterbalance to the discontinuation rule since children who scored unrealistically high or low on one form due to chance would presumably not do so also on the other form.

The results from the administration of both sets clearly illustrate that the PPVT-IV is not straightforwardly applicable to an L2 setting. The correlation between the two forms is expected to be high for the test's reference group of English speakers, e.g., r=.86 for ages 5-6 (Dunn & Dunn, 2007b, p. 56). For the subjects in the present study, however, this correlation was lower; Pearson's r=.614, p<.001 for all three L2 groups combined (i.e., immersion, bilingually based, and native-language based). The correlation is not the same in the three groups; r=.904, p=.003 in the immersion group, r=.405, p=.006 in the bilingually-based group, and r=.315, p=.048 in the native-language based group. A parametric test such as Pearson's r may not be reliable in the immersion group because of the small group size (n=7), but the same pattern is found for the three groups with non-parametric Spearman's ρ, with ρ =.716, p=.035 (immersion group), ρ =.361, p=.023 (bilingually based), and ρ =.343, p=.034 (native-language based). The higher correlation in the immersion group and the lower correlation in the native language-based group are indications that the correlation between the forms does not depend on language status (L1 vs. L2) as much as on language proficiency. Internal consistency was checked for the test in the context of the present study. For this test, as well as for all other tests in the present project with a discontinuation rule, reliability was tested using split-half (odd-even) reliability. For the PPVT-

[15] One alternative which was considered was the British Picture Vocabulary Scale (Dunn, Dunn, Whetton, & Burley, 1997), a version of the PPVT adapted for use in the UK. However, this would entail the same problem. Furthermore, although the UK is geographically closer to Norway than is the US, it is likely that the input that young children receive in English in various contexts is more often American than British, and so such a move would probably not have made the test any more reliable.

IV form A, the odd-even correlation was .944. For form B, the odd-even correlation was .914 in the pre-test and .966 for the post-test. These are all considered good correlations.

In the pre-test session, only form B of the PPVT-IV was administered in English. There are two reasons why form A was not used: Firstly, as already noted, an important consideration in test design was the generally limited attention and motivation spans of young children. Since it was expected that performance on this test would be very low, an expectation which was largely borne out, it was decided that administering two forms to establish that performance was low would cause more harm in the form of test fatigue than it would benefit the study. Secondly, as will be described in more detail in section 5.3.4.1, a Norwegian vocabulary test based on form A was used to test L1 vocabulary in the same session. Thus, form A could not be reliably used also to test English vocabulary in the pre-test.

Since only form B was distributed both in the pre- and the post-test, these are the scores reported in Paper 1, which discusses vocabulary development in a repeated-measures design in the native-language based and the bilingually-based based group. In Papers 2 and 3, evaluating the repetition test and comparing the native-language based and bilingually-based groups on all three English post-test measures, mean scores for the two forms of the PPVT-IV in the post test are reported. In Paper 4, the individual scores for each form are discussed for the immersion group.

5.3.3.2 The sentence comprehension tests

Both in the pre- and the post-test session, sentence comprehension tests were administered. However, the tests used were not the same in the two sessions. In the pre-test, an online test was used (Elllo, 2013a; 2013b, see also Appendix 1). This test consists of 12 sets of three pictures accompanied by linguistic stimuli, and each stimulus consists of several sentences, spoken by a male native speaker of American English. For each stimulus, the participant is asked to select the corresponding picture. Half of the sets included pictures of fruit, and the other half included pictures of pets. The fruit or the pet in the picture was described, but not named, in the linguistic stimuli. High performance in the English native-speaker control group indicated that this was not a problem in itself. Pilot testing found this test to perform relatively well, and test administration was successful in all four groups in the project. Most children perceived it as a computer game, and very few needed any encouragement to be able to complete the test. Approximate test time was 10 minutes for each child. However, after the pre-test sessions were

concluded, there was some concern about how fine-grained the results from this test would be when the children's competence had increased, with only 12 trials and only three alternatives per trial. Furthermore, the fact that stimuli consisted of more than one sentence provided the participants with many clues to the right answer, which made the interpretation of results difficult. Furthermore, internal consistency was found to be fairly poor, with Cronbach's α=.540. Thus, a new test was designed and administered in the post-test session.

In the sentence comprehension test for the post-test, there were 30 sets of four pictures, and linguistic stimuli consisted of one sentence per set of pictures (see Appendix 2). In each set of pictures, only one corresponded completely to the linguistic stimuli, while two other pictures contained material corresponding to some lexical elements from the sentence, and the last image was completely unrelated. For example, one stimulus sentence was "The horse is running." For this sentence, the picture set consisted of the target picture of a running horse, and one totally unrelated picture of a face. The remaining two pictures were a horse standing still and a person running, i.e., containing something corresponding to a part of the stimulus sentence. Similarly, for the question "Can you see the yellow flower," the target picture was of a yellow flower, the unrelated picture was of a gray cat, while the partially corresponding pictures contained a purple flower and a yellow rabbit, respectively. Thus, correct answers could not be provided based on isolated words only, and the child would minimally have to combine the understanding of several lexical items to select the appropriate picture.

Like the test used in the fall, the new sentence comprehension test was computerized; sentences were read alternately by two female native speakers of English, one speaker of British English and one of American English. Children were asked to click on the image which best corresponded to the sentence. If the subject preferred, he or she was given the option of pointing to the screen rather than operating the mouse, in which case the test administrator clicked on the appropriate image. Test time was again approximately 10 minutes, and the use of the computer worked as a motivating factor for most of the children. The child's answers for each sentence as well as the the total score were automatically recorded.

Two trial items were administered before the start of the test, to familiarize the child with the task and to make sure she understood that selecting an image based on one individual word might lead to the wrong result. Whether or not the child selected the appropriate images in the trials, there was a conversation where the test administrator explained how we "tried to trick" the

child by putting a red chair and a green monster to go along with the correct picture of a green chair for the sentence "The chair is green", and a dancing girl and a pig which was not dancing for the sentence "The pig is dancing". By the time the real test started, the child knew that it was important to listen to the entire sentence, and to look carefully at each picture, before making a decision. For the trial, both voices (the British and the American speaker) were introduced, so that the switch in speaker would not be confusing during the test. The test used basic vocabulary and simple sentences, and, as discussed in Paper 3, was probably mainly a test of segmentation and lexical retrieval, i.e., what S. E. Carroll (2004) calls "hearing words".

Reliability analysis using Cronbach's Alpha showed that internal consistency in the sentence comprehension test used in the post-test session was greatly improved from that in the pre-test session, and was in fact very good, α=.906. Because of the problems found with the pre-test sentence comprehension test described above, its results cannot be used to draw firm conclusions. In the present project, this test is mainly treated as a control measure which, together with the pre-test administration of the PPVT-IV form B, indicates that initial English comprehension in the two state school groups (i.e., native-language based and bilingually based) did not differ significantly. Still, sentence comprehension can be studied as a repeated measures design by dividing each participant's percentage of correct responses by the number of alternatives (3 and 4, respectively), which is done for the immersion group in Paper 4. However, the problems with the pre-test as well as possible ceiling effects in the post-test for sentence comprehension in this group prevents us from drawing any firm conclusions about their development on sentence comprehension.

5.3.3.3 The sentence repetition test

In the post-test session, a sentence repetition test consisting of 17 sentences[16] was administered. Such tests have been used in various types of language assessment, but there is some debate concerning exactly what they can be used to test and to what extent they are valid language measures. For this reason, a separate paper (Paper 2) in the present thesis was devoted to evaluating the use of this test. In this paper, the performance of all four groups of participants in the present project was analyzed. Since it was the English competence level at the time of the

[16] The slightly odd number of sentences was a result of findings in the pilot testing, where the test originally consisted of 15 sentences. Since pilot testing indicated that some children had problems with some of the shortest items intended as warm-ups, these were moved into the main test.

post-test, not the teaching to which they had been exposed, which mattered for the two state school groups (i.e., the native-language based and the bilingually-based group), the details of their school experience were not discussed in this paper. For this reason, they are given different labels in Paper 2 from the labels in papers 1 and 3 and in the rest of this chapter. The bilingually-based group is referred to as the high-input group (HG), while the native-language based group is referred to as the low-input group (LG). The third group in Paper 2 is the group discussed in Paper 4, namely the seven children who had started international schools with very little knowledge of English. They are referred to as the immersion group (IG). The fourth group is the group of 15 native speakers of English, i.e., the control group (CG). In the test, the participants repeated sentences read by me. Their responses were recoded and were scored by two independent raters based on correctly pronounced morphemes (see Appendix 3), and the test was at least expected to tap sensitivity to morphology. The conclusions of Paper 2 will not be repeated here, but certain issues not addressed in the paper will be discussed in the following.

For Paper 2 and the evaluation of this test, the unequal sizes of the four participant groups constitute a problem. For many of the evaluations, analyses of variance (ANOVAs) are the most appropriate method to look for differences in performance between groups. However, with group sizes varying from seven (immersion group) to 31 (bilingually-based group), the ANOVA may not be robust enough. Several steps were taken to check that the results reported in this paper can in fact be trusted. For all one-way ANOVAs, Levene's test of homogeneity of variance was applied, and when found significant, Welch's statistic for the F ratio is reported, as well as values not assuming equal variances for the planned contrasts. Furthermore, while the immersion and the bilingually-based group, i.e., the two groups most different in size, are in fact compared for all group-by-group comparisons, these are invariably the comparisons where differences in means are the largest. Thus, while unequal sample sizes may in theory yield too conservative or too liberal results, the differences between these two groups are so large that even if tests are not completely accurate, a significant difference found is highly likely to be real. Finally, all one-way ANOVAs were duplicated with non-parametric Kruskal-Wallis tests which found the same pattern of significant differences, although these are not reported in Paper 2. These results are summarized in the following.

Table 1 shows results for the relevant tests involving only three groups, namely the background measures and the post-test PPVT-IV, which were not administered in the control

group of native speakers of English. The point of the comparison of background measures is to make sure that the groups are not significantly different on these measures, since differences here could indicate that differences on the repetition test are not results of differences in English competence but of other factors. The comparison of post-test PPVT-IV results, on the other hand, is intended to establish that there are in fact significant differences between all groups in English competence, to ensure that they represent different levels of L2 proficiency at the time when the repetition test was administered.

Table 1: Results of Kruskal-Wallis tests for the means of all three L2 English groups (immersion, bilingually based, native-language based) on control measures and English vocabulary

	Non-verbal intelligence	Verbal ability	Visio-spatial memory	Norwegian vocabulary	English vocabulary (post-test)
Chi-Square	2.933	4.221	1.226	9.850	23.239
df	2	2	2	2	2
Asymp. Sig.	.231	.121	.542	.007	.000

While table 1 shows that there are no significant differences between the immersion group, the bilingually-based/high-input group (HG), and the native-language based/low input group (LG) for non-verbal intelligence, verbal ability, or visio-spacial memory, we see a significant difference for Norwegian vocabulary and an even larger difference for English vocabulary. This is exactly what the ANOVAs reported in Paper 2 found. Step-down follow-up analysis showed that between the immersion group and the two other groups there was a significant difference at the .05 level[17] for Norwegian vocabulary, while the bilingually-based and the native-language based group formed a homogenous subset with no significant difference, p=.276. This is exactly the pattern found in Paper 2.

Table 2 shows the Kruskal-Wallis test results for the measures tested in Paper 2 for all four groups, including the English native-speaker control group; these are the (post-test) sentence comprehension and sentence repetition tests. As with the PPVT-IV scores, these measures were compared to establish that the four groups did in fact represent four different English competence levels, and a significant difference is expected between all groups on both tests.

[17] The step-down follow-up procedure in the software used, IBM SPSS Statistics version 21, does not provide exact significance values for differences between groups that are found to differ significantly.

Table 2: Results of Kruskal-Wallis tests for the means of all four English groups (control, immersion, bilingually based, native-language based) on (post-test) sentence comprehension and repetition

	Sentence comprehension	Sentence repetition
Chi-Square	56.249	49.326
df	3	3
Asymp. Sig.	.000	.000

Again, the overall pattern follows that of the ANOVAs reported in Paper 2, in that there is a significant overall difference on both measures. Step-down follow-up analysis again found the same pairwise differences as those reported in the paper; it found the differences between the CG and the IG, between the IG and the HG, and between the HG and the LG, respectively, all to be significant at the .05 level for both sentence comprehension and repetition.

For all these comparisons, it thus seems that we can trust the results reported in Paper 2 concerning group differences. For all tests, ANOVAs are reported in the paper because they are more powerful and thus more appropriate to look for differences between the two state school groups (i.e., bilingually based/HG and native-language based/LG). They were in fact roughly equal in size (n=31, n=29, respectively), and large enough for parametric statistics to be appropriate.

Scoring of the repetition test by the two independent raters followed a scoring guide for awarding points per morpheme (see Appendix 3). Also for appropriate zero morphemes, such as no plural -s on a singular noun, one point was awarded. This is marked as Ø in the scoring guide. For inflected forms of pronouns and auxiliaries, one point was awarded for providing the appropriate word, and one for the appropriate form, so that *has* in an appropriate context would be awarded a total of two points. This is marked by *word+x* in the scoring guide. For the indefinite determiner (a/an), half a point was awarded for providing the determiner, and half a point for the appropriate form. This is marked with a slash (/) in the scoring guide. Both raters were thoroughly trained on the scoring procedure before rating responses on the test.

5.3.3.4 The yes/no question test

One test of English competence administered in the post-test session is not reported in any of the four papers because it was found to be problematic. This was a comprehension test with a delay condition. The test consisted of 14 simple yes/no questions, the answers to which could be

found by looking at a picture (see Appendix 4). However, the participant did not see the picture until three seconds after hearing the question, the idea being that this would force processing of the question in order to keep it in working memory.

The rationale for this test was the role of (phonological) working memory for spoken language comprehension, and for L2 in particular (e.g., Dufva & Voeten, 1999; N. Ellis, 2001; S. E. Gathercole & Baddeley, 1993; Hummel, 2009; Juffs & Harrington, 2011; Martin & Ellis, 2012; J. N. Williams, 2012). Memory resources taken up by processing may depend on automatization, and proficiency influences working memory tasks such as digit span in an L2 (Bley-Vroman & Chaudron, 1994; Olsthoorn, Andringa, & Hulstijn, 2012). It is thus possible that immediate comprehension and comprehension after a short delay are slightly different processes, in that the latter may need more memory resources, which depend on proficiency.

The questions were read by me. The children were told to answer "Yes" or "No"; they could respond in English or in Norwegian ("Ja" and "Nei") as they chose. Before the test there was a brief training session, where the child was presented with two sentences and, after a three second delay, with their accompanying images. The appropriate answer was "Yes" to the first practice question, and "No" to the second. The test was scored awarding one point for a correct answer and no points for a wrong answer. Test time per child was less than 5 minutes, including instructions and practice trials.

For this test, Cronbach's Alpha was only .553 for the group as a whole, which means that internal consistency in the test was rather poor. Other problems in the test were the level of difficulty, sensitivity, and bias for positive responses in some of the children. Even in the lowest competence (native-language based) group, there were subjects who obtained a perfect score on this test, which raises issues of ceiling effects. The means were high for all groups; out of a top possible score of 14, the means ranged from from 13.67 and 13.86 for the control and the immersion group, respectively, to 11.16 and 10.86 in the bilingually-based and native-language based groups respectively. The fact that the immersion group's mean is slightly higher than that of the native speaker control group may also indicate ceiling effects. The test thus discriminated between the two high-competence groups (control and immersion) on the one hand and the two state-school FL groups (bilingually-based and native-language based) on the other. However, this is not a very sensitive measure, and it is clear that competence differences were better discerned using the other measures in the study.

It is likely that more and more varied questions, including more complex ones, would have helped make this test more difficult and hence more sensitive. It seems that the three-second delay between the linguistic and the visual stimuli did not serve to increase difficulty level sufficiently to measure comprehension reliably. Furthermore, the fact that possible responses were only "Yes" or "No" created a 50% probability of selecting the correct response by chance. This further decreased the sensitivity of the measure.

Finally, another problem in the test is that it is possible that some children had a bias towards one response, usually "Yes"[18]. There were seven sentences in the test for which the correct response was *yes*, and seven for which it was *no*. While there was an approximately equal number of children who gave more "Yes" responses and those who gave more "No" responses, total erroneous "Yes" responses outnumbered erroneous "No" responses with 99 against 83; in other words, for those children who did give more "Yes" responses, the bias was probably stronger.

As many as 24 out of the 83 erroneous "No" responses were given to the same question, which indicates a specific problem with this item. This question was the relatively simple question 3, "Is the sun shining?", and errors occurred even in the control and immersion groups. This was surprising, but the explanation is probably to be found in the image accompanying the question. The sun in the picture was yellow and, to an adult eye, clearly a sun, and the sky was blue. However, the picture was relatively dark, and framed in black, which may have led the children to feel that it did not represent sunshine. One child in the native speaker control group pointed out this problem with the comment: "No, the sun is not shining – it's dark!". This highlights how children's interpretations may differ from those of adults, which is a factor we need to control for when creating tests for young children. The reliability issues discussed in the previous section make it clear that this was not a good measure of language competence, and thus, its results are not reported in any of the papers.

5.3.4 Control measures

Based on what we know about the relationship between L2 learning and other verbal and non-verbal abilities (see section 3.1.2.3), a number of control measures were used in the project. Their purpose was mainly to ensure that the two state-school FL groups (native-language based

[18] The bias for subjects, especially children, for "yes" responses is, for example, discussed by Crain and Thornton (1998).

and bilingually based) were as closely matched as possible on such factors, so that differences in L2 acquisition could safely be attributed to their learning contexts. However, the background measures were also used to look for relationships between background factors and SLA. For example, the effect of group membership (native-language based or bilingually based) was compared to the effect of L1 vocabulary and verbal ability in Paper 3, while correlation analysis was performed between a number of background measures and English development in the immersion group in Paper 4. Furthermore, the relationship between background factors and English measures was further explored through correlation analysis in section 4.3.

5.3.4.1 The Norwegian vocabulary test

In the pre-test session, a Norwegian test of receptive vocabulary based on the Peabody Picture Vocabulary Test form A was administered in order to check that L1 vocabulary development was comparable across groups. Since then, the British Picture Vocabulary Scale, second edition (BVPS) (Dunn et al., 1997) has been translated and normed for Norwegian (cf. Lyster, Horn, & Rygvold, 2010). However, this was not available when the testing took place in the present project. The test thus had to be developed specifically for this project.

The test used the vocabulary of the PPVT-IV form A as its starting point. However, direct translation of the English test was not always possible. English and Norwegian are are closely related languages, but their lexicons are relatively different, as discussed in section 3.2.1. This is a result of English being different from the other Germanic languages, in having been especially influenced by French and Latin. Norwegian, on the other hand, has mostly retained its original Germanic vocabulary. This creates a problem in particular for the translation of a test such as the PPVT-IV since it means that many words that are opaque and must be acquired independently in English are compounds in Norwegian, whose meaning can often be inferred from its components (see V. C. M. Gathercole, Thomas, & Hughes, 2008 for a similar problem with translation into Welsh). An example is the word *orchard*, which is *frukthage* (literally "fruit-garden") in Norwegian. This word appears in the starting set for the age range 14-16 in Form B, at a point where other words are clearly more advanced than the components *fruit* and *garden*. For this reason, words completely different from those of the English PPVT-IV sometimes had to be found to retain the appropriate level of the particular stage of the test.

Since this test was developed specifically for the present project, it was not normed or standardized. Thus, it can only be used for comparisons within the project, and vocabulary scores

in Norwegian and English cannot be directly compared. However, internal consistency was good, with an odd-even reliability of .872.

5.3.4.2 The verbal ability test

For verbal ability, the Riddles section of the Kaufmann Brief Intelligence Test, 2nd edition (K-Bit 2) was adapted to be used in Norwegian. The problems discussed for the adaption of the PPVT-IV are less relevant for this test, and more direct translation was possible. However, as with the Norwegian vocabulary test in this project, no norming is available for the verbal ability test, and scores can thus be compared within the project only. Odd-even reliability was checked, and was found to be relatively good, at .733.

The K-Bit 2 is normed for children from the age of four, and tests of verbal abilities similar to the Riddles section are standard parts of intelligence test batteries. However, tasks in this test entail accessing concepts on the basis of linguistic descriptions, which may be argued to be a complex cognitive process also involving metalinguistic aspects, which may be a competence which develops during school years. Thus, exactly what it tests in very young children is not entirely clear. However, a Pearson correlation with the Norwegian vocabulary test of .487, $p<.001$ indicates that this test does measure verbal ability, but is different enough from a receptive vocabulary test to yield relevant results on its own.

5.3.4.3 The non-verbal intelligence test

For non-verbal intelligence, the participants completed the Matrices section of (K-Bit 2). This section required no adaption except giving instructions in Norwegian rather than in English, given that the test itself is non-verbal. For this reason, there is no reason to assume that the reliability reported in the test manual (Kaufman, 2004) should not hold, and independent reliability analysis was not conducted within the project. The correlation between scores on the Riddles section and the Matrices section in the project was checked and found significant, $r=.264$, $p=.016$.

5.3.4.4 Test of working memory

As a measure of short-term memory, a memory game was used where picture cards were laid out in a pre-planned pattern, and the child was given a fixed amount of time to memorize the pattern before the cards were turned upside-down and the child was asked to find the pairs in as

few attempts as possible. The smallest set consisted of four cards, i.e., two pairs, and sets were incrementally expanded by one pair until the final set of 12 cards, or six pairs. The number of attempts the child used to find all pairs was recorded. The nature of this test does not allow for testing internal consistency since it does not consist of test items. However, the task was exactly the same at all levels and can therefore be assumed to have tested one underlying ability. This test can be assumed to have tested visio-spatial memory, e.g., the visio-spatial sketchpad in Baddeley & Hitch's (1974) model of working memory. It is a shortcoming of the present project that phonological working memory was not tested; this is discussed extensively in Paper 2.

5.3.4.5 Measures of social factors

As mentioned in section 3.2.2, there is ample access to English-language media in Norway, and even though most media for young children is in Norwegian, it was likely that some parents may have deliberately exposed their children to English in preparation for starting school. In a questionnaire at the beginning of the year (see Appendix 5), parents were asked about this, and also about whether they had been practicing English with the child. They were asked to indicate the approximate number of hours per week of such exposure. There were also questions about each child's international experience prior to starting school, on the assumption that for stays abroad, the child would have been exposed to at least some English as it is the normal lingua franca for tourism and travel.

This information was quantifiable, and it was first and foremost used to check that the two state school FL groups (native-language based and bilingually based) had not had significantly different levels of exposure. In section 4.3 the information was also used to look for relationships between outside exposure and English proficiency. However, it is important to note that these measures were crude and should not be used to draw firm conclusions.

5.3.5 Methods for analysis

5.3.5.1 Considerations

The first consideration to take into account when choosing methods for analysis was participant numbers in each group. In the two state school groups, the numbers of participants were 31 and 29 respectively. In the immersion group, there were seven participants. This low number is a result of the specific characteristics needed for inclusion in the group, i.e., no

knowledge of English, but Norwegian as an L1. For the control group, children needed to be native speakers of English but live in Norway. This group has 15 participants.

The state school FL groups are large enough for parametric statistic tests to be conducted. However, parametric tests rely on certain assumptions, such as a normal distribution of data. In large samples, this assumption tends to be met, but even the largest groups in the present project are small enough that normality may be a problem (Field, 2013, pp. 42, 156). There were issues with the distributions of data in the present project, with non-normal distributions for one or both groups on a number of measures, and this has affected how analyses were carried out.

The low number of children in the immersion group is in itself a problem for the use of parametric tests. For this reason, Paper 4 looks at these children's development individually in addition to statistical group analyses. However, this makes the results of this group less generalizable than those of the two state FL school groups. Furthermore, the fact that group sizes are so different makes comparison across groups difficult. In particular, this is an issue in Paper 2, which is the reason why analyses were checked with non-parametric tests as described in section 5.3.3.3.

5.3.5.2 Methods of analysis

The research questions in the present project were primarily related to differences in means between the groups of participants. However, secondary questions dealt with the relationship between background factors and L2 competence, or between aspects of L2 competence. Thus, a variety of statistical methods for analysis were employed, both methods for comparing means, such as t-tests and analyses of variance (ANOVA), and correlation and regression methods. Two methods were employed for overcoming problems with normality of data, namely non-parametric tests and bootstrapping. Non-parametric tests avoid the problems of low numbers and non-normality of data, usually by basing the analysis on ranked scores rather than on absolute values. Non-parametric tests are used in Paper 1 because of concerns related to the normality of data, in Paper 2 to check some of the findings of the parametric tests involving the small immersion group, and in Paper 4, which addresses the results of this group in particular.

However, although non-parametric tests are appropriate with small groups, their drawback is that they may be less powerful than their parametric counterparts, and that they are not appropriate for all experimental designs (Dancey & Reidy, 2002, p. 512; Field, 2013, pp. 214, 199). Therefore, another technique was used in Papers 2 and 3, namely parametric tests with

bootstrapping. Bootstrapping is a resampling technique where multiple samples are taken from the overall data sample, and their mean is calculated. This can be used to estimate robust standard errors and confidence intervals around the mean, estimated from the limits within which 95% of the sample means fall. Thus, we get a robust test of significance, and the analysis is reliable even when assumptions of normality are not met (Field, 2013, pp. 198 – 200). In Paper 3, both the t-tests for the differences between the native language-based and the bilingually-based groups and the regression analysis for the impact of L2 input thus have robust significance levels. In Paper 2, bootstrapping was used for confidence intervals and standard errors in t-tests and Pearson's correlation analyses. For ANOVAs, the software used, i.e., IBM SPSS Statistics version 21, only offers bootstrapping for contrasts and post-hoc tests, not for the main analysis, and it has problems performing this bootstrapping with very small sample sizes (cf. Field, 2013, pp. 465-466). For this reason, one-way ANOVAs in this paper were checked with non-parametric tests as reported in section 5.3.3.3. For reliability tests, i.e., Cronbach's Alpha and the Intraclass Correlation Coefficient, and for mixed designs ANOVAs, neither parametric alternatives nor bootstrapping options are available in the software.

5.3.5.3 Criteria and definitions of success

A possible methodological problem in the present project is to define criteria for what is to be considered successful acquisition. The perspective is that SLA is a process of becoming multilingual, and that L2 users are not failed native speakers (cf. Cook, 1999). This is in line with arguments concerning the consequences of the "bilingual turn", i.e., the realization that multilingualism is natural and normal (Ortega, 2014). However, the results of the L2 learners in this thesis are in various ways compared to the performance of native speakers. Yet, unlike in many discussions of native speaker comparisons (cf. Birdsong & Gertken, 2013; Ortega, 2014), no assumption is made in the present thesis about "native" necessarily equaling "monolingual". In Papers 1 and 4, age equivalents for the PPVT-IV are used; these are based on norms for children growing up in the US. The majority of a normal sample of American children will be monolingual speakers of English, but the test does not assume monolingualism in its reference group. Scores on the sentence comprehension and repetition tests are compared to the project's control group of English native speakers, where 10 out of 15 children were reported by their parents to be bilingual. The remaining five were reported to know English only, most having arrived in Norway shortly before starting school, but they all had at least some multilingual

experience in that they now lived in Norway, and that Norwegian as L2 was a subject in their schools.

The comparison to native speaker controls in the present project is not to be taken as an indication that the native speaker is seen as the ideal for successful L2 acquisition. The native-speaker comparisons are made for two reasons: First, in the absence of native speaker norms, it is not clear how one should measure L2 acquisition; we simply do not have any other objective criteria for language competence. Second, comparing L2 learners to native speakers is not the same as saying that any difference is a sign of failure. At least for the types of competence measured in the present project, i.e., receptive vocabulary, comprehension, and automaticity of language processing, it is possible to make comparisons to native speakers without unfairly biasing the conclusions towards native speaker standards. Crucially, the present studies are of very early language acquisition. These children are still developing, and the research questions do not ask about near-nativeness. We are not asking whether detailed representations are native-like; we are asking about the extent to which the learners approximate native speakers in broad areas involved in comprehension. In this sense, the native controls are to be seen as a yardstick, but not as an ideal towards which participants necessarily need to strive.

5.3.6 Limitations and suggestions for further research

A number of limitations to the present study have already been pointed out. One limitation is the number of participants. Not only is the immersion group too small to draw any firm conclusions; the fact that all findings for the bilingually-based and the native-language based groups are based on children from only two schools also makes generalizing the results difficult. For example, while it is assumed that the native-language based group represents the norm for Norwegian English teaching, in reality, diversity is obviously to be expected among different schools. There is no reason to assume that the school is not representative, but more groups of children would need to be studied for us to be certain.

Furthermore, the assumption is that the bilingually-based group differed from the native-language based group only in the amount of English input to which they were exposed, and that such an increase in input would yield similar results in other schools. Again, we do not know for certain that teachers in this school were not simply particularly effective in other ways which were not controlled. In particular, of course, the fact that one teacher was a native speaker of English may have been an advantage. While there is no theoretical reason to assume that children

cannot acquire language equally well from a fluent L2 speaker, again, more groups should be studied adopting a "bilingual" approach to establish whether the effect is generalizable. The formal English competence and proficiency of the other teachers in either school was not evaluated, and it may very well be that the lack of target-language use in the native-language based group was a result of low teacher proficiency. However, since low formal English competence among Norwegian grade 1 teachers is common, this would not make this school less representative. A natural recommendation based on the present findings is that teacher competence in English in the earliest grades in Norwegian schools needs to be improved to make sure the target language is used in the classroom.

There is also a limitation to the generalizability of the findings connected to the weak control over the classrooms in which English classes took place, as discussed above. This was a conscious choice in the project, and it may not be possible to study naturalistic acquisition by controlling classrooms more carefully. However, with studies of more classrooms adopting similar approaches, we may obtain more robust results. The short duration of the project is another reason why we cannot say very much about the exact benefits of the increased input in the bilingually-based group, and more longitudinal study of exposure may provide more insights.

Furthermore, the tests included in each test session could have been more comprehensive. In particular, testing phonological working memory would have benefited the evaluation of the repetition test, allowed us to be even more certain that the two state school FL groups were similar, and it might have shed more light on individual differences in these groups as well as in the immersion group. Measures of competence in all reported L1s in the immersion group would also have been useful for conclusions concerning what factors contribute to successful immersion SLA. In the immersion group, more advanced tests of English might have given more detailed results; in particular, the comprehension test may have been too easy for this group in the post-test. Furthermore, administering all L2 tests both in the pre- and the post-test session, and administering the same sentence comprehension test in both sessions would have allowed repeated-measure designs across the board, which would have yielded clearer results.

Also language pairs which are more different than Norwegian and English may yield interesting results. This is not least the case in the immersion group since their development is discussed with reference to other groups of children starting school in an L2, e.g., children with immigrant backgrounds, whose L1 may be more different from the language of instruction. Both

in the FL and the immersion contexts, studying languages which belong to different prosodic classes, e.g., stress vs. syllable-timed languages would be particularly interesting since there seems to be a benefit of increased exposure on word segmentation in particular, and since segmentation strategies may differ between prosodic classes. Further research in different contexts would also be useful, e.g., investigating whether the same benefits as this project has found can be found in FL contexts where the target language is less ubiquitous in the community, and where target language competence in teachers and parents can be assumed to be lower. In this connection, a particularly interesting question is whether the type of benefit found in the present project of increased input could also be found with input provided by media in the classroom.

6 Conclusion

The focus of the present project is second language acquisition in young children in school contexts, both in traditional foreign language and immersion settings. The project provides no answers to the question of the optimal starting age for a foreign language since no comparison is made to later starters. It does, however, contribute to our knowledge of what factors facilitate successful acquisition in the early language classroom.

One important finding is the clear indication that such early language learning, also in an FL classroom, is similar to other child SLA, in that volume and quality of input play a crucial role for acquisition, and that early foreign language learners benefit from implicit learning. For the Norwegian context, the findings show that English teaching within the national curriculum for grade 1 does not guarantee substantial acquisition, but that relatively modest adjustments to such teaching can have a significant effect. The project furthermore provides insight into the early language development of young foreign language learners, particularly in highlighting the benefits from target language exposure on processes such as phoneme perception and lexical retrieval.

For the immersion group, the project contributes to our knowledge about bilingual children attending school in their L2, by studying individual development in a group of students whose L2 proficiency when starting school was by definition very low. This part of the project highlights the danger of basing conclusions about bilinguals on generalizations since great individual variation was found even within this small group on children in the same context. It is worth noting, however, that rapid L2 development was found in all seven children.

Finally, the project contributes to the methodology in testing very young L2 learners, by establishing that tests such as the PPVT-IV can be validly used to test English competence even at very low levels, and also by administering and evaluating the tests developed especially for this project. This includes the yes/no question test, which did not provide useful data, and the sentence comprehension and sentence repetition tests, which did. Such test development is important to enable us to meaningfully test L2 development even in very young learners, and even at very early stages, to learn more about how young school-age children acquire new languages in different classroom contexts.

Conclusion

The four chapters which follow consist of the four separate papers summarized in section 2. First, Paper 1 discusses the vocabulary development of the two state-school groups (i.e., native-language based and bilingually based), concluding that in the former group, no significant development can be found, while in the latter group development is both significant and substantial. Next, Paper 2 evaluates the use of the sentence repetition test for all four groups. The paper concludes that the test is a valid measure of L2 competence even at the lowest competence level, and that in addition to sensitivity to morphology, it tests phoneme perception and lexical knowledge. Paper 3 compares the two state school groups (native-language based and bilingually based) on all three L2 measures in the post-test, i.e., receptive vocabulary (PPVT-IV forms A and B), sentence comprehension, and sentence repetition, and finds a significant difference on all three measures. The most substantial difference is found on sentence comprehension, and it is proposed that this reflects a particular benefit in "good enough" or "shallow" processing in the bilingually-based group. The effect of group membership on sentence comprehension is then compared to the effect of L1 receptive vocabulary and verbal ability, and found to make a much greater contribution, indicating that volume of input during the first year of school is essential for the development of such comprehension. Finally, Paper 4 investigates the individual English development of the seven children in the immersion group over their first year of school, finding very rapid development for all participants, but also great individual variation, and, predictably, general performance below the English native speaker control group. These results highlight both the need to treat children in such contexts as individuals, and the importance of distinguishing between different categories of bilingual children, given that these children are clearly still best characterized as L2 learners.

Appendix 1

Sentence comprehension test, pre-test

Images and transcripts from http://www.elllo.org/english/Games/G008-Fruit.html

and http://www.elllo.org/english/Games/G039-Pets.html

1.

OK, this fruit is big, it's round, it's pretty heavy. Usually it's green on the outside. It's hard on the outside, but very soft on the inside. You need to cut it - before you can eat it.

2.

OK, this fruit is round. It's usually hard on the outside - and hard on the inside. You sometimes cut it before you eat it, but - you don't have to cut it. You can just bite into it - <crunch> - and eat it naturally.

3.

OK, this fruit is soft on the outside and soft on the inside - kind of. And - you must peel it or cut it before you eat it. Usually people peel it with their hands before they eat it. It's round and it comes from a tree, and you can use it to make juice.

4.

OK, this fruit is pretty popular around the world. It's oblong - it's not round - it's oblong, and you must peel it before you eat it. You can cut it, but - you still must peel it a little bit before you can eat it. It's kind of soft on the outside, but very soft on the inside.

5.

OK, this fruit is small, and you can peel it, but most people don't. You can cut it but - most people don't. This fruit you pick, and throw in your mouth, and eat. It's very easy. Maybe the easiest fruit to eat of all fruits.

6.

OK, this last fruit is red - and small - and usually sweet. Ehm, its really good on ice cream, - and - ehm, you can eat it with cream, it's really good, some people like to put chocolate on it. And - if you ever have a milkshake, you can get this - fruit's flavor in the milkshake.

7.

Hi, I'm Conrad and - I'm gonna talk about pets today. Let's start with the first pet. This pet is such a lovable pet, it will wag its tail and lick you on your face when you come home!

8.

The next pet is not as big as a dog, but it's - bigger than a rat. In fact, it likes to chase rats. It's very furry.

9.

The next pet likes to swim. It swims around in a little bowl very fast. And it will eat, and eat and eat if you keep feeding it.

10.

This next pet is - a very - slow - animal. It walks and moves very slowly. It also has a very hard shell.

11.

And our last pet is a cute little critter. It's furry and feels nice and soft to touch. Every day it exercises in his cage.

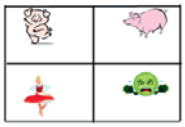
12.

And this pet - loves to fly. In fact, it wants to fly out of its cage - but it can't. It - makes - very - cheerful noises like 'tweet-tweet', 'chirp-chirp'.

Appendix 2

Sentence comprehension test, post-test

Trial 1: The chair is green



Trial 2: The pig is dancing

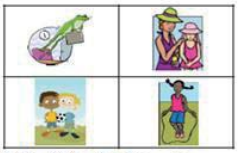
1. The girl is eating an apple

2. The boy is out in the rain

3. There are trees outside the house

4. The man is wearing a white shirt

5. This boy has a nice bird



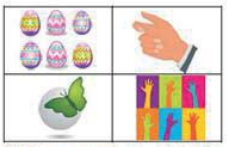
6. The little girl is jumping

7. The car in this picture is red

8. The people in this picture are happy

9. The boy is very sad

10. Where is the black shoe?

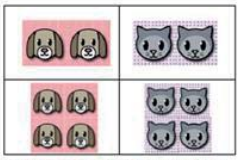


11. There are six hands in this picture

12. The girl really likes to sing

13. She has washed the dog

14. Daddy is reading to the boy



15. There are four cats in this picture

16. This funny rabbit has pink ears.

17. Can you see the yellow flower?

18. This elephant is playing with a ball.

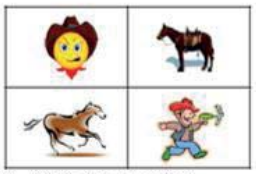
19. The teacher is very old.

20. The mother is feeding the baby.

21. The boy has drawn his mother a picture.

22. The cow jumped over the moon.

23. Three children are playing.



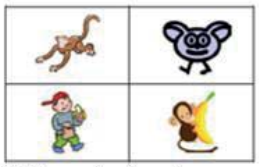
24. The horse is running.

25. She is holding the baby.

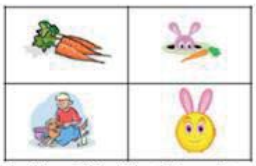
26. He is sleeping under a tree.

27. They are playing in the water.

28. The girl is holding the cat.



29. The monkey has a banana.



30. The rabbit is looking at the carrot.

Appendix 3

Sentence repetition test scoring guide

Trials

Pen
Fish
Eight
Shoulder
Pencil
Seven
I am a boy/girl.
You are three years old.
Is it raining outside?

Test sentences and scoring guide

Sentence	*Morphemes*	*Total*
1. Birds fly.	bird+s fly+Ø	4
2. Horses run fast.	Horse+s run+Ø fast	5
3. The baby drinks milk.	the baby+Ø drink+s milk+Ø	7
4. She is jumping.	she+x is+x jump+ing	6
5. He is eating a banana.	he+x is+x eat+ing a/ banana+Ø	9
6. She has washed the dog.	she+x has+x wash+ed the dog+Ø	9
7. They have played.	they+x have+x play+ed	6
8. He is not young.	he+x is+x not young	6
9. Cats do not fly.	cat+s do+Ø not fly+Ø	7
10. Is the elephant playing?	is+x the elephant+Ø play+ing	7
11. Does she sing?	do+es she+x sing+Ø	6
12. Do you like to read?	do+Ø you+x like+Ø to read+Ø	9
13. Are the children at school?	are+x the child+ren at school+Ø	8
14. The cat is playing with a ball.	the cat+Ø is+x play+ing with a/ ball+Ø	11
15. The children's dog always sleeps on the floor.	the child+ren+'s dog+Ø always sleep+s on the floor+Ø	13
16. The boy has been playing with his friends for an hour.	the boy+Ø has+x be+en play+ing with his+x friend+s for an/ hour+Ø	18
17. The house with the blue door looks nice.	the house+Ø with the blue door+Ø look+s nice	11
Total:		142

Ø = point for zero morpheme
x = 1 point for auxiliary/pronoun
/ = 0.5 point for indefinite article, 0.5 for correct form

Appendix 4

Yes/no question test

Trial 1: Is there a book in this picture?

Trial 2: Is it snowing?

Q1: Is this a baby?

Q2: Is there a dog in this picture?

Q3: Is the sun shining?

Q4: Is the man wearing a shirt?

Q5: Is this flower yellow?

Q6: Are the children eating?

Q7: Is this a foot?

Q8: Is the baby crying?

Q9: Is the strawberry red?

Q10: Is this a fish?

Q11: Is the mouse sleeping?

Q12: Is there a cat in the chair?

Q13: Has the boy hurt his knee?

Q14: Is the man in the house?

Appendix 5

i) Parent information and consent form, distributed to parents in native-language based, bilingually-based, and immersion groups.

Forespørsel om deltakelse i forskningsprosjekt

Hei alle foreldre i førsteklasse på xxx skole!

Skolen der dere har barn i første klasse hjelper meg i år med et doktorgradsprosjekt jeg holder på med på NTNU. Her undersøker jeg elevenes språkutvikling i engelsk i løpet av første skoleår. I den forbindelse ber jeg om deres tillatelse til å teste deres barn to ganger; en gang litt etter skolestart i høst, og en gang til våren, for å undersøke språkutviklingen i engelsk.

Med "test" menes altså ikke "prøve" i skoleforstand, men derimot en forskningsmessig test. Jeg bruker velutprøvde testmetoder som ungene syns er artige og ikke oppfatter som skremmende. Nå i høst testes ungenes ordforståelse både i norsk og engelsk, først og fremst ved at de hører ord og peker på bilder, og også ved at de svarer på noen enkle spørsmål etter en kort video på engelsk. Til våren vil testene ha større fokus på forståelse i engelsk, og også på andre faktorer som kan påvirke språklæring, som hukommelse og generelle problemløsningsevner, og vil fortsatt bestå av å peke på bilder og å svare på enkle oppgaver.

Under testene fokuserer vi ikke på riktige og gale svar overfor ungene, og testene er designet for å gi ungene en mestringsfølelse uavhengig av teknisk score. Under testene er bare det enkelte barnet og én eller noen ganger flere prosjektmedarbeidere til stede. Prosjektmedarbeiderne er meg selv, min veileder professor Mila Vulchanova, og muligens også en eller noen få masterstudenter fra NTNU. Det er viktig for den forskningsmessige kvaliteten at ungene føler seg trygge i testsituasjonen, og vi vil derfor sørge for at de kjenner minst en av de voksne til stede før testingen. Hver test vil ta omtrent en times tid, og foregå i skoletiden, eventuelt i SFO-tid når det passer.

Jeg er altså ute etter statistiske data, ikke etter å se hvor "flinke" de enkelte barna er. Resultatene av testingen vil bare håndteres av meg og mine medarbeidere, og vil være fullstendig anonymiserte før de brukes i offentligheten. I første omgang lagres alle resultatene med en personkode som tilsvarer hver elev på en atskilt navneliste slik at navn på eleven og resultater på testene ikke oppbevares på samme sted. Ved prosjektets slutt 1. august 2014 vil dataene anonymiseres fullstendig ved at elevenes navn og andre personopplysninger slettes helt.

Deltakelse i prosjektet er selvsagt frivillig, og hvis dere ikke svarer på denne forespørselen, vil deres barn ikke bli involvert i studien. Hvis dere har spørsmål eller ønsker mer informasjon om prosjektet, må dere svært gjerne kontakte meg. Godtar dere at barnet deres deltar i prosjektet, kan dere selvsagt likevel ombestemme dere og reservere dere når som helst. Dere behøver i så fall ikke gi noen begrunnelse for hvorfor dere ønsker å trekke dere.

Hvis dere velger alternativet "ja" under, vil dere i løpet av kort tid få et enkelt spørreskjema der jeg ber om litt mer informasjon om barnets bakgrunn som kan være relevant for språklæringen. Denne informasjonen vil behandles konfidensielt på lik linje med all annen personlig informasjon som kommer fram gjennom prosjektet. Skulle dere se at dere ikke ønsker å fylle ut dette skjemaet, kan dere la være å returnere det, og dere vil da regnes som å ha trukket dere fra prosjektet uten at dere trenger å foreta dere noe mer.

Med aller beste hilsen,

Anne Dahl,

Universitetslektor/PhD-student
NTNU, Institutt for moderne fremmedspråk
Tlf. 73596794, epost anne.j.dahl@ntnu.no

☐ Ja, jeg godtar at mitt barn deltar i forskningsprosjektet

_____	_____	_____
Barnets navn	Sted og dato	Foresattes underskrift

ii) Translation of parent information and consent form, distributed to parents in native-language based, bilingually-based, and immersion groups. English version not distributed to any group.

Dear first-grade parents at xxx School!

Your child's school is currently helping me with a PhD project conducted at NTNU. I am investigating the students' acquisition of English during their first year of school. For this, I hereby ask permission to test your child twice; once shortly after school starts this fall, and once in the spring, to investigate his/her development in English.

The word "test" does not mean examination in an academic sense, but rather a research test. I will be using well-tried methods which children find engaging and which will not make them anxious. This fall, the children's vocabulary comprehension is tested in Norwegian and English, first and foremost by them listening to words and pointing to pictures, but also by answering simple questions after a brief video in English. In the spring, the tests will have a stronger focus on English comprehension, and also on other factors which may influence language learning, such as memory and general problem solving skills, and they will still consist of pointing to pictures and solving simple tasks.

During tests, we will not focus on right and wrong answers with the children, and the tests are designed for them to have a sense of mastery independently of their technical score. During testing, only each individual child and one or sometimes several project participants will be present. Project participants are myself, my professor Mila Vulchanova, and possibly also one or a few master students from NTNU. It is important for research quality that the children feel safe during testing, and we will therefore make sure that they know at least one of the adults before testing. Each test will take approximately one hour and take place during school hours or sometimes during the after school program.

I am interested in statistical data, not in assessing the academic performance of individual children. The results of the tests will only be handled by me and my partners, and will be completely anonymized before they are used in public. Initially, all results will be saved with a personal code corresponding to each pupil in a separate list of names, so that names of pupils and results of tests will not be kept in the same place. At the end of the project on August 1, 2014, data will be completely anonymized by deleting pupils' names and other personal information.

Participation is of course voluntary, and if you do not respond to this request, your child will not be included in the study. If you have questions or want more information about the project, you are more than welcome to contact me. If you agree to let your child participate in the project, you may of course still change your mind and withdraw at any point. If you choose to do so, you do not need to give me any reason for withdrawing.

If you select the alternative "yes" below, you will shortly be given a simple questionnaire where I ask for more information about the child's background which may be relevant for language learning. This information will be handled confidentially just like all other personal information which I acquire through the project. If you decide that you do not wish to fill in this form, you simply do not return it, and I will then register that you have withdrawn from the project with no further action required from you.

Sincerely,

Anne Dahl,

University lecturer/PhD student
NTNU, Department of Modern Foreign Languages
Tlf. 73596794, anne.j.dahl@ntnu.no

☐ Yes, I will allow my child to participate in the study

The child's name	Place and date	Parent's signature

iii) Background information sheet, distributed to parents in native-language based, bilingually-based, and immersion groups.

Bakgrunnsinformasjon for forskningsprosjekt om engelsk i første klasse

Jeg er svært takknemlig for at dere har sagt ja til at deres barn kan delta i mitt språkforskningsprosjekt! I dette skjemaet ber jeg om litt bakgrunnsinformasjon som er nødvendig for at resultatene jeg får skal kunne brukes. Jeg håper dere vil bruke noen få minutter til å svare på disse spørsmålene.

Spørsmålene er ganske generelle, og hvis det noen steder er vanskelig å avgjøre hvilken svaralternativ som passer best, er det bare å forklare mer utfyllende i boksene under, eller på eget ark hvis nødvendig.

Legg merke til at skjemaet har to sider.

Med svært takknemlig hilsen,

Anne Dahl
Universitetslektor/PhD-student, NTNU

1. Barnets navn: _____ **Barnets fødselsdato**_____

2. Barnets språklige bakgrunn

☐ Barnet har ingen spesielle kunnskaper i andre språk enn norsk.

☐ Barnet snakker ___engl språk_____ bedre enn han/hun snakker norsk.

☐ Barnet snakker ___engl språk _____ omtrent like godt som han/hun snakker norsk.

☐ Barnet snakker ___engl språk _____, men han/hun snakker norsk bedre.

3. Barnets tidligere møter med engelsk og andre fremmedspråk

I skjemaet under, vennligst fyll inn hvilke land barnet har vært i utenom Norge i løpet av sitt liv, inkludert både korte og lengre opphold, omtrent hvor lenge barnet var der, og barnets alder da han/hun var der.

Hvis barnet ikke er født i Norge føres det også opp sammen med alder for når han/hun flyttet til Norge.

	Land barnet har besøkt:	Oppholdets varighet:	Barnets alder ved oppholdet:

4. Annen erfaring med engelsk

Har barnet jevnlig møtt engelsk på andre måter, f.eks. gjennom film, TV-programmer eller dataspill? ☐ Ja Nei ☐

Hvis ja, vennligst spesifiser hva slags aktivitet(er) det dreier seg om, og omtrentlig omfang (anslå f.eks. gjennomsnittlig antall timer i uka barnet har brukt på aktiviteten og fra hvilken alder dette har foregått).

5. Forberedelser til skolestart

Har barnet før skolestart på noen måte øvd eller forberedt seg spesielt på møtet med engelsk i skolen? ☐ Ja ☐ Nei

Hvis ja, vennligst spesifiser i her:

6. Barnets søsken (med søsken menes her andre barn som bor sammen med barnet i alle fall deler av tiden)

Antall brødre: ☐ Brødres alder: ☐ Antall søstre: ☐ Søstres alder: ☐

Dersom barnet har søsken han/hun ikke bor sammen med hele tiden, kan tid de bor sammen spesifiseres her:

7. Andre forhold som kan påvirke språklæringen

Har barnet, eller har det hatt, hørselsproblemer, alvorlige synsproblemer eller andre diagnoser som kan tenkes å påvirke språklæring (f.eks. spesifikke språkvansker, dysleksi, ADHD, autisme, osv.)? ☐ Ja ☐ Nei

Hvis ja, vennligst spesifiser i her:

Kryss av her hvis barnet har en tilstand eller diagnose dere ikke ønsker å beskrive på skjemaet, men som dere kan tenke dere å fortelle meg om personlig: ☐

8. Andre opplysinger

Er det andre opplysninger eller mer utfyllende informasjon som gjelder barnet ditt som du tror jeg kan ha nytte av å vite om i prosjektet mitt? Dette kan være utfyllende informasjon om barnets språkkunnskaper utover norsk, om barnets møter med engelsk, eller andre ting du tror kan påvirke resultatene av testene. Bruk gjerne eget ark om nødvendig.

Tusen takk ☺

iv) Translation of background information sheet, distributed to parents in native-language based, bilingually-based, and immersion groups. English version not distributed to any groups.

Background information for my research project about English in the first grade

I am very grateful that you have agreed to let your child participate in my language research project! In this form I ask for some background information necessary for me to be able to use the results. I hope you will spend a few minutes responding to these questions.

The questions are fairly general, and if in some places you find it difficult to decide which alternative is the best response, you are welcome to elaborate in the boxes provided, or on a separate sheet if necessary.

Please note that the form has two pages.

Very gratefully,

Anne Dahl
University lecturer/PhD student, NTNU

1. The child's name: _____The child's date of birth_____

2. Child's linguistic background

☐ The child has no particular knowledge of languages other than Norwegian.

☐ The child speaks ___language_____ better than he/she speaks Norwegian.

☐ The child speaks ___language_____ approximanely equally well as he/she speaks Norwegian.

☐ The child speaks ___language_____, but he/she speaks Norwegian better.

3. The child's previous encounters with English and other foreign languages

In the form below, please provide which countries the child has visited in addition to Norway in his/her life, including both short and longer stays, approximately how long the child stayed there, and the child's age when he/she was there.

If the child was not born in Norway, this information should be provided along with the age when he/she moved to Norway.

	Countries which the child has visited:	Duration of stay:	Child's age during stay:

4. Other experience with English

Has the child been exposed to English in other ways, e.g. through film, TV programs or computer games? ☐ Yes ☐ No

If yes, please specify which activitie(s) these were, and approximate volume (e.g. estimate average hours per week the child has been spending on the activity and from which age this has been taking place.

5. Preparation for starting school

Has the child been preparing before starting school for encountering in English in particular?

Yes ☐ No ☐

If yes, please specify below:

6. The child's siblings ('siblings' refers to children who live with the child at least part of the time)

Number of brothers: ☐ Brothers' ages: ☐ Number of sisters: ☐ Sisters' ages: ☐

If the child has siblings with whom he/she does not live all the time, the time when they live together can be specified here:

7. Other circumstances which may influence language learning

Does the child have, or has the child had, hearing problems, serious eyesight problems, or other diagnoses which may possibly influence language learning (e.g. specific language impairment, dyslexia, ADHD, autism, etc.)?? Yes ☐ No ☐

If yes, please specify below:

Please check this box if the child has a condition or diagnosis which you do not which to describe in this form, but which you are willing to tell me about in person: ☐

8. Other information

Is there any other information about your child which you think may be useful for me in my project? This may be information about the child's language knowledge in languages other than Norwegian, about the child's encounters with English, or other things you thing may infulence the results of the tests. You may use a separate sheet if needed.

Thank you ☺

v) Parental permission request and background information sheet distributed to parents in control group.

Research project permission request

Dear first-grade parents at xxx School!

xxx is currently helping me with a PhD project conducted at NTNU where I am investigating the English acquisition of Norwegian 6-year-olds. For the project, I need a control group of children for whom English is (one of) the main language(s), and I am hoping that you will allow me to include your child in this control group.

Participation entails allowing me to meet your child once during school hours this year to complete a simple set of research tests. These tests focus on English listening comprehension, and are not for assessment in an educational sense, meaning that there is no focus on right or wrong answers. Test methods have already been tried with a large number of children and they all enjoyed participating. During the test session only I and the child will be present, and the session will take approximately 30-60 minutes. Sessions will be organized in agreement with the children's teacher.

The results of the tests will be kept anonymous with a test subject code, and nobody but me will have access to a list of names linking the children to the results. At the end of my project in 2014, this list of names will be destroyed ensuring complete anonymity.

Participation is of course voluntary, and if you do not respond to this request, your child will not be included in the study. If you have any questions or concerns, please do not hesitate to contact me.

If you do agree to let me include your child in the study, I ask you to please answer the questions on the back of this sheet about the background of your child.

Sincerely,

Anne Dahl,

University lecturer/PhD student
NTNU, Department of Modern Foreign Languages
Tlf. 73596794, anne.j.dahl@ntnu.no

☐ Yes, I will allow my child to participate in the study

The child's name	Place and date	Parent's signature

Bacground information:

Some of the questions below may seem too general; please do not hesitate to use the boxes or even a separate sheet to provide additional information if necessary.

1. The child's name: _____ **The child's date of birth:** _____

2. The child's linguistic competence:

☐ The child mainly only speaks English

☐ The child speaks ___language_____ better than he/she speaks English

☐ The child speaks ___language _____ and English approximately equally well.

☐ The child speaks ___language _____, but less well than English.

3. The child's home language(s):_____

4. The parents' native language(s):_____

5. Stays abroad:

Countries outside of Norway where the child has spent time:	Duration of each stay abroad:	The child's age during stays abroad:

6. Would you consider English to be one of your child's main languages? ☐ **Yes** ☐ **No**

7. Does the child have difficulties with hearing or vision, or other diagnoses which may influence his/her linguistic development (e.g. specific language impairment, dyslexia, ADHD, autism, etc.) or has he/she had such difficulties in the past? ☐ **Yes** ☐ **No**

If yes, please specify below:

8. If you have any other information about your child which you think may be relevant to my study, please supply it below. (Use a separate sheet if necessary.)

Thank you for your help! ☺

References

Abello-Contesse, C., Chacón-Beltrán, R., López-Jiménez, M. D., & Torreblancaa-López, M. M. (2006). Introduction and overview. In C. Abello-Contesse, R. Chacón-Beltrán, M. D. López-Jiménez & M. M. Torreblancaa-López (Eds.), *Age in L2 acquisition and teaching* (pp. 7-27). Bern: Peter Lang.

Abildgaard, E., & Helland, T. (2011). Sammenligning av engelskkunnskaper hos to grupper 6.-7. klassinger undervist etter henholdsvis L97 og K06. *Acta Didactica Norge, 5*(1), 1-23.

Abrahamsson, N., & Hyltenstam, K. (2006). Inlärningsålder och uppfattad införddhet i andraspråket – lyssnarexperiment med avancerade L2-talare av svenska. *Nordand : nordisk tidsskrift for andrespråksforskning, 1*, 9-36.

Abrahamsson, N., & Hyltenstam, K. (2008). The robustness of aptitude effects in near-native second language acquisition. *Studies in Second Language Acquisition, 30*(4), 481-509. doi: 10.1017/S027226310808073X

Adams, A. M., & Gathercole, S. E. (1995). Phonological working memory and speech production in preschool children. *Journal of Speech Language and Hearing Research, 38*(2), 403-414. doi: 10.1044/jshr.3802.403

Adams, A. M., & Gathercole, S. E. (2000). Limitations in working memory: implications for language development. *International Journal of Language and Communication Disorders, 35*(1), 95-116.

Alexiou, T. (2009). Young learners' cognitive skills and their role in foreign language vocabulary learning. In M. Nikolov (Ed.), *Earlyl learning of modern foreign languages: Processes and outcomes* (pp. 46-61). Bristol: Multilingual Matters.

Álvarez, E. (2006). Rate and Route of Acquisition in EFL Narrative Development at Different Ages. In C. Muñoz (Ed.), *Age and the rate of foreign language learning* (Vol. 127-155). Clevedon: Multilingual Matters.

Aronin, L., & Singleton, D. (2012). *Multilingualism*. Amsterdam: John Benjamins.

Baddeley, A. D. (2003). Working memory and language: an overview. *Journal of Communication Disorders, 36*(3), 189-208. doi: 10.1016/S0021-9924(03)00019-4

Baddeley, A. D., Gathercole, S. E., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review, 105*(1), 158-173. doi: 10.1037/0033-295X.105.1.158

Baddeley, A. D., & Hitch, G. (1974). Working Memory. In G. H. Bower (Ed.), *Psychology of Learning and Motivation* (pp. 47-89). New York: Academic Press.

Behrens, B. (2008). Explaining advanced L2: Discourse properties of cordinating conjunction in English L1 and advanced L2. In W. Ramm & C. Fabricius-Hansen (Eds.), *Linearization and Segmentation in Discourse Multidisciplinary Approaches to Discourse 2008* (pp. 17-31). Oslo: University of Oslo.

Berulfsen, B., & Gundersen, D. (2000). *Fremmedord blå ordbok*. Oslo: Kunnskapsforl.

Bialystok, E. (2009). Bilingualism: The good, the bad, and the indifferent. *Bilingualism: Language and Cognition, 12*(1), 3-11. doi: 10.1017/S1366728908003477

Bialystok, E., & Barac, R. (2012). Emerging bilingualism: Dissociating advantages for metalinguistic awareness and executive control. *Cognition, 122*(1), 67-73. doi: 10.1016/j.cognition.2011.08.003

Bialystok, E., Craik, F. I. M., & Freedman, M. (2007). Bilingualism as a protection against the onset of symptoms of dementia. *Neuropsychologia, 45*(2), 459-464. doi: 10.1016/j.neuropsychologia.2006.10.009

Bialystok, E., & Hakuta, K. (1999). Confounded age: linguistic and cognitive factors in age differences for second language acquisition. In D. Birdsong (Ed.), *Second Language Acquisition and the Critical Period Hypothesis* (pp. 161-181). Mahwah, NJ: Lawrence Erlbaum Associates.

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. doi: 10.1017/S1366728909990423

Bialystok, E., McBride-Chang, C., & Luk, G. (2005). Bilingualism, language proficiency, and learning to read in two writing systems. *Journal of Educational Psychology, 97*(4), 580-590. doi: 10.1037/0022-0663.97.4.580

Bialystok, E., Peets, K. F., & Moreno, S. (2014). Producing bilinguals through immersion education: Development of metalinguistic awareness. *Applied Psycholinguistics, 35*(01), 177 - 191. doi: 10.1017/S0142716412000288

Birdsong , D. (1999). Introduction: Whys and why nots of the Critical Period Hypothesis for second language acquisition. In D. Birdsong (Ed.), *Second language acquisition and the Critical Period Hypothesis* (pp. 1-22). Mahwah, NJ: Lawrence Erlbaum.

Birdsong, D. (2006). Age and second language acquisition and processing: A selective overview. *Language Learning, 56*(S1), 9-49. doi: 10.1111/j.1467-9922.2006.00353.x

Birdsong , D., & Gertken, L. M. (2013). In faint praise of folly: A critical review of native/non-native speaker comparisons, with examples from native and bilingual processing of French complex syntax. *Language, Interaction and Acquisition, 4*(2), 107-133. doi: 10.1075/lia.4.2.01bir

Birketveit, A., & Rimmereide, H. E. (2012). Does reading stories enhance language learning? In A. Hasselgreen, I. Drew & B. Sørheim (Eds.), *The young language learner: Research-based insights into teaching and learning* (pp. 37-50). Bergen: Fagbokforlaget.

Bishop, D. (2003). *Test for Reception of Grammar, Version 2* London: The Psychological Corporation.

Bjørkeng, P. K. (2012). Barne-TV har skapt norsk dubbing-industri. *Aftenposten*. Retrieved from http://www.aftenposten.no/kultur/Barne-TV-har-skapt-norsk-dubbing-industri-6742235.html#.UvPl37StL1M

Blair, B., Moe, E., & Barsnes, B. (2011). Impact of Norwegian government policies on ESL classroom practices. In D. Tsagari & I. Csépes (Eds.), *Classroom-based language assessment* (pp. 111-128). Frankfurt am Main: Peter Lang.

Bley-Vroman, R. (1983). The comparative fallacy in interlanguage studies: The case of systematicity. *Language Learning, 33*, 1-17. doi: 10.1111/j.1467-1770.1983.tb00983.x

Bley-Vroman, R. (1988). The fundamental character of foreign language learning. In W. Rutherford & M. S. Smith (Eds.), *Grammar and second language teaching* (pp. 19-30). Rowley, MA: Newbury House.

Bley-Vroman, R. (1989). What is the logical problem of foreign language learning? In S. Gass & J. Schachter (Eds.), *Linguistic Perspectives on Second Language Acquisition* (pp. 41-68). Cambridge, MA: Cambridge University Press.

Bley-Vroman, R. (1990). The logical problem of foreign language learning. *Linguistic analysis, 20*(1-2), 3-49.

Bley-Vroman, R. (2009). The evolving context of the fundamental difference hypothesis. *Studies in Second Language Acquisition, 31*(2), 175-198. doi: 10.1017/S0272263109090275

Bley-Vroman, R., & Chaudron, C. (1994). Elicited imitation as a measure of second-language competence. In E. Tarone, S. M. Gass & A. D. Cohen (Eds.), *Research methodology in second-language acquisition* (pp. 245-261). Hillsdale, NJ: L. Erlbaum.

Bloomfield, L. (1933). *Language*. New York and Chicago: Holt, Rinehart and Winston.

Boas, A. (2013, 21. februar 2013). Dataspill kan gi bedre undervisning. *forskning.no*.

Bohnacker, U. (2006). When Swedes begin to learn German: From V2 to V2. *Second Language Research, 22*(4), 443-486. doi: 10.1191/0267658306sr275oa

Bonnet, G. (2004). The assessment of pupils' skills in English in eight European countries 2002: The European Network of Policy Makers for the Evaluation of Education Systems.

Bosch, L., & Sebastian-Galles, N. (1997). Native-language recognition abilities in 4-month-old infants from monolingual and bilingual environments. *Cognition, 65*, 33-69. doi: 10.1016/s0010-0277(97)00040-1

Bosch, L., & Sebastián-Gallés, N. (2001). Early language differentiation in bilingual infants. In J. Cenoz & F. Genesee (Eds.), *Trends in Bilingual Acquisition* (pp. 71-93). Philadelphia, PA: John Benjamins Publishing Company.

Breivik, L. E., & Hasselgren, A. (2002). *From the COLT's mouth - and others': language corpora studies in honour of Anna-Brita Stenström*. Amsterdam: Rodopi.

Brevik, L. M. M., Eli. (2012). Effects of CLIL teaching on language outcomes. In D. Tsagari & I. Csépes (Eds.), *Collaboration in language testing and assessment* (pp. 213-227). Frankfurt am Main: Peter Lang.

Broch-Utne, B. (2001). The growth of English for academic comminication in the Nordic countries. *International review of Education, 47*, 221-233. doi: 10.1023/A:1017941523866

Brubæk, S. (2012). Pragmatic competence in English at the VG1 level: To what extent are Norwegian EFL students able to adapt to contextual demands when making requests in English? *Acta Didactica Norge, 6*(1), 1-19.

Bruskeland, P. A., & Ranke, C. T. (2005). *Junior Scoop 1-2* (Vol. 1-2). Oslo: Samlaget.

Bruskeland, P. A., & Ranke, C. T. (2006). Junior Scoop 1-2 CD. *Junior Scoop*: Samlaget.

Burstall, C. (1975). French in the primary school: The British experiment. *Canadian Modern Language Review, 31*, 388-402.

Butler, Y. G. (2012). Bilingualism/multilingualism and second-language acquisition. In T. K. Bhatia & W. C. Ritchie (Eds.), *Blackwell Handbooks in Linguistics: Handbook of Bilingualism and Multilingualism* (2 ed., pp. 109-136). Somerset, NJ: Wiley.

Carroll, J. B. (1962). The prediction of success in intensive foreign language training. In R. Glaser (Ed.), *Training research and education* (pp. 87-136). Pittsburgh, PA: University of Pittsburgh Press.

Carroll, J. B. (1993). *Human cognitive abilities*. Cambridge: Cambridge University Press.

Carroll, S. E. (1999). Putting 'input' in its proper place. *Second Language Research, 15*(4), 337-388. doi: 10.1191/026765899674928444

Carroll, S. E. (2004). Segmentation: Learning how to 'hear words' in the L2 speech stream. *Transactions of the Philological Society, 102*(2), 227-254. doi: 10.1111/j.0079-1636.2004.00136.x

Cenoz, J. (2002). Age differences in foreign language learning. *ITL Review of Applied Linguistics*(135-136), 125-142.

Cenoz, J. (2003). The influence of age on the acquisition of English: General proficiency, attitudes and code-mixing. In M. García Mayo & M. García Lecumberri (Eds.), *Age and the acquisition of English as a foreign language* (pp. 77-93). Clevedon: Multilingual Matters.

Cenoz, J., & Genesee, F. (2001). First words. In J. Cenoz & F. Genesee (Eds.), *Trends in bilingual acquisition* (pp. 1-10). Amsterdam: J. Benjamins.

Charboneau, R. (2012). Approaches and practices relating to the teaching of EFL reading at the Norwegian primary level. In A. Hasselgreen, I. Drew & B. Sørheim (Eds.), *The young language learner: Research-based insights into teaching and learning* (pp. 51-70). Bergen: Fagbokforlaget.

Chechik, G., Meilijson, I., & Ruppin, E. (1997). Synaptic pruning in development: A computational account. *Neural Computation, 10*, 1759-1777.

Chvala, L. (2011). Examining oral skills in lower secondary school in Norway – a closer look at genre and situation. In M. Misztal & M. Trawiński (Eds.), *Annales universitatis paedagogicae Cracoviensis: Studia Anglica I* (pp. 80-90). Kraków: Wydawnictwo Naukowe UP.

Chvala, L., & Graedler, A.-L. (2010). Assessment in English. In S. Dobson & R. Engh (Eds.), *Vurdering for læring i alle fag* (pp. 75-89). Kristiansand: Høyskoleforlaget.

Clahsen, H. (1999). Lexical entries and rules of language: A multidisciplinary study of German inflection. *Behavioral and Brain Sciences, 22*(6), 991-1013; discussion 1014-1060.

Clahsen, H., & Felser, C. (2006a). Grammatical processing in language learners. *Applied Psycholinguistics, 27*(1), 3-42. doi: 10.1017/S0142716406060024

Clahsen, H., & Felser, C. (2006b). How native-like is non-native language processing? *Trends in Cognitive Sciences, 10*(12), 564-570. doi: 10.1016/j.tics.2006.10.002

Colledge, E., Bishop, D. V., Koeppen-Schomerus, G., Price, T. S., Happé, F. G., Eley, T. C., . . . Plomin, R. (2002). The structure of language abilities at 4 years: A twin study. *Developmental Psychology, 38*(5), 749-757. doi: 10.1037//0012-1649.38.5.749

Cook, V. (1991). The poverty-of-the-stimulus argument and multicompetence. *Second Language Research, 7*(2), 103-117. doi: 10.1177/026765839100700203

Cook, V. (1992). Evidence for multicompetence. *Language Learning, 42*(4), 557-591. doi: 10.1111/j.1467-1770.1992.tb01044.x

Cook, V. (1999). Going beyond the native speaker in language teaching. *TESOL Quarterly, 33*(2), 185-209. doi: 10.2307/3587717

Cook, V. (2008). *Second language learning and language teaching* (4 ed.). London: Hodder Education.

Cook, V., & Bassetti, B. (2011). Relating language and cognition: The second language user. In V. Cook & B. Bassetti (Eds.), *Language and bilingual cognition* (pp. 143-190). New York, NY: Psychology Press.

Corder, P. (1967). The significance of learners' errors. *International Review of Applied Linguistics, 5*, 161-169. doi: 10.1515/iral.1967.5.1-4.161

Crain, S., & Thornton, R. (1998). *Investigations in universal grammar: a guide to experiments on the acquisition of syntax and semantics*. Cambridge, MA: MIT Press.

Crystal, D. (2003). *A dictionary of linguistics and phonetics* (5 ed.). London: Blackwell.

Cummins, J. (1979). Linguistic interdependence and the educational development of bilingual children. *Review of Educational Research, 49*(2), 222-251. doi: 10.3102/00346543049002222

Cummins, J. (1998). Immersion education for the millennium: What we have learned from 30 years of research on second language immersion. In M. R. Childs & R. M. Bostwick (Eds.), *Learning through two languages: Research and practice. Second Katoh Gakuen International Symposium on Immersion and Bilingual Education* (pp. 34-47). Katoh Gakuen.

Dahl, A. (2004). Negative evidence in L2 acquisition. *Nordlyd*, *32*(1), 28-45. doi: 10.3389/fpsyg.2014.00329

Dahl, A., & Vulchanova, M. (2014). Naturalistic acquisition in an early language classroom. *Frontiers in Psychology. Special issue. Learning a non-native language in a naturalistic environment: Insights from behavioural and neuroimaging research.*, *5*, 1-9.

Dale, P. S., Harlaar, N., Haworth, C. M., & Plomin, R. (2010). Two by two: A twin study of second-language acquisition. *Psychological Science*, *21*(5), 635-640. doi: 10.1177/0956797610368060

Dancey, C. P., & Reidy, J. (2002). *Statistics without maths for psychology: using SPSS for Windows*. Harlow: Prentice Hall.

de Bot, K. (2006). The plastic bilingual brain: Synaptic pruning or growth? Commentary on Green, et al. *Language Learning*, *56*(S1), 127-132. doi: 10.1111/j.1467-9922.2006.00358.x

De Graaff, R. (1997). The EXperanto experiment: effects of explicit instruction on second language acquisition. *Studies in Second Language Acquisition*, *17*, 379-410.

de Groot, A. M. B. (1992). Bilingual lexical representation: A closer look at conceptual representations. In R. F. L. Katz (Ed.), *Orthography, phonology, morphology, and meaning* (pp. 389-412). Oxford: North-Holland.

de Groot, A. M. B. (1995). Determinants of bilingual lexicosemantic organisation. *Computer Assisted Language Learning*, *8*(2-3), 151-180. doi: 10.1080/0958822940080204

De Houwer, A., Bornstein, M. H., & Putnick, D. L. (2013). A bilingual–monolingual comparison of young children's vocabulary size: Evidence from comprehension and production. *Applied Psycholinguistics, FirstView*, 1-23. doi: 10.1017/S0142716412000744

DeKeyser, R. M. (2000). The robustness of Critical Period effects in second language acquisition. *Studies in Second Language Acquisition*, *22*(4), 499-533.

DeKeyser, R. M., & Larson-Hall, J. (2005). What does the critical period really mean? In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism: psycholinguistic approaches* (pp. 88-108). Oxford: Oxford University Press.

Djigunović, J. M. (2009). Impact of learning conditions on young FL learners' motivation. In M. Nikolov (Ed.), *Early learning of modern foreign languages: Processes and outcomes* (pp. 75-89). Bristol: Multilingual Matters.

Djigunović, J. M., & Vilke, M. (2000). Eight years after: Wishful thinking vs. the facts of life. In J. Moon & M. Nikolov (Eds.), *Research into teaching English to young learners* (pp. 67-86). Pécs: University Press Pécs.

Doughty, C. (2003). Instructed SLA: Constraints, compensation, and enhancement. In C. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 256-310). Malden, MA: Blackwell.

Doughty, C., & Williams, J. (1998). *Focus on form in classroom second language acquisition*. Cambridge: Cambridge University Press.

Drew, I. (2003). A comparison of early writing in Norwegian L1 and English L2. In K. Fraurud & K. Hyltenstam (Eds.), *Multilingualisn in global and local perspectives: Selected papers from the 8th Nordic conference on bilingualism, November 1-3, 2001, Stockholm* -

Rinkeby (pp. 341-355). Stockholm: Stockholm University/Rinkeby Institute of Multilingual Research.
Drew, I. (2004). Survey of English teaching in Norwegian primary schools. Stavanger: Stavanger University College.
Drew, I. (2008). Lesing og skriving: den australske måten. In T. M. L. Guldal, Ole F., G. Løkken, N. Naastad & F. Rønning (Eds.), *FoU i Praksis 2007: Rapport fra konferanse om praksisrettet FoU i lærerutdanning* (pp. 93-102). Trondheim: Tapir akademisk forlag.
Drew, I. (2009). Using the early years literacy programme in primary EFL Norwegian classrooms. In M. Nikolov (Ed.), *Early learning of modern foreign languages: Processes and outcomes* (pp. 108-120). Bristol: Multilingual Matters.
Drew, I. (2010). A longitudinal study of young language learners' development in written English. *Norsk Lingvistisk Tidsskrift, 28*, 193-226.
Drew, I. (2013). Using Content and Language Integrated Learning (CLIL) in the 9th grade: Teaching about the Second World War from a global to local perspective in English lessons. In I. Pareliussen, B. B. Moen & A. B. S. Reinertsen, Trond (Eds.), *FoU i praksis 2012 conference proceedings* (pp. 69-77). Trondheim: Akademika forlag.
Drew, I., & Hasselgreen, A. (2008). Young language learner (YLL) research: An overview of some international and national approaches *Acta Didactica Norge, 2*(1), 1-18.
Drew, I., Oostdam, R., & Toorenburg, H. v. (2007). Teachers' experiences and perceptions of primary EFL in Norway and the Netherlands: A comparative study. *European Journal of Teacher Education, 30*(3), 319-341.
Drew, I., & Pedersen, R. R. (2010). Readers Theatre: A different approach to English for struggling readers. *Acta Didactica Norge, 4*(1), 1-18.
Drew, I., & Pedersen, R. R. (2012). Readers Theatre: A group reading approach to texts in mainstream classes. In A. Hasselgreen, I. Drew & B. Sørheim (Eds.), *Understanding the language classroom: Research-based insights into language teaching and learning in primary and lower secondary school* (pp. 71-84). Bergen: Fagbokforlaget.
Drew, I., & Vigrestad, A. (2008). Choosing the path to excellence - Kunnskapsløftet i engelskfaget i Rogaland: Framgang, stagnasjon eller tilbakegang? Stavanger: Endelig rapport til Nasjonalt senter for fremmedspråk i opplæringen og Universitetet i Stavanger.
Dufva, M., & Voeten, M. J. M. (1999). Native language literacy and phonological memory as prerequisites for learning English as a foreign language. *Applied Psycholinguistics, 20*(3), 329-348.
Dunn, L. M., & Dunn, D. M. (2007a). Peabody Picture Vocabulary Test. Minneapolis, MN: Pearson Education, Inc.
Dunn, L. M., & Dunn, D. M. (2007b). *Peabody Picture Vocabulary Test Manual* (4 ed.). Minneapolis, MN: Pearson Education, Inc.
Dunn, L. M., Dunn, D. M., Whetton, C., & Burley, J. (1997). British Picture Vocabulary Scale. Nelson Publishing Company: London, UK.
Edelenbos, P., Johnstone, R., & Kubanek, A. (2006). *The main pedagogical principles underlying the teaching of languages to very young learners: Languages for the children of Europe. Published Research, Good Practice & Main Principles* Brussels: European Commission.
Edwards, E. M. T., Flognfeldt, M. E., Moen, E. S., Nilsen, M., & Møkleby, P. R. (2013). *Explore: engelsk for barnetrinnet*. Oslo: Gyldendal undervisning.
EF Education First. (2013). EF English Proficiency Index: EF Education First.

Eikrem, B. O. (2006). *Between 'filling up hard disks' and 'looking for talents and creativity': Teaching English as a foreign language in the 21st century: perceptions, attitudes, dilemmas, types of talk and educational needs.* (PhD thesis), University of Oslo, Oslo.

Eikrem, B. O. (2012). "Betre å vere stor enn liten". Korleis elevane opplever engelskfaget, sett i lys av arbeidsmåtane som vert nytta. In P. Haug (Ed.), *Kvalitet i opplæringa: Arbeid i grunnskulen observert og vurdert* (pp. 101-121). Oslo: Det norske samlaget.

Ellis, N. (2001). Memory for language. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 33-68). Cambridge: Cambridge University Press.

Ellis, N. C. (1995). Consciousness in second language acquisition: A review of field studies and laboratory experiments. *Language Awareness, 4*(3), 123-146. doi: 10.1080/09658416.1995.9959876

Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition, 24*(02), 143-188. doi: 10.1017/S0272263102002024

Ellis, N. C. (2006a). Language acquisition as rational contingency learning. *Applied Linguistics, 27*(1), 1-24. doi: 10.1093/applin/ami038

Ellis, N. C. (2006b). Selective attention and transfer phenomena in L2 acquisition: Contingency, cue competition, salience, interference, overshadowing, blocking, and perceptual learning. *Applied Linguistics, 27*(2), 164-194. doi: 10.1093/applin/aml015

Ellis, R. (2006). Current issues in the teaching of grammar: An SLA perspective. *TESOL Quarterly, 40*(1), 83-107. doi: 10.2307/40264512

Ellis, R. (2012). *Language teaching research and language pedagogy*. Chichester: Wiley-Blackwell.

Elllo. (2013a). Listening Games: Fruit. Retrieved 08.29.13, 2013, from http://www.elllo.org/english/Games/G008-Fruit.html

Elllo. (2013b). Listening Games: Pets. Retrieved 08.29.13, 2013, from http://www.elllo.org/english/Games/G039-Pets.html

Enever, J. (2009). Can today's early language learners in England become tomorrow's plurilingual European citizens? In M. Nikolov (Ed.), *Early learning of modern foreign languages: Processes and outcomes*. Bristol: Multilingual Matters.

Enever, J., Moon, J., & Raman, U. (Eds.). (2009). *Young learner English language policy and implementation: International perspectives*. Reading: Garnet Publishing.

Eurobarometer. (2005). Europeans and languages. Brussels: European Commission.

European Commission. (2009). European language policy and CLIL: A selection of EU–funded projects. Brussels: European Commission.

Eurydice. (2006). Content and Language Integrated Learning (CLIL) at School in Europe Directorate-General for Education and Culture, European Commission.

Eurydice, & Eurostat. (2012). Key data on teaching languages in Europe.

Faarlund, J. T., Vannebo, K. I., & Lie, S. (1997). *Norsk referansegrammatikk*. Oslo: Universitetsforl.

Felix, S. W. (1981). The effect of formal instruction on second language acquisition. *Language Learning, 31*(1), 87-112. doi: 10.1111/j.1467-1770.1981.tb01374.x

Felix, S. W. (1985). More evidence on competing cognitive systems. *Second Language Research, 1*(1), 47-72.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics: and sex and drugs and rock 'n' roll*. Los Angeles, CA: SAGE.

Flemmen, R. L. F. (2006). "Head and shoulders, knees and toes". In P. Haug (Ed.), *Begynnaropplæring og tilpassa undervisning - kva skjer i klasserommet?* Bergen: Caspar Forlag.

Flognfeldt, M. E. (2011). Emerging precision in lower secondary student writing in English. In M. Misztal & M. Trawiński (Eds.), *Annales universitatis paedagogicae Cracoviensis: Studia Anglica I* (pp. 101-108). Kraków: Wydawnictwo Naukowe UP.

Foyn, C., Vulchanova, M., Nilsen, R. A., & Sigmundsson, H. (under revision). Links between phonological memory, first language competence and second language competence in 10 year-old children. *Learning and Individual Differences*.

García Lecumberri, M. L., & Gallardo, F. (2003). English FL sounds in school learners of different ages. In M. d. P. García Mayo & M. L. García Lecumberri (Eds.), *Age and the acquisition of English as a foreign language* (pp. 115-135). Clevedon: Multilingual Matters.

García Mayo, M. d. P. (2003). Age, length of exposure and grammaticality judgements in the acquisition of English as a foreign language. In M. L. García Lecumberri & M. d. P. García Mayo (Eds.), *Age and the acquisition of English as a foreign language* (pp. 94-114). Clevedon: Multilingual Matters.

García Mayo, M. d. P., & García Lecumberri, M. L. (Eds.). (2003). *Age and the acquisition of English as a foreign language*. Clevedon: Multilingual Matters.

Garcia, R. L. (1974). Mexican American bilingualism and English language development. *Journal of Reading, 17*(6), 467-473.

Gass, S., Fleck, C., Leder, N., & Svetics, I. (1998). Ahistoricity revisited: Does SLA have a history? *Studies in Second Language Acquisition, 20*(3), 407-421.

Gathercole, S. E., & Baddeley, A. D. (1993). *Working memory and language*. Hove: Erlbaum.

Gathercole, S. E., Hitch, G., Service, E., & Martin, A. J. (1997). Phonological short-term memory and new word learning in children. *Developmental Psychology, 33*(6), 966-979.

Gathercole, S. E., Tiffany, C., Briscoe, J., Thorn, A., & team, T. A. (2005). Developmental consequences of poor phonological short-term memory function in childhood: A longitudinal study. *Journal of Child Psychology and Child Psychiatry, 46*(6), 598-611. doi: 10.1111/j.1469-7610.2004.00379.x

Gathercole, V. C. M., Thomas, E. M., & Hughes, E. (2008). Designing a normed receptive vocabulary test for bilingual populations: A model from Welsh. *International Journal of Bilingual Education and Bilingualism, 11*(6), 678-720. doi: 10.1080/13670050802149283

Genesee, F. (2004). What do we know about bilingual education for majority-language students? In T. K. Bhatia & W. C. Ritchie (Eds.), *The Handbook of Bilingualism* (pp. 547-576). Malden, MA: Blackwell Pub.

Gonzàlez Davies, M., & Taronna, A. (2013). *New trends in early foreign language learning: the age factor, CLIL and languages in contact : bridging research and good practices*. Newcastle: Cambridge Scholars.

Goorhuis-Brouwer, S., & de Bot, K. (2010). Impact of early English language teaching on L1 and L2 development in children in Dutch schools. *International Journal of Bilingualism*, 1-14. doi: 10.1177/1367006910367846

Gooskens, C. (2006). Which do young Swedes and Norwegians understand better, English or Danish? *Tidsskrift for Sprogforskning, 4*(1-2), 221-244.

Graedler, A.-L. (2002). Norwegian. In M. Görlach (Ed.), *English in Europe* (pp. 57-81). Oxford: Oxford University Press.

Grosjean, F. (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language*(36), 3-15. doi: 10.1016/0093-934X(89)90048-5
Grosjean, F. (2008). *Studying bilinguals*. Oxford: Oxford University Press.
Grosjean, F. (2010). *Bilingual: Life and reality*. Cambridge, MA: Harvard University Press.
Hansen, C. R., Mørner, I. O., Næss, T. A. B., & Pritchard, P. (2012). Quest 1-4.
Harley, B., & Hart, D. (1997). Language aptitude and second language proficiency in classroom learners of different starting ages. *Studies in Second Language Acquisition, 19*(03), 379-400.
Harley, B., & Hart, D. (2002). Age, aptitude, and second language learning on a bilingual exchange. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 302-330). Amsterdam: Benjamins.
Harrington, M. (2001). Sentence processing. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 91-124). Cambridge: Cambridge University Press.
Hasselgreen, A. (1994). Lexical teddy bears and advanced learners: A study into the ways Norwegian students cope with English vocabulary. *International Journal of Applied Linguistics, 4*(2), 237-258. doi: 10.1111/j.1473-4192.1994.tb00065.x
Hasselgreen, A. (2000). The assessment of the English ability of young learners in Norwegian schools: An innovative approach. *Language Testing, 17*, 261-277.
Hasselgreen, A. (2002). Sounds a bit foreign. In L. E. Breivik & A. Hasselgreen (Eds.), *From the COLT's mouth - and others': Language corpora studies in honour of Anna-Brita Stenström* (pp. 103-123).
Hasselgreen, A. (2003). What learners (know they) can do: A research-based approach to portfolio assessment in the foreign/second language classroom. *Acta Didactica*(1), 54-67.
Hasselgreen, A. (2004). *Testing the spoken English of young Norwegians: a study of test validity and the role of "smallwords" in contributing to pupils' fluency*. Cambridge: Cambridge University Press.
Hasselgreen, A. (2005). Assessing the language of young learners. *Language Testing, 22*(3), 337-354. doi: http://dx.doi.org/10.1191/0265532205lt312oa
Hasselgreen, A. (2012). Developing a framework for assessing writing in primary school. In A. Hasselgreen, I. Drew & B. Sørheim (Eds.), *The young language learner: Research-based insights into teaching and learning* (pp. 221-232). Bergen: Fagbokforlaget.
Hasselgreen, A., Drew, I., & Sørheim, B. (2012). *The young language learner: Research-based insights into teaching and learning*. Bergen: Fagbokforlaget.
Hasselgård, H. (2009a). Temporal and spatial structuring in English and Norwegian student essays. In R. Bowen, M. Mobärg & S. Ohlander (Eds.), *Corpora and discourse - and stuff. Papers in honour of Karin Aijmer* (pp. 93-104). Gothenburg: Acta Universitatis Gothoburgensis.
Hasselgård, H. (2009b). Thematic choice and expressions of stance in English argumentative texts by Norwegian learners. In K. Aijmer (Ed.), *Corpora and language teaching* (pp. 121-139). Amsterdam: John Benjamins Pub. Co.
Hasselgård, H. (2012). Facts, ideas, questions, problems, and issues in advanced learners' English. *Nordic Journal of English Studies, 11*(1), 22-54.
Haugen, E. (1969). *The Norwegian language in America: A study in bilingual behavior*. Bloomington, IN: Indiana University Press.
Hayiou-Thomas, M. E., Dale, P. S., & Plomin, R. (2012). The etiology of variation in language skills changes with development: A longitudinal twin study of language from 2 to 12 years. *Developmental Science, 15*(2), 233-249. doi: 10.1111/j.1467-7687.2011.01119.x

Helland, T. (2008). Second language assessment in dyslexia: Principles and practice. In J. Kormos & E. H. Kontra (Eds.), *Language learners with special needs: An international perspective* (pp. 63-85). Clevedon: Channel View Publications.

Helland, T., & Kaasa, R. (2005). Dyslexia in English as a second language. *Dyslexia, 11*(1), 41-60. doi: 10.1002/dys.286

Hellekjær, G. O. (1994). Developing reading skills in the bilingual classroom: A pilot project. *Språk og Språkundervisning, 27*(3), 32-47.

Hellekjær, G. O. (2005). *The acid test: Does upper secondary EFL instruction effectively prepare Norwegian students for the reading of English textbooks at colleges and universities?* (PhD thesis), University of Oslo, Oslo.

Hellekjær, G. O. (2007). Fremmedspråk i norsk næringsliv - engelsk er ikke nok! Halden: Fremmedspråksenteret.

Hellekjær, G. O. (2008). A case for improved reading instruction for academic English reading proficiency. *Acta Didactica Norge, 2*(1), 1-17.

Hellekjær, G. O. (2009). Academic English reading proficiency at the university level: A Norwegian case study. *Reading in a Foreign Language, 21*(2), 198-222.

Hellekjær, G. O. (2010). Lecture comprehension in English-medium higher education. *Hermes, 45*, 11-34.

Hellekjær, G. O. (2012). A survey of English use and needs in Norwegian export firms. *Hermes, 48*, 7-18.

Helness, H. L. (2012). A comparative study of the vocabulary of 7th and 10th graders in scripts from the National Tests in English. In A. Hasselgreen, I. Drew & B. Sørheim (Eds.), *The young language learner: Research-based insights into teaching and learning* (pp. 145-158). Bergen: Fagbokforlaget.

Hestetræet, T. (2012). Teacher cognition and the teaching and learning of EFL vocabulary. In A. Hasselgreen, I. Drew & B. Sørheim (Eds.), *The young language learner: Research-based insights into teaching and learning* (pp. 177-190). Bergen: Fagbokforlaget.

Hestnes, H. (2006). Internasjonalisering i lærerutdanningen: Hvordan opplever norske elever og lærere å ha en engelsktalende lærerstudent i klasserommet? *Norsk Pedagogisk Tidskrift*(3), 218-229.

Hoad, T. F. (1996). *The concise Oxford dictionary of English etymology*. Oxford: Oxford University Press.

Holmstrand, L. S. E. (1982). *English in the elementary school: Theoretical and empirical aspects of the early teaching of English as a foreign language*. Uppsala: Acta Universitatis Upsaliensis.

Hulstijn, J., & De Graaff, R. (1994). Under what conditions does explicit knowledge of a second language facilitate the acquisition of implicit knowledge? A research proposal. *AILA Review, 11*(97-113).

Hummel, K. M. (2009). Aptitude, phonological memory, and second language proficiency in nonnovice adult learners. *Applied Psycholinguistics, 30*(02), 225-249. doi: 10.1017/S0142716409090109

Hyltenstam, K. (1992). Non-native features of near-native speakers: On the ultimate attainment of childhood L2 learners. In R. J. Harris (Ed.), *Cognitive Processing in Bilinguals* (pp. 351-368). Amsterdam: Elsevier Science.

Hyltenstam, K., & Abrahamsson, N. (2003). Maturational constraints in SLA. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 539-588). Oxford: Blackwell.

Håkenstad, H., & Vestgård, M. U. (2006). *Stairs 1+2*. Oslo: Cappelen.
Ibsen, E. (2004). Engelsk i Europa - 2002 : Norsk rapport fra en europeisk engelskundersøkelse om holdninger til og ferdigheter i engelsk ved utgangen obligatoriske grunnskolen. *Acta Didactica*, 1-124.
Ibsen, E., & Hellekjær, G. O. (2003). A profile of Norwegian teachers of English in the 10th grade. *Acta Didactica*, 68-86.
Iverson, P., & Evans, B. G. (2007). Learning English vowels with different first-language vowel systems: Perception of formant targets, formant movement, and duration. *The Journal of the Acoustical Society of America*, *122*(5), 2842-2854. doi: 10.1121/1.2783198
Jalkanen, J. (2009). Language proficiency and age-appropriate methodology at the English Kindergarten of Kuopio. In M. Nikolov (Ed.), *The age factor and early language learning* (pp. 97-118). Berlin: Mouton de Gruyter.
Jiang, N. (2000). Lexical representation and development in a second language. *Applied Linguistics, 21*(1), 47-77. doi: 10.1093/applin/21.1.47
Jiang, N. (2002). Form-meaning mapping in vocabulary acquisition in a second language. *Studies in Second Language Acquisition, 24*(4), 617-637. doi: 10.1017.S0272263102004047
Jiang, N. (2004). Semantic transfer and its implications for vocabulary teaching in a second language. *The Modern Language Journal, 88*(3), 416-432. doi: 10.1111/j.0026-7902.2004.00238.x
Johansson, S., & Graedler, A.-L. (2002). *Rocka, hipt og snacksy: Om engelsk i norsk språk og samfunn*. Kristiansand: Høyskoleforlaget.
Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology, 21*(1), 60-99.
Juffs, A., & Harrington, M. (2011). Aspects of working memory in L2 learning. *Language Teaching, 44*(02), 137-166. doi: 10.1017/S0261444810000509
Kachru, B. (1985). Standards, codification and sociolinguistic realism: The English language in the outer circle. In R. Quirk & H. G. Widdowson (Eds.), *English in the world: Teaching and learning the language and literatures* (pp. 11-30). Cambridge: Cambridge University Press.
Kachru, B. (1992). Models for non-native Englishes. In B. Kachru (Ed.), *The other tongue: English across cultures* (pp. 48-74). Urbana, IL: University Of Illinois Press.
Kaufman, A. S. K., Nadeen L. (2004). *Kaufman Brief Intelligence Test, second edition manual*. Minneapolis, MN: Education, Inc.
Kay, S. (2014). *The multilingual turn: Implications for SLA, TESOL and bilingual education*. New York, NY: Routledge.
Klein, W. (1998). The contribution of second language acquisition research. *Language Learning, 48*(4), 527-549. doi: 10.1111/0023-8333.00057
Krashen, S. D. (1981a). Aptitude and attitude in relation to second language acquisition and learning. In K. C. Diller (Ed.), *Individual differences and universals in language learning aptitude* (pp. 155-175). Rowley, MA: Newbury House.
Krashen, S. D. (1981b). *Second language acquisition and second language learning*. Oxford: Pergamon Press.
Krashen, S. D. (1985). *The Input Hypothesis: Issues and implications*. New York, NY: Longman.
Krashen, S. D., Long, M. H., & Scarcella, R. C. (1979). Age, rate and eventual attainment in second language acquisition. *TESOL Quarterly, 13*(4), 573-582.

Krashen, S. D., Long, M. H., & Scarcella, R. C. (1982). *Child-adult differences in second language acquisition*. Rowley, MA: Newbury House.

Krave, M. F. (2012). The English -ing participial adjuncts in first and second language written discourse. *Discourse and Interaction, 5*(2), 17-34. doi: 10.5817/DI2012-2-17

Kroll, J. F., Gullifer, J. W., & Rossi, E. (2013). The multilingual lexicon: The cognitive and neural basis of lexical comprehension and production in two or more languages. *Annual Review of Applied Linguistics, 33*, 102-127. doi: 10.1017/S0267190513000111

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language, 33*(2), 149-174. doi: 10.1006/jmla.1994.1008

Kroll, J. F., & Tokowicz, N. (2001). The development of conceptual representation for words in a second language. In J. Nicol (Ed.), *One mind, two languages: Bilingual language processing* (pp. 50-71). Malden, MA: Blackwell.

Kroll, J. F., Van Hell, J. G., Tokowicz, N., & Green, D. W. (2010). The Revised Hierarchical Model: A critical review and assessment. *Bilingualism: Language and Cognition, 13*(3), 373-381.

KUF. (1997). *Læreplanverket for den 10-årige grunnskolen*. Oslo: Ministry of Church, Education, and Research; Nasjonalt læremiddelsenter.

Kunnskapsdepartementet. (2006). *Forskrift til opplæringslova*. Retrieved from http://lovdata.no/dokument/SF/forskrift/2006-06-23-724/*#*.

Kunnskapsdepartementet. (2008). *Stortingsmelding nr. 31 (2007-2008): Kvalitet i skolen*. (31). Retrieved from http://www.regjeringen.no/nb/dep/kd/dok/regpubl/stmeld/2007-2008/stmeld-nr-31-2007-2008-.html?id=516853.

Lagerstrøm, B. O. (2007). Kompetanse i grunnskolen: Hovedresultater 2005/2006 *Rapporter*. Oslo-Kongsvinger: Statistisk sentralbyrå.

Langeland, A. S. (2012). Investigating vocabulary development in English from grade 5 to grade 7 in Norwegian primary school. In A. Hasselgreen, I. Drew & B. Sørheim (Eds.), *The young language learner: Research-based insights into teaching and learning* (pp. 131-144). Bergen: Fagbokforlaget.

Langeland, A. S. (2013). Action research into the use of popular music: A goldmine worth exploring in the ELT secondary-school classroom? *Nordic Journal of Modern Language Methodology, 2*(1), 1-24.

Larson-Hall, J. (2008). Weighing the benefits of studying a foreign language at a younger starting age in a minimal input situation. *Second Language Research, 24*(1), 35-63. doi: 10.1177/0267658307082981

Larssen, D. L. S. (2010). The effects of using the European Language Portfolio - Learning strategy awareness and use in Young English Learners. In T. M. Guldal, C. F. Dons, S. Sagberg, T. Solhaug & R. Tromsdal (Eds.), *FoU i praksis 2009: Rapport fra konferanse om praksisrettet FoU i lærerutdanning* (pp. 119-128). Trondheim: Tapir akademisk forlag.

Lasagabaster, D., & Doiz, A. (2003). Maturational constraints on foreign-language written production. In M. L. García Lecumberri & M. d. P. García Mayo (Eds.), *Age and the acquisition of English as a foreign language* (pp. 136-160). Clevedon: Multilingual Matters.

Laufer, B. (2005). Focus on Form in second language vocabulary learning. *EUROSLA Yearbook, 5*, 223-250.

Lee, J. (2011). Size matters: Early vocabulary as a predictor of language and literacy competence. *Applied Psycholinguistics, 32*(1), 69-92. doi: 10.1017/S0142716410000299

Lehmann, T. (1999a). *Literacy and the tertiary student. Why has the Communicative Approach failed.* (PhD), University of Bergen, Bergen.

Lehmann, T. (1999b). Why do they write as they speak? Interference from speech reflected in tertiary EFL students' writing, and the consequent need to teach writing as a separate language variety. *Anglicana Turkuensia, 20*, 385-395.

Lenneberg, E. H. (1967). *Biological foundations of language*. New York, NY: Wiley.

Leopold, W. (1939). *Speech development of a bilingual child. A linguist's record: Vol. I: Vocabulary growth in the first two years*. Evanston, IL: Northwestern University Press.

Leopold, W. (1947). *Speech development of a bilingual child. A linguist's record: Vol. II: Sound-learning in the first two years*. Evanston, IL: Northwestern University Press.

Leopold, W. (1949a). *Speech development of a bilingual child. A linguist's record: Vol. III: Grammar and general problems in the first two years*. Evanston, IL: Northwestern University Press.

Leopold, W. (1949b). *Speech development of a bilingual child. A linguist's record: Vol. IV: Diary from age 2*. Evanston, IL: Northwestern University Press.

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 Child Psychiatry, 51*(5), 612-620. doi: 10.1111/j.1469-7610.2009.02185.x

Lewis, P. M., Simons, G. F., & Fennig, C. D. (Eds.). (2013). *Ethnologue: Languages of the world* (17 ed.). Dallas, TX: SIL International.

Lie, S. (1993). *Kontrastiv grammatikk - med norsk i sentrum*. Oslo: Novus forlag.

Lie, S., Hopfenbeck, T. N., Ibsen, E., & Turmo, A. (2005). Nasjonale prøver på ny prøve: rapport fra en utvalgsundersøkelse for å analysere og vurdere kvaliteten på oppgaver og resultater til nasjonale prøver våren 2005. *Acta Didactica*(1), 1-160.

Lie, T. (2004). Hvorfor synger vi på engelsk? *Språknytt*.

Lieven, E. (2010). Input and first language acquisition: Evaluating the role of frequency. *Lingua, 120*(11), 2546-2556. doi: 10.1016/j.lingua.2010.06.005

Lightbown, P. M. (1983). Exploring relationships between developmental and instructional sequences in L2 acquisition. In H. W. Seliger & M. H. Long (Eds.), *Classroom-oriented research in second language acquisition* (pp. 217-243). Rowley, MA: Newbury House.

Lightbown, P. M. (2000). Classroom SLA research and second language teaching. *Applied Linguistics, 21*(4), 431-462. doi: 10.1093/applin/21.4.431

Lightbown, P. M. (2012). Intensive L2 istruction in Canada: Why not immersion? In C. Muñoz (Ed.), *Intensive exposure experiences in second language learning* (pp. 25-44). Clevedon: Multilingual matters.

Lightbown, P. M., & Spada, N. (1990). Focus-on-Form and corrective feedback in communicative language teaching. *Studies in Second Language Acquisition, 12*(04), 429-448. doi: 10.1017/S0272263100009517

Lillevold, A. J., & Moseng, E. (2006). *Doors 1*. Oslo: Aschehoug.

Lillevold, A. J., & Whittaker, F. (2006). Doors 1-2 Lærer-CD. Oslo: Aschehough.

Lindgren, E., & Muñoz, C. (2012). The influence of exposure, parents, and linguistic distance on young European learners' foreign language comprehension. *International Journal of Multilingualism, 10*(1), 105-129. doi: 10.1080/14790718.2012.679275

Lindqvist, C. (2010). Inter- and intralingual lexical influences in advanced learners' French L3 oral production. *IRAL, 48*(2-3), 131-157. doi: 10.1515/iral.2010.007

Ljosland, R. (2007). English in Norwegian academia: a step towards diglossia? *World Englishes, 26*(4), 395–410. doi: 10.1111/j.1467-971X.2007.00519.x

Ljosland, R. (2010). Teaching through English: Monolingual policy meets multilingual practice. *Hermes, 45*, 99-114.

Ljosland, R. (2011). English as an academic Lingua Franca: Language policies and multilingual practices in a Norwegian university. *Journal of Pragmatics, 43*(4), 991–1004. doi: 10.1016/j.pragma.2010.08.007

Locke, J. L. (1997). A theory of neurolinguistic development. *Brain and Language, 58*(2), 265-326. doi: 10.1006/brln.1997.1791

Long, M. H. (1990). Maturational constraints on language development. *Studies in Second Language Acquisition, 12*(3), 251-285. doi: 10.1017/S0272263100009165

Long, M. H. (1991). Focus on form: A design feature in language teaching methodology. In K. De Bot, R. Ginsberg & C. Kramsch (Eds.), *Foreign language research in cross-cultural perspective* (pp. 39-52). Amsterdam: John Benjamins.

Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413-468). New York, NY: Academic Press.

Lund, A. (2006). The multiple contexts of online language teaching. *Language Teaching Research, 10*(2), 181-204. doi: 10.1191/1362168806lr191oa

Lund, R. E. (2012). Sing English: Om sangens potensial i engelskopplæringen. *Acta Didactica Norge, 6*(1), 1-15.

Lund, R. E. (2014). Writing in EFL teachers' education. *Acta Didactica Norge, 8*(1), 1-18.

Lyster, S.-A. H., Horn, E., & Rygvold, A.-L. (2010). Ordforråd og ordforrådsutvikling hos norske barn og unge. Resultater fra en utprøving av British Picture Vocabulary Scale, Second Edition (BPVS II). *Spesialpedagogikk*(09), 35-43.

Mackey, W. (1962). The description of bilingualism. *Canadian Journal of Linguistics, 71*, 51-85.

MacWhinney, B. (2001). The Competition Model: The input the context and the brain. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 69-90). New York, NY: Cambridge University Press.

Martin, K. I., & Ellis, N. C. (2012). The roles of phonological short-term memory and working memory in L2 grammar and vocabulary learning. *Studies in Second Language Acquisition, 34*(3), 379-413. doi: 10.1017/S0272263112000125

MCG Media Consulting Group. (2009). Study on the use of subtitling: The potential of subtitling to encourage foreign language learning and improve the mastery of foreign languages: European Commission.

McKay, P. (2006). *Assessing young language learners*. Cambridge: Cambridge University Press.

Mechelli, A., Crinion, J. T., Noppeney, U., O'Doherty, J., Ashburner, J., Frackowiak, R. S., & Price, C. J. (2004). Neurolinguistics: Structural plasticity in the bilingual brain. *Nature, 431*(7010), 757-757. doi: 10.1038/431757a

Meisel, J. (2001). The simultaneous acquisition of two first languages: Early differentiation and subsequent development of grammars. In J. Cenoz & F. Genesee (Eds.), *Trends in bilingual acquisition* (pp. 11-42). Amsterdam/Philadelphia, PA: John Benjamins.

Meisel, J. (2004). The bilingual child. In T. K. Bhatia & W. C. Ritchie (Eds.), *The handbook of bilingualism* (pp. 91-113). Malden, MA: Blackwell.

Meisel, J. (2009). Second language acquisition in early childhood *Zeitschrift für Sprachwissenschaft* (Vol. 28, pp. 5-34).

Melby-Lervåg, M., & Lervåg, A. (2013). Reading comprehension and its underlying components in second-language learners: A meta-analysis of studies comparing first- and second-language learners. *Psychological Bulletin*, 409-433. doi: 10.1037/a0033890

Mellegård, I., & Pettersen, K. D. (2012). Curriculum practice: English teachers' understanding and realisation of the new national curriculum. In A. Hasselgreen, I. Drew & B. Sørheim (Eds.), *The young language learner: Research-based insights into teaching and learning* (pp. 207-218). Bergen: Fagbokforlaget.

Merkies, W. J., & Tuin, D. (1987). Foreign language education in Norway: A comparison. *Levende Talen, 419*(Mar), 184-187.

Milton, J., & Alexiou, T. (2006). Language aptitude development in young learners. In C. Abello-Contesse, R. Chacón-Beltrán, M. D. López-Jiménez & M. M. Torreblancaa-López (Eds.), *Age in L2 acquisition and teaching* (pp. 177-192). Bern: Peter Lang.

Miralpeix, I. (2006). Age and vocabulary acquisition in English as a foreign language. In C. Muñoz (Ed.), *Age and the rate of foreign language learning* (pp. 89-106). Clevedon: Multilingual Matters.

Mirman, D., Magnuson, J. S., Estes, K. G., & Dixon, J. A. (2008). The link between statistical segmentation and word learning in adults. *Cognition, 108*(1), 271-280.

Mitchell, R., & Myles, F. (1998). *Second language learning theories*. London: Arnold.

Mora, J. C. (2006). Age effects on oral fluency development. In C. Muñoz (Ed.), *Age and the rate of foreign language learning* (pp. 65-88). Clevedon: Multilingual Matters.

Morgan-Short, K., Steinhauer, K., Sanz, C., & Ullman, M. T. (2012). Explicit and implicit second language training differentially affect the achievement of native-like brain activation patterns. *Journal of Cognitive Neuroscience, 24*(4), 933-947. doi: 10.1162/jocn_a_00119

Munden, J., Elind, M., Engvall, J., & Oscarsson, M. (2006). *Steps 1-2*. Oslo: Gyldendal.

Muñoz, C. (2003). Variation in oral skills development and age of onset *Age and the acquisition of English as a foreign language* (pp. 161-181). Clevedon: Multilingual Matters.

Muñoz, C. (2006). The BAF Project: Research on the effects of age on foreign language acquisition. In C. Abello-Contesse, R. Chacón-Beltrán, M. D. López-Jiménez & M. M. Torreblancaa-López (Eds.), *Age in L2 acquisition and teaching* (pp. 81-92). Bern: Peter Lang.

Muñoz, C. (2008a). Age-related differences in foreign language learning. Revisiting the empirical evidence. *IRAL - International Review of Applied Linguistics in Language Teaching, 46*(3), 197-220. doi: 10.1515/IRAL.2008.009

Muñoz, C. (2008b). Symmetries and asymmetries of age effects in naturalistic and instructed L2 learning. *Applied Linguistics, 29*(4), 578-596. doi: 10.1093/applin/amm056

Muñoz, C. (2009). Input and long-term effects of early learning in a formal setting. In M. Nikolov (Ed.), *The age factor and early language learning*. Berlin: Mouton de Gruyter.

Muñoz, C., & Singleton, D. (2011). A critical review of age-related research on L2 ultimate attainment. *Language Teaching, 44*(01), 1-35. doi: doi:10.1017/S0261444810000327

Murphy, S. (2003). Second language transfer during third language acquisition. *Teachers College, Columbia University Working Papers in TESOL & Applied Linguistics, 3*(1), 1-21.

Murphy, V. A. (2010). The relationship between age of learning and type of linguistic exposure in children learning a second language. In E. Macaro (Ed.), *Continuum companion to*

second language acquisition (pp. 158-178). London/New York, NY: Continuum International Publishing Group.

Nacey, S. L., & Graedler, A.-L. (2013). Communication strategies used by Norwegian students of English. In S. Granger, G. Gilquin & F. Meunier (Eds.), *Twenty years of learner corpus research. Looking back, moving ahead. Corpora and language in use - Proceedings 1* (pp. 345-356). Louvain-la-Neuve: Presses universitaires de Louvain.

Nassaji, H., & Fotos, S. (2011). *Teaching grammar in second language classrooms: Integrating form-focused instruction in communicative context*. New York: Routledge.

Navés, T., Torras, M. R., & Celaya, M. L. (2003). Long term effects of an earlier start: An analysis of EFL written production. *EUROSLA Yearbook, 3*, 103-130. doi: 10.1075/eurosla.3.08nav

Newport, E. L. (1988). Constraints on learning and their role in language acquisition: Studies of the acquisition of American Sign Language. *Language Sciences, 10*(1), 147-172. doi: 10.1016/0388-0001(88)90010-1

Newport, E. L. (1990). Maturational constraints on language learning. *Cognitive science, 14*(1), 11-28. doi: 10.1016/0364-0213(90)90024-Q

Nicolaisen, K. (2011). *Teachers' L1 vs. L2 use in the English classroom in Norwegian schools: A study of how often, when and why teachers of English in Norway use Norwegian over English in the English classroom.* (Master thesis), NTNU Norwegian University of Science and Technology.

Nicolay, A.-C., & Poncelet, M. (2013). Cognitive advantage in children enrolled in a second-language immersion elementary school program for three years. *Bilingualism, 16*(3), 597-607. doi: 10.1017/S1366728912000375

Nikolov, M. (2009a). *The age factor and early language learning*. Berlin: Mouton de Gruyter.

Nikolov, M. (2009b). The age factor in context. In M. Nikolov (Ed.), *The age factor and early language learning* (pp. 1-38). Berlin: Mouton de Gruyter.

Nikolov, M. (2009c). *Early learning of modern foreign languages: Processes and outcomes*. Bristol: Multilingual Matters.

Nikolov, M., & Djigunović, J. M. (2006). Recent research on age, second language acquisition, and early foreign language learning. *Annual Review of Applied Linguistics, 26*(Jan), 234-260.

Nikolov, M., & Djigunović, J. M. (2011). All shades of every color: An overview of early teaching and learning of foreign languages. *Annual Review of Applied Linguistics, 31*, 95-119. doi: doi:10.1017/S0267190511000183

Nikolov, M., Djigunović, J. M., Mattheoudakis, M., Lundberg, G., & Flanagan, T. (2007). *Teaching modern languages to young learners: teachers, curricula and materials*. Strasbourg: Council of Europe.

Nordbø, N. (2013, 07.03.2013). Oversettere fra engelsk under press *nrk.no*. Retrieved from http://www.nrk.no/kultur/litteratur/oversettere-fra-engelsk-under-press-1.10939619

Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning, 50*(3), 417-528. doi: 10.1111/0023-8333.00136

Norås, S. (2007). *"Det høres så mye mer fancy ut å plotte inn litt engelsk..." Our Norwegian-English language - are we just borrowing words, or are we actually bilingual?* (MA thesis), NTNU Norwegian University of Science and Technology.

Odlin, T. (1994). *Perspectives on pedagogical grammar*. Cambridge: Cambridge University Press.

Oller, D. K., & Pearson, B. Z. (2002). Assessing the effects of bilingualism: A background. In D. K. Oller & R. E. Eilers (Eds.), *Language and literacy in bilingual children* (pp. 3-21). Clevedon, England: Multilingual Matters.

Oller, D. K., Pearson, B. Z., & Cobo-Lewis, A. B. (2007). Profile effects in early bilingual language and literacy. *Applied Psycholinguistics, 28*(2), 191-230. doi: 10.1017/S0142716407070117

Oller, J. W., & Nagato, N. (1974). The long-term effects of FLES: An experiment. *The Modern Language Journal, 58*(1-2), 15-19.

Olsen, S. (1999). Errors and compensatory strategies: A study of grammar and vocabulary in texts written by Norwegian learners of English. *System, 27*(2), 191-205. doi: 10.1016/S0346-251X(99)00016-0

Olsthoorn, N. M., Andringa, S., & Hulstijn, J. H. (2012). Visual and auditory digit-span performance in native and non-native speakers. *International Journal of Bilingualism*, 1-11. doi: 10.1177/1367006912466314

Orosz, A. (2009). Factors influencing young learners' vocabulary acquisition. In M. Nikolov (Ed.), *Early learning of modern foreign languages: Processes and outcomes* (pp. 181-194). Bristol: Multilingual Matters.

Ortega, L. (2009). *Understanding second language acquisition*. London: Hodder Education.

Ortega, L. (2014). Ways forward for a bi/multilingual turn in SLA. In S. May (Ed.), *The multilingual turn: Implications for SLA, TESOL and bilingual education* (pp. 32-53). New York: Routledge.

Peal, E., & Lambert, W. E. (1962). The relation of bilingualism to intelligence. *Psychological Monographs: General and Applied, 76*(27), 1-23. doi: 10.1037/h0093840

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Statistical learning in a natural language by 8-month-old infants. *Child Development, 80*(3), 674-685. doi: 10.1111/j.1467-8624.2009.01290.x

Penfield, W., & Roberts, L. (1959). *Speech and brain-mechanisms*. Princeton, NJ: Princeton University Press.

Pérez-Cañado, M. L. (2012). CLIL research in Europe: Past, present, and future. *International Journal of Bilingual Education and Bilingualism, 15*(3), 315-341. doi: 10.1080/13670050.2011.630064

Phillipson, R. (1992). *Linguistic imperialism*. Oxford: Oxford University Press.

Pienemann, M. (1984). Psychological constraints on the teachability of languages. *Studies in Second Language Acquisition, 6*(02), 186-214. doi: 10.1017/S0272263100005015

Pienemann, M. (1989). Is language teachable? Psycholinguistic experiments and hypotheses. *Applied Linguistics, 10*(1), 52-79. doi: 10.1093/applin/10.1.52

Pinker, S. (1999). *Words and rules: The ingredients of language*. London: Weidenfeld & Nicolson.

Piske, T., & Young-Scholten, M. (2009). Introduction. In T. Piske & M. Young-Scholten (Eds.), *Input matters in SLA* (pp. 1-28). Bristol: Multilingual Matters.

Poplack, S. (1980). Sometimes I'll start a sentence in English Y TERMINO EN ESPAÑOL: Toward a typology of code-switching. *Linguistics, 18*, 581-618. doi: 10.1515/ling.1980.18.7-8.581

Pulvermuller, F. F., & Schumann, J. H. (1994). Neurobiological mechanisms of language acquisition. *Language Learning, 44*(4), 681-734. doi: 10.1111/j.1467-1770.1994.tb00635.x

Raaen, G. F. (2009). "Fluency is for everybody, accuracy for the few, and absolute accuracy for the very, very few". In B. Groven, T. M. Guldal, O. F. Lillemyr, N. Naastad & F. Rønning (Eds.), *FoU i praksis 2008: Rapport fra konferanse om praksisrettet FoU i lærerutdanning* (pp. 307-315). Trondheim: Tapir akademisk forlag.

Raaen, G. F. (2010). A comparison of basic English skills in 7th and 10th grade. In T. M. Guldal, C. F. Dons, S. Sagberg, T. Solhaug & R. Tromsdal (Eds.), *FoU i praksis 2009: Rapport fra konferanse om praksisrettet FoU i lærerutdanning* (pp. 205-213). Trondheim: Tapir akademisk forlag.

Raaen, G. F., & Guldal, T. M. (2012). Basic skills: Orthography and sentence complexity in written English from grades 7 to 10. In A. Hasselgreen, I. Drew & B. Sørheim (Eds.), *The young language learner: Research-based insights into teaching and learning* (pp. 101-118). Bergen: Fagbokforlaget.

Rhodes, N. C., & Pufahl, I. (2008). *Foreign Language Teaching in U.S. Schools: Results of a National Survey*. Washington DC: Center for Applied Linguistics.

Robinson, P. (1995). Attention, memory, and the "Noticing" Hypothesis. *Language Learning, 45*(2), 283-331. doi: 10.1111/j.1467-1770.1995.tb00441.x

Robinson, P. (1997). Individual differences and the fundamental similarity of implicit and explicit adult second language learning. *Language Learning*, *47*, 45-99. doi: 10.1111/0023-8333.21997002

Roe, A. W., Pallas, S. L., Hahm, J.-O., & Sur, M. (1990). A map of visual space induced in primary auditory cortex. *Science, 250*, 818-820. doi: 10.1126/science.2237432

Romaine, S. (1999). Bilingual language development. In M. Barrett (Ed.), *The development of language* (pp. 251-275). Philadelphia, PA: Psychology Press.

Ronjat, J. (1913). *Le développement du langage observé chez un enfant bilingue*. Paris: Champion.

Rosansky, E. J. (1975). The Critical Period for the acquisition of language: Some cognitive developmental considerations. *Working Papers on Bilingualism*(6), 1-12.

Ruben, R. J. (1997). A time frame of critical/sensitive periods of language development. *Acta oto-laryngologica*, *117*(2), 202-205. doi: 10.1007/BF02996542

Rugesæter, K. N. (1995). The missing phonemes revisited (Or: "What the language lab did not teach them"). *Språk og Språkundervisning, 28*(1), 3-8.

Rugesæter, K. N. (2012). Phonological competence in English among Norwegian pupils and implications for the teaching of pronunciation in the classroom. In A. Hasselgreen, I. Drew & B. Sørheim (Eds.), *The young language learner: Research-based insights into teaching and learning* (pp. 119-130). Bergen: Fagbokforlaget.

Ruiz-González, G. (2006). Age effects on single phoneme perception. In C. Abello-Contesse, R. Chacón-Beltrán, M. D. López-Jiménez & M. M. Torreblancaa-López (Eds.), *Age in L2 acquisition and teaching* (pp. 155-173). Bern: Peter Lang.

Rørvik, S. (2013). *Texture in learner language.* UiO, Oslo. (PhD thesis)

Rørvik, S., & Egan, T. (2013). Connectors in the argumentative writing of Norwegian novice writers. In S. Granger, G. Gilquin & F. Meunier (Eds.), *Twenty years of learner corpus research: Looking back, moving ahead. Corpora and language in use – Proceedings 1*. (pp. 401-410). Louvain-la-Neuve: Presses universitaires de Louvain.

Saer, D. J. (1923). The effect of bilingualism on intelligence. *British Journal of Psychology, 14*(1), 25-38. doi: 10.1111/j.2044-8295.1923.tb00110.x

Saffran, J. R. (2001). The use of predictive dependencies in language learning. *Journal of Memory and Language, 44*(4), 493-515. doi: http://dx.doi.org/10.1006/jmla.2000.2759

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language, 35*(4), 606-621. doi: 10.1006/jmla.1996.0032

Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental language learning: listening (and learning) out of the corner of your ear. *Psychological Science, 8*(2), 101-105. doi: 10.1111/j.1467-9280.1997.tb00690.x

Schachter, J. (1998). Recent research in language learning studies: Promises and problems. *Language Learning, 48*(4), 557-583. doi: 10.1111/0023-8333.00059

Schei, L. K. (1977). Modern language teaching in Norway. *Modern Languages in Scotland, 14*(Sept), 85-98.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (pp. 3-32). Cambridge: Cambridge University Press.

Schwartz, B. D. (1993). On explicit and negative data effecting and affecting competence and linguistic behavior. *Studies in Second Language Acquisition, 15*(2), 147-163. doi: 10.1017/S0272263100011931

Sebastian-Galles, N., & Diaz, B. (2012). First and second language speech perception: Graded learning. *Language Learning, 62*(S2), 131-147. doi: 10.1111/j.1467-9922.2012.00709.x

Seip, B. E., & Arnt, M. (2004). The Invisible Text: A report on subtitling for Norwegian television. Oslo: IMK services, Institute of Media and Communication, University of Oslo.

Seliger, H. W. (1978). Implications of a multiple critical periods hypothesis for second language learning. In W. C. Ritchie (Ed.), *Second language acquisition research* (pp. 11-19). New York, NY: Academic Press.

Selinker, L. (1972). Interlanguage. *IRAL - International Review of Applied Linguistics in Language Teaching, 10*(1-4), 209-231. doi: 10.1515/iral.1972.10.1-4.209

Serratrice, L. (2012). The bilingual child. In T. K. Bhatia & W. C. Ritchie (Eds.), *Handbook of bilingualism and multilingualism* (2 ed., pp. 87-108). Somerset, NJ: Wiley.

Shintani, N. (2011). A comparative study of the effects of input-based and production-based instruction on vocabulary acquisition by young EFL learners. *Language Teaching Research, 15*(2), 137-158. doi: 10.1177/1362168810388692

Simensen, A. M. (2010). Fluency: an aim in teaching and a criterion in assessment. *Acta Didactica Norge, 4*(1), 1-13.

Simensen, A. M. (2011). The relationship between an applied discipline and its 'parent disciplines' in English teaching in Norway. *Histoire Epistemologie Langage, 33*(1), 39-52.

Singleton, D., & Ryan, L. (2004). *Language acquisition: The age factor*. Clevedon: Multilingual Matters.

Sivertzen, T. G. (2013). *The role of input in early second language acquisition: A study of the long-term effects of initial extra English input on Norwegian 4th graders' receptive vocabulary* (Master thesis), NTNU.

Skehan, P. (1986). Where does language aptitude come from? In P. Meara (Ed.), *Spoken Language* (pp. 95-113). London: Centre for Information on Language Teaching.

Skehan, P. (1989). *Individual differences in second-language learning*. New York, NY: Edward Arnold.

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press.

Skehan, P. (2002). Theorising and updating aptitude. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 69-94). Amsterdam/Philadelphia, PA: John Benjamins.

Skehan, P. (2012). Language aptitude. In A. Mackey & S. M. Gass (Eds.), *The Routledge Handbook of Second Language Acquisition* (pp. 381-395). London: Routledge.
Skeiseid, G. (1995). Lyttedelen i avgangsprøva i engelsk i 9. klasse. *Språk og Språkundervisning, 28*(1), 6-14.
Skinner, B. F. (1957). *Verbal behavior*. New York: Appleton-Century-Crofts.
Snow, C. E., & Hoefnagel-Höhle, M. (1977). Age differences in the pronunciation of foreign sounds. *Language and speech, 20*(4), 357-365. doi: 10.1177/002383097702000407
Sparks, R. L., Patton, J., Ganschow, L., & Humbach, N. (2009). Long-term relationships among early first language skills, second language aptitude, second language affect, and later second language proficiency. *Applied Psycholinguistics, 30*(4), 725-755. doi: 10.1017/S0142716409990099
Sparks, R. L., Patton, J., Ganschow, L., Humbach, N., & Javorsky, J. (2006). Native language predictors of foreign language proficiency and foreign language aptitude. *Annals of Dyslexia, 56*(1), 129-160.
Språkrådet, & NHO. (2012). Språkvett for næringslivet.
Stern, H. H. (1983). *Fundamental concepts of language teaching*. Oxford: Oxford University Press.
Strand, O. L. (2014). *Do we have to speak English? A longitudinal perspective on the impact of early English input on the development of English comprehension skills in Norwegian pupils.* (Master thesis), NTNU.
Sunderman, G., & Kroll, J. F. (2006). First language activation during second language lexical processing: An investigation of lexical form, meaning, and grammatical class. *Studies in Second Language Acquisition, 28*(03), 387-422. doi: 10.1017/S0272263106060177
Svenhard, B. W. (2010). *CLIL - engelsk og fremmedspråk i andre fag* (Vol. Nummer 24/2010). Halden: Fremmedspråksenteret.
Svenhard, B. W. (2012). *CLIL: Kombinert fag- og engelskopplæring i videregående skole* (Vol. nr. 28/april 2012). Halden: Fremmedspråksenteret.
Svenhard, B. W., Servant, K., Hellekjær, G. O., & Bøhn, H. (2007). Norway. In A. Maljers (Ed.), *Windows on CLIL : content and language integrated learning in the European spotlight* (pp. 139-146). The Hague: European Platform for Dutch education, 2007.
Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. M. Gass & C. G. Madden (Eds.), *Input in second language acquisition* (pp. 235-253). Rowley, MA: Newbury House.
Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seildlhofer (Eds.), *Principle & practice in applied linguistics: studies in honour of H.G. Widdowson* (pp. 125-144). Oxford: Oxford University Press.
Swain, M., & Lapkin, S. (1995). Problems in output and the cognitive processes they generate: A step towards second language learning. *Applied Linguistics, 16*, 371-391. doi: 10.1093/applin/16.3.371
Thiery, C. (1978). True bilingualism and second language learning. In D. Gerver & H. W. Sinaiko (Eds.), *Language interpretation and communication* (pp. 145-153). New York, NY: Plenum.
Thompson, S. P., & Newport, E. L. (2007). Statistical learning of syntax: The role of transitional probability. *Language Learning and Development, 3*(1), 1-42. doi: 10.1080/15475440709336999

Torgersen, E. N. (2002). Temporal factors in perception of the voicing contrast: Immediate semantic effects on speech processing and the L2 learner. *Proceedings from Temporal Integration in Speech* (Vol. 50). Cambridge: Cambridge University Press.

Torgersen, E. N. (2003). Temporal and semantic factors in perception of the voicing contrast: Different processing strategies in L1 and L2. In M. J. Solé, D. Recasens & J. Romero (Eds.), *Proceedings from the 15th International Congress of Phonetic Sciences, Barcelona* (pp. 2897-2900). Universitat Autònoma de Barcelona: Barcelona.

Toro, J. M., Sinnett, S., & Soto-Faraco, S. (2005). Speech segmentation by statistical learning depends on attention. *Cognition, 97*(2), B25-B34. doi: http://dx.doi.org/10.1016/j.cognition.2005.01.006

Torras, M. R., Navés, T., Celaya, M. L., & Pérez-Vidal, C. (2006). Age and IL development in writing. In C. Muñoz (Ed.), *Age and the rate of foreign language learning* (Vol. 156-182). Clevedon: Multilingual Matters.

Treffers-Daller, J. (2011). Operationalizing and measuring language dominance. *International Journal of Bilingualism, 15*(2), 147-163. doi: 10.1177/1367006910381186

Tremblay, M.-C. (2006). Cross-linguistic influence in third language acquisition: The role of L2 proficiency and L2 exposure. *Cahiers linguistiques d'Ottawa, 34*(Jan), 109-119.

Truscott, J. (1998). Noticing in second language acquisition: A critical review. *Second Language Research, 14*(2), 103-135. doi: 10.1191/026765898674803209

Tucker, G. R. (1998). A global perspective on multilingualism and multilingual education. In J. Cenoz & F. Genesee (Eds.), *Beyond bilingualism: multilingualism and multilingual education* (pp. 3-15). Clevedon: Multilingual Matters.

Ullman, M. T. (2001). The declarative/procedural model of lexicon and grammar. *Journal of Psycholinguistic Research, 30*(1), 37-69. doi: 10.1023/A:1005204207369

Unsworth, S., Persson, L., Prins, T., & De Bot, K. (2014). An investigation of factors affecting early foreign language learning in the Netherlands. *Applied Linguistics*, 1-23. doi: 10.1093/applin/amt052

Ursin, L. H., Aursland, T., & Ripegutu, H. (2012). Advarer mot økt dubbing på TV. *bt.no*. Retrieved from http://www.bt.no/bergenpuls/Advarer-mot-okt-dubbing-pa-TV-2752121.html#.UhILHZIyb6M

Utdanningsdirektoratet. (2006). Knowledge Promotion: Curriculum in English: Utdanningsdirektoratet, www.udir.no.

Utdanningsdirektoratet. (2007). Ansvaret for fordeling av skoledager utover året: Utdanningsdirektoratet, www.udir.no.

Utdanningsdirektoratet. (2013a). *English subject curriculum*. Retrieved from http://www.udir.no/kl06/ENG1-03/Hele/?lplang=eng.

Utdanningsdirektoratet. (2013b). *Norwegian subject curriculum*. Retrieved from http://www.udir.no/kl06/ENG1-03/Hele/?lplang=eng.

Utdanningsdirektoratet. (2013c). *Udir-1-2013 Kunnskapsløftet fag- og timefordeling og tilbudsstruktur*. Retrieved from http://www.udir.no/Regelverk/Rundskriv/2013/Udir-1-2013-Kunnskapsloftet-fag--og-timefordeling-og-tilbudsstruktur/Udir-1-2013-Vedlegg-1/2-Grunnskolen/.

Van Buskirk, W. R. (1979). Aspects of the English language; its position and teaching in Norway. *English Around the World, 20*(May), 1-2.

van Dommelen, W. A., & Hazan, V. (2010). Perception of English consonants in noise by native and Norwegian listeners. *Speech Communication, 52*(11-12), 968-979. doi: 10.1016/j.specom.2010.05.001

VanPatten, B. (1996). *Input processing and grammar instruction in second language acquisition*. Norwood, NJ: Ablex.

VanPatten, B. (2012). Input processing. In A. Mackey & S. M. Gass (Eds.), *The Routledge handbook of second language acquisition* (pp. 268-281). London: Routledge.

VanPatten, B., & Cadierno, T. (1993a). Explicit instruction and input processing. *Studies in Second Language Acquisition, 15*(02), 225-243. doi: 10.1017/S0272263100011979

VanPatten, B., & Cadierno, T. (1993b). Input processing and second language acquisition: A role for instruction. *Modern Language Journal, 77*, 45-57. doi: 10.1111/j.1540-4781.1993.tb01944.x

Volterra, V., & Taeschner, T. (1978). The acquisition and development of language by bilingual children. *Journal of Child Language, 5*, 311-326. doi: 10.1017/S0305000900007492

Vouloumanos, A. (2008). Fine-grained sensitivity to statistical information in adult word learning. *Cognition, 107*(2), 729-742. doi: http://dx.doi.org/10.1016/j.cognition.2007.08.007

Vulchanova, M., Vulchanov, V., Sarzhanova, D., & Eshuis, H. (2012). The role of input in early bilingual lexical development. *Lingue e linguaggio*(2), 181-198. doi: 10.1418/38785

Waara, R. (2003). A Norwegian study: Teachers' and learners' perceptions of upper secondary foreign language education. In D. Newby (Ed.), *Mediating between theory and practice in the context of different learning cultures and languages* (pp. 131-136). Strasbourg: European Centre for Modern Languages.

Waara, R. (2004). Construal, convention, and constructions in L2 speech. In M. Achard & S. Niemeier (Eds.), *Cognitive linguistics, second language acquisition, and foreign language teaching* (pp. 51-75). Berlin: Mouton de Gruyter.

Waara, R. (2009). Connecting pupils, students and theory in a project: Planning time in the classroom, a crucial element in lesson design. In B. Groven, T. M. Guldal, O. F. Lillemyr, N. Naastad & F. Rønning (Eds.), *FoU i praksis 2008: Rapport fra konferanse om praksisrettet FoU i lærerutdanning* (pp. 317-326). Trondheim: Tapir akademisk forlag.

Waara, R. (2011). Correlations between task complexity and measures of language development in the foreign language classroom. In T. L. Hoel, T. M. Guldal, C. F. Dons, S. Sagberg, T. Solhaug & K. Wæge (Eds.), *FoU i praksis 2010: Rapport fra konferanse om praksisrettet FoU i lærerutdanning* (pp. 531-540). Trondheim: Tapir akademisk forlag.

Wei, L. (2000). Dimensions of bilingualism. In L. Wei (Ed.), *The bilingualism reader* (pp. 2-21). London: Routledge.

Weinreich, U. (1953/1968). *Languages in contact*. The Hague: Mouton.

Werker, J. (2012). Perceptual foundations of bilingual acquisition in infancy. *Ann N Y Acad Sci, 1251*, 50-61. doi: 10.1111/j.1749-6632.2012.06484.x

Werker, J. F., & Byers-Heinlein, K. (2008). Bilingualism in infancy: first steps in perception and comprehension. *Trends in Cognitive Sciences, 12*(4), 144-151. doi: 10.1016/j.tics.2008.01.008

Wesche, M. B. (2002). Early French immersion: How has the original Canadian model stood the test of time? In P. Burmeister, T. Piske & A. Rohde (Eds.), *An integrated view of language development: Papers in honour of Henning Wode* (pp. 357-379). Trier: Wissenschaftlicher Verlag Trier.

Westergaard, M. (2002). Minimal input in L2 acquisition: How Norwegian children learn English word order. *Acta Didactica, 3*, 199-218.

Westergaard, M. (2003). Unlearning V2: Transfer, markedness, and the importance of input cues in the acquisition of word order in English by Norwegian children. *EUROSLA Yearbook, 3*, 77-101. doi: 10.1075/eurosla.3.07wes

Williams, J. N. (2012). Working memory and SLA. In S. M. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 427-441). London: Routledge.

Williams, K. T. (2007). *Expressive Vocabulary Test, Second Edition*. Minneapolis, MN: Pearson Education, Inc.

Wilson, D., & Mollergard, E. (1981). Errors in the production of vowel no. 10 /(turned v)/ by Norwegian learners of English. *IRAL, International Review of Applied Linguistics in Language Teaching, 19*(1), 69-76.

Wode, H. (1996). Bilinguale Kindergärten und Vorschulklassen: Struktur, Funktionsweise, Leistungsfähigkeit, psychologische Grundlagen. Kiel: Englisches Seminar.

Wode, H. (1997). Bilinguale Kindergärten: Wieso? Weshalb? Warum? *Kita, 10*, 203-207.

Wode, H. (1999). Language learning in European immersion classes. In J. Masih (Ed.), *Learning through a foreign language: Models, methods and outcomes* (pp. 16-29). Grantham: Grantham Book Services.

Yu, C., & Ballard, D. H. (2007). A unified model of early word learning: Integrating statistical and social cues. *Neurocomputing, 70*(13–15), 2149-2165. doi: http://dx.doi.org/10.1016/j.neucom.2006.01.034

Yurovsky, D., Yu, C., & Smith, L. B. (2012). Statistical speech segmentation and word learning in parallel: Scaffolding from child-directed speech. *Frontiers in Psychology, 3*, 1-9. doi: 10.3389/fpsyg.2012.00374

Ørevik, S. (2012). From 'essay' to 'personal text': The role of genre in Norwegian EFL exam papers 1996-2011. *Acta Didactica Norge, 6*(1), 1-21.

Paper I

Naturalistic acquisition in an early language classroom

Anne Dahl*, Mila Dimitrova Vulchanova

Department of Language and Literature, NTNU Norwegian University of Science and Technology, Trondheim, Norway

*** Correspondence:** Anne Dahl, Department of Language and Literature, NTNU Norwegian University of Science and Technology, 7491 Trondheim, Norway
anne.j.dahl@ntnu.no

Keywords: second language acquisition, naturalistic acquisition, input, early-start, classroom.

Abstract

This study investigated whether it is possible to provide naturalistic second language acquisition (SLA) of vocabulary for young learners in a classroom situation without resorting to a classical immersion approach. Participants were 60 first-grade pupils in two Norwegian elementary schools in their first year. The control group followed regular instruction as prescribed by the school curriculum, while the experimental group received increased naturalistic target language input. This entailed extensive use of English by the teacher during English classes, and also during morning meetings and for simple instructions and classroom management throughout the day. Our hypothesis was that it is possible to facilitate naturalistic acquisition through better quality target language exposure within a normal curriculum. The students' English vocabulary knowledge was measured using the Peabody Picture Vocabulary Test, version 4 (PPVT-IV, Dunn & Dunn, 2007a), at the beginning and the end of the first year of school. Findings are that 1) early-start second-language (L2) programs in school do not in themselves guarantee vocabulary development in the first year, 2) a focus on increased exposure to the L2 can lead to a significant increase in receptive vocabulary comprehension in the course of only 8 months, and 3) even with relatively modest input, learners in such an early-start L2 program can display vocabulary

acquisition comparable in some respects to that of younger native children matched on vocabulary size. The overall conclusion is that naturalistic vocabulary acquisition is in fact possible in a classroom setting.

Introduction

Over the past decades, there has been a trend in many countries of lowering starting ages for learning foreign languages, especially English. One reason is globalization and the role of English as an international lingua franca; another is increased knowledge of the benefits of young starting ages for language acquisition. However, the relationship between what we know about language acquisition and what goes on in early language classrooms is not straightforward, and it is not obvious that such classrooms make the best possible use of the learners' young age. A number of studies (e.g., Burstall, 1975; Cenoz, 2003; García Lecumberri & Gallardo, 2003; García Mayo, 2003; Holmstrand, 1982; Lasagabaster & Doiz, 2003; Muñoz, 2006) indicate that an early start in a foreign language does not necessarily make a difference in terms of the pupils' attained competence.

Even though the common assumption that children always acquire languages more easily than adults has been contested (see e.g., D. Singleton & Ryan, 2004, p. 72 ff. for an overview), the conclusion from findings in research is generally that the earlier one starts acquiring a language before adulthood, the better the chances are of attaining target competence (DeKeyser, 2000; Hyltenstam, 1992; Hyltenstam & Abrahamsson, 2003; Johnson & Newport, 1989; D. Singleton & Ryan, 2004 ch. 7). This is often attributed to a difference in learning style, as well as maturational constraints related to a sensitive period in language learning (Bley-Vroman, 1989; Felix, 1985; Newport, 1990). Yet little is known about how the factors known to impact on language acquisition interact in the course of development, and what their relative weighting is.

Nikolov (2009) hypothesizes that a possible explanation for the lack of an early-start advantage in previous studies may be that classroom activities employed in that research were better suited to older learners. Quite often the maturational facts in language acquisition link naturally to the learning style differences, namely that younger learners are more likely to employ implicit learning, whereas older learners outperform them on explicit learning (Muñoz, 2006). It then follows that what younger learners need above and beyond all else is exposure to the target language not explicit instruction and formal training. We know that L2 learners are fully capable

of acquiring linguistic knowledge without intentional effort or instruction, and that reading and listening alone can lead to acquisition especially in young learners (cf. Lightbown, 2000; Lightbown, Halter, White, & Horst, 2002). Amount and quality of input are undoubtedly crucial factors in SLA (cf. Gass, 2003; Hyltenstam, 1992), and there is evidence that sensitivity to frequency is relevant for the acquisition of grammatical items (cf. Goldschneider & DeKeyser, 2001; Larsen-Freeman, 1975). Frequency of language items and volume of language exposure have also been demonstrated to influence vocabulary size, at least in L1 acquisition (Childers & Tomasello, 2002; Hart & Risley, 1995; Hoff & Naigles, 2002; Vulchanova, Vulchanov, Sarzhanova, & Eshuis, 2012), e.g., contributing to learning from distributional cues, a mechanism found in both L1 and L2 acquisition (e.g., Pelucchi, Hay, & Saffran, 2009; Saffran, Newport, & Aslin, 1996; Saffran, Newport, Aslin, Tunick, & Barrueco, 1997).

As Wode (1981, p. 302) points out, "[t]here is no learner on record who learned a language or even part of it without receiving some language input". Instruction and explicit knowledge may play a role in SLA, specifically in compensating for the limited time and opportunity for exposure in the language classroom (cf. Lightbown, 2000). However, it is likely that explicit instruction is less relevant for young learners, and that cognitive maturity may be necessary in order for explicit forms of instruction to make up for impoverished input (see e.g., DeKeyser, 2000; DeKeyser & Larson-Hall, 2005; Larson-Hall, 2008; Muñoz, 2001). There is thus reason to believe that high-input child SLA contexts are the successful ones, and that both intensity and continuation of exposure are decisive factors (Abello-Contesse, Chacón-Beltrán, López-Jiménez, & Torreblancaa-López, 2006; Burstall, 1975; Larson-Hall, 2008: 56; Lightbown, 2000; Ruiz-González, 2006; Stern, 1983).

The crucial question is whether acquisition in early-start L2 classrooms can be significantly improved even with only a limited increase in the amount and density of exposure to English. This can be achieved by giving the language itself a more central place in the English classroom, e.g., in conducting classroom management in English and in prioritizing input-heavy activities such as the teacher reading aloud. In addition, L2 input can be increased also outside of English class by providing classroom management and simple instructions in English throughout the school day. The present study is, to our knowledge, the first study to use such an approach, and to

investigate the effect of such increased input on vocabulary acquisition in the context of English as an L2 in Norway.

Norwegian children start learning English systematically in school from age 6. However, the number of teaching hours is low, normally less than one per week (Utdanningsdirektoratet, 2006, 2007), and the English input to which students are exposed is thus necessarily limited. Furthermore, the Norwegian early English classroom typically does not provide very much L2 input, since it largely uses Norwegian as the language of instruction. One reason for this situation may be that English is not an obligatory subject in teacher training in Norway, although most teachers in lower primary school will normally have to teach the subject (cf. Guldal in Trønder-Avisa, 2007). Also, the curriculum and teaching materials commonly used in the classroom reflect a teaching style where the target language is the object, but not the medium of instruction.

Vocabulary acquisition in an L2 has traditionally been associated with rote learning and memorization of words (cf. Kersten, 2010). However, L2 vocabulary can obviously also be acquired from naturalistic input, as is the case in L1. In fact, vocabulary acquisition may not be subject to age effects. While it is the area of language first evident in young children, we all continue to learn new words throughout life. We know that vocabulary is acquired at a fast pace in school (see e.g., Berman, 2007; Bloom, 2000, 2004; Clark, 1993; Nagy & Herman, 1987; Pinker, 1994). On the other hand, vocabulary is an aspect of language for which L1 and L2 acquisition may be assumed to differ. In L1, vocabulary acquisition entails the daunting task of learning concepts and words at once. The L2 learner, on the other hand, will generally have acquired the concepts already. Many theories have been proposed about bilingual vocabulary acquisition, some involving the L1 as a mediator, while others assume a direct link to the concept. Without engaging in a discussion of the extent to which cross-linguistic lexical variation reflects deeper conceptual differences, we assume that L2 vocabulary acquisition, at least at early stages, and at least when the L1 and the L2 represent similar cultures, does to a large extent entail learning the new labels for familiar concepts (see e.g., MacWhinney, 2005; D. Singleton, 1999, p. 48).

There is thus no reason to believe that neither age nor the presence of an already acquired L1 should have a detrimental effect on vocabulary acquisition, and we should expect that increased exposure to an L2 during the first year of school will lead to naturalistic acquisition and

significantly increased vocabulary comprehension. Specifically, it is likely that input alone is particularly beneficial for vocabulary acquisition in young L2 learners. Shintani (2011) explicitly investigated whether input-only instruction may be as effective as production-based instruction for 6-8-year old Japanese learners of English, hypothesizing that mechanisms such as fast mapping are still available at this age. The conclusions of her study are indeed that in this age group, the effect of input-based instruction on vocabulary acquisition is as good as, or better than, that of production-based instruction.

Method

The present study investigates whether employing a bilingual approach to an otherwise normal Norwegian first-grade English classroom will lead to improved acquisition over one year, compared to a standard, i.e., largely native language-based, first grade class. The research questions are whether children in each of the classes improve significantly in vocabulary acquisition over their first year of school, and whether there is a measurable difference in the two groups' vocabulary comprehension at the end of the first grade.

Conditions

Two different schools were recruited for the experiment. In one school, teachers were told to do nothing out of the ordinary, and to teach English to their first-graders the way they would normally do, with the L1 as the main medium of instruction. In the other school, teachers agreed to use English more extensively with the children in and outside of English class, such as for morning meetings, simple instructions during the day, and reading aloud. However, they were not instructed to avoid the use of the L1; this school's approach to English teaching can thus be said to be bilingually-based.

The two schools were both standard state schools, situated in similar suburban areas in one of Norway's largest towns. The areas from which the schools recruit their pupils are socioeconomically comparable; they are both relatively affluent, with mean incomes slightly above the national average. The ethnic makeup of the two neighborhoods is also comparable, with a low percentage of families with immigrant backgrounds. On the national tests of English for 5th grade in 2008 the two schools scored similarly at or (in the case of the native language-based classroom's school) a little above the average. Thus, there is every reason to believe that

these two schools are comparable in terms of student population and quality of teaching, and that they are representative of Norwegian state schools. In addition, a parent questionnaire asked for background information about the children concerning factors such as foreign travel, English-speaking friends and relatives, and input received through media. None of the participants included in the study had extended stays beyond normal vacations abroad, and none had close English-speaking family or other special circumstances which might make their English competence atypical for a Norwegian 6-year-old. Although the parental reports were relatively crude, information was quantified by counting weeks spent outside of Scandinavia and hours per week with English exposure from games and media prior to starting school. This information is summarized for both groups in Table 1.

Table 1. Mean, minimum and maximum values and standard deviations (SD) for weeks spent outside of Scandinavia and hours of exposure from media, games and music prior to starting school in the bilingually-based and the native language-based groups

	Weeks outside Scandinavia		Exposure from media, games, music	
	BB	NB	BB	NB
Min	0.0	0.0	0.0	0.0
Max	14.0	24.0	11.0	7.0
Mean	4.2	6.4	1.8	1.2
SD	3.6	6.8	2.5	1.6

BB= Bilingually-based
NB= Native language-based

A Mann-Whitney U test found no significant difference for weeks spent outside Scandinavia ($U(59) = 399.0$, $Z = -0.751$, $p = 0.453$), nor on previous exposure to English ($U(59) = 407.0$, $Z = -0.652$, $p = 0.515$). It thus seems safe to assume that these children's English exposure outside of school was similar.

Children in the two groups were also similar on a number of factors that may potentially influence English acquisition, which will be more closely described in the test materials section.

In each school, three different classes and class teachers participated in the project. In the bilingually-based school, one teacher had the main responsibility for English classes in all groups. In this school, groups were often organized across classes for various subjects, and this

teacher was a natural choice for English classes since she was a native speaker of English. However, all class teachers participated in providing input throughout the school day. In each school, one teacher was responsible for recording information on time spent on English, and about activities and materials used, and to report to the researcher. These reports were frequent and relatively informal during the two periods of test sessions in September and May. In the middle of the spring term, both teachers formally reported on the same three questions (time, activities, and materials) in emails to the researcher. Information from both schools indicated that they consistently followed the pattern described below throughout the test period.

The native language-based condition school reported formally spending 30 minutes a week on English class. They also reported spending a few minutes in morning meetings every day talking about the weather and the names of the days in English, but these meetings were otherwise conducted entirely in Norwegian. The time spent on English in this group was thus of around 45 minutes per week, which is representative of the average that normal Norwegian schools spend in the first grade. Also representative is the fact that communication during this time took place mainly in Norwegian. Activities in this group included the use of the workbook Junior Scoop 1-2 (Bruskeland & Ranke, 2005) which is intended for use in first grade, and which contains simple activities, including routine instructions, rhymes and songs. Furthermore, teachers reported a number of other English songs used in class.

The bilingually-based group spent about 30-40 minutes per week on English class. While this school also uses the Scoop series of work- and textbooks, it was not used in this group of children. The teachers instead used various other materials, including simple stories and books, which the teacher read aloud, often with illustrations. Teachers would also spend time talking about pictures or objects. Furthermore, this group spent more time speaking English during morning meeting time; the teacher estimated about 5-10 minutes per day. While the native language-based group's morning meetings were conducted in Norwegian, with only routine discussion of words for the weather and the days of the week in English, morning meetings in the bilingually-based group were more or less conducted in English on the part of the teacher, while the pupils were free to answer in either language as they wished. Teachers in the bilingually-based school also used English for simple classroom management throughout the day, often with

Norwegian translations instantly following, such as the reminder "No running in the corridors – *ikke løpe i gangen*!".

It is important to point out, then, that the change in the English classroom of the bilingually-based group did not consist of more formal instruction or an increase in teaching hours for English. Time spent on English was a little higher than is normal in Norwegian schools, but with an average of around 70 minutes per week including morning meetings, it still is a small proportion of the total time spent at school. Furthermore, there was no focus on pupils' production, even though increased L2 production may have been a natural consequence of the increased input. In other words, the change in this school consisted solely of an increased focus on providing target language exposure in a natural context.

Participants

All parents of students in the relevant first grades were contacted in writing and asked for written consent for their child to participate. Approximately 80% of the parents provided consent in each group. In the bilingually- based group, the total number of volunteers was 59. 10 participants were excluded because they were bilingual, and one because he had participated in another, related study. From the remaining 49 children, 31 were randomly selected for the project by the researcher. The final test group consisted of 17 boys and 14 girls, all monolingual speakers of Norwegian with no known diagnosis which might influence acquisition. In the native-language based group there were 35 volunteers. Three were excluded because of bilingualism, and one because of hearing problems. Two children participated in the pre-test only; one because he was not available during the post-test, the other because he did not want to participate in it. The final test group consisted of 15 boys and 14 girls, all monolingual and with no known diagnosis which may have had consequences for the study.

Mean age at the time of pre-testing was 6;1 in both groups, with no significant difference ($U(59) = 433.0$, $Z = -0.245$, $p = 0.806$). The project was approved by the Norwegian Social Science Data Services (NSD). Table 2 summarizes descriptive statistics for age and scores on background measures in the two groups.

Table 2. Mean, minimum and maximum values and standard deviations (SD) for age, vocabulary, verbal and non-verbal intelligence and memory scores (raw) in the bilingually-based and the native language-based groups

	Age at pretest		Norw. vocabulary		English vocabulary (pre-test)		English sentences (pre-test)		Non-verbal intelligence		Verbal intelligence		Memory	
	BB	NB	BB	NB	BB	NB	BB	NB	BB	NB	BB	NB	BB	NB
Min	5;6	5;6	97	85	2	3	3	2	12	11	11.0	10.0	40	40.0
Max	6;6	6;5	157	145	56	61	8	10	32	38	24.0	26.0	59	62.0
Mean	6;1	6;1	119.9	113.8	25.4	23.7	5.7	5.2	18.7	17.7	16.6	16.1	46.3	48.0
SD	0.027	0.028	14.8	14.6	11.4	13.6	1.5	1.9	4.6	5.2	3.9	4.7	4.6	5.9

BB= Bilingually-based
NB= Native language-based

Test materials

English vocabulary comprehension was tested using Form B of the Peabody Picture Vocabulary Test, Fourth Edition (PPVT™-4). This test measures vocabulary comprehension by means of pictures; the subject hears a word and selects the corresponding picture from a set of four options. This means that issues related to literacy can be avoided, and no L2 production is necessary on the part of the participant. Both these criteria made the test particularly well suited to these young learners, whose level both of literacy and of English was too low, especially in the pre-test, for more comprehensive tests to yield reliable results.

Pre-testing took place within the first six weeks of the children's first school year. During the pre-test session Norwegian vocabulary comprehension was tested in addition to initial English vocabulary comprehension, using a translated version of Form A of the Peabody Picture Vocabulary Test, Fourth Edition (PPVT™-4). There were no significant differences between the groups on either of these tests.

Post-testing took place during the last six weeks of the school year. During this test session, in addition to the post-test of English vocabulary comprehension, visio-spatial working memory was tested using a memory game where the child memorized sets of picture cards which were then turned face down, and was asked to find the pairs in as few attempts as possible. Furthermore, non-verbal intelligence was tested using the Matrices section of the Kaufmann Brief Intelligence Test, Second Edition (KBIT-2), and verbal intelligence using a version of the Riddles

section of the KBIT-2 translated into Norwegian. These particular control measures, including L1 vocabulary, were chosen firstly to control for group differences on the outset, and secondly to provide measures that are believed to correlate with L2 acquisition. There are consistent findings in research suggesting that L2 language competence correlates highly with working memory, non-verbal intelligence, and, most importantly, L1 competence and skills (e.g., Colledge et al., 2002; Dale, Harlaar, Haworth, & Plomin, 2010; Foyn, Vulchanova, Nilsen, & Sigmundsson, under revision; Gathercole, 2006; Hayiou-Thomas, Dale, & Plomin, 2012; Sparks, Patton, Ganschow, & Humbach, 2009). No significant differences between the two groups were found on any of the control measures; see Table 3.

Table 3: Mann-Whitney U, Z, and *p* for between-groups comparison of vocabulary, verbal and non-verbal intelligence and memory in the bilingually-based and the native language-based groups

	Norwegian vocabulary	English vocabulary (pre-test)	English sentences (pre-test)	Non-verbal intelligence	Verbal intelligence	Memory
Mann-Whitney U	352.000	363.000	344.000	333.000	389.500	375.500
Z	-1.443	-1.281	-1.402	-1.747	-0.891	-1.098
p (two-tailed)	0.149	0.200	0.161	0.081	0.373	0.272

Each test session was conducted at the child's school, during school hours or in the after-school program. Testing took place in a quiet room, with only the child, the researcher, and sometimes an assistant present. Each test session lasted for approximately one hour. Most children were able to complete test sessions without signs of fatigue; if they did show signs of losing concentration, they were given a short break. Average time between pre- and post-testing was eight months in both groups.

3 Results

Because the sample is relatively small (native language-based: n=29, bilingually-based: n=31) and because the data are not normally distributed, data were analyzed with non-parametric tests.

Results from the pre-test reveal that the children in general knew very little English when starting school. The mean raw score of the native language-based group was 23.72, which according to

the PPVT™-4 manual has an age equivalent for native English speakers of 2;4. The mean raw score in the bilingually-based group was 25.39, with a native age equivalent of 2;5. In short, these Norwegian children demonstrated English comprehension comparable to very young English-speaking children. Both groups' age equivalents are in fact below the chronological age for which the PPVT™-4 is normed, which has a lower bound of 2;6, although they are above the lower bound for age equivalents, which is 2;0. Competence was very similar between the two groups, even though the bilingually-based group did score slightly higher. An independent samples Mann-Whitney U test reveals that this difference is not significant (U (59) = 363, Z = -1.281, p = 0.20). This is confirmed by the PPVT™-4 Manual, which allows raw scores to be converted into growth scale value (GSV) scores. According to the manual (Dunn & Dunn, 2007b, p. 205), between chronological ages 2;6 and 12;0, a GSV point difference of eight is considered significant. The difference between an average score of 23.72 (GSV 84) and one of 25.39 (GSV 85) is only one point, and is thus not significant. Figure 1 illustrates the scores on the pre-test and the post-test in the two groups.

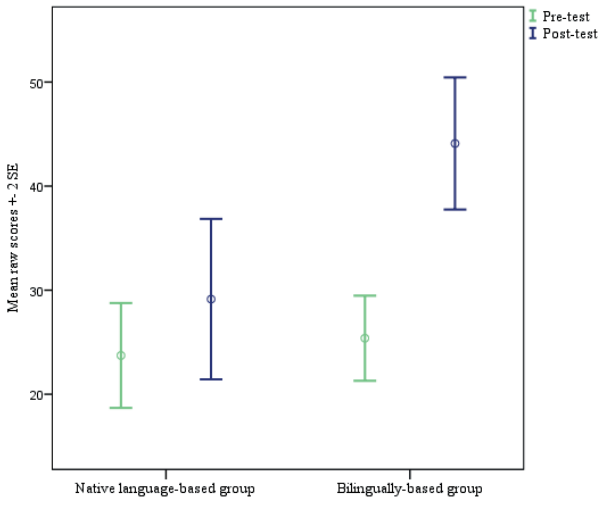


Figure 1: PPVT-IV receptive vocabulary development, pre-test to post-test

After 8 months, the mean raw score on the PPVT™-4 had increased for both groups; to 29.14 for the native language-based group, and to 44.10 for the bilingually-based group. A Mann-Whitney U test found the difference between the two groups at this time to be significant ($U(59) = 207.5$, $Z =$
-3.582, $p < 0.01$), and this finding is confirmed by the PPVT™-4 Manual, since the difference between a mean of 29.14 (GSV 89) and a mean of 44.10 (GSV 101) is 12 points.

Group development

For the repeated-measures test, the Wilcoxon signed ranks test was used. For the native language-based group, the test did not reveal a significant difference between the pre- and post-tests, with $W(28) = 143.50$, $Z = -1.356$, $p = 0.0875$ (one-tailed). This finding is confirmed by GSV scores; the difference between mean pre- and post-test GSV scores is only five points, from 84 to 89 GSV points.

For the bilingually-based group, the median score was significantly different from the pre-test to the post-test with $W(30) = 19.50$, $Z = -4.479$, $p < 0.01$ (one-tailed). Again, the GSV scores confirm the finding, since the difference between mean pre-test (GSV 85) and post-test (GSV 101) score is 16 points, which, according to the PPVT™-4 Manual, is a significant difference (Dunn & Dunn, 2007b, p. 205). The effect size for the bilingually-based group was -0.8, indicating that the change in these pupils' average receptive vocabulary from the beginning to the end of the school year was not only significant but also substantial.

Successful L2 acquisition does not necessarily equal near-nativeness, but comparison to L1 acquisition may nevertheless be useful for purposes of illustration. One measure of the meaningfulness of the development in the bilingually-based group is illustrated in table 4, which summarizes the results and their age equivalents, as given by the PPVT™-4 manual. Thus, the native language-based group's (non-significant) mean increase in receptive vocabulary translates into an equivalent of only three months' development in native English children, from age 2;4 to age 2;7.

Table 4: Age equivalents of pre- and post-test vocabulary scores (raw) in the bilingually-based and the native language-based groups

	PPVT™-4, pre-test		PPVT™-4, post-test	
	mean score	L1 age equivalent	mean score	L1 age equivalent
NB	23.72	2;4	29.14	2;7
BB	25.39	2;5	44.10	3;3

BB= Bilingually-based
NB= Native language-based

The mean age equivalent of the bilingually-based group, however, has increased by ten months in the course of an average time span of 8 months. This means that these L2 learners have, on average, been acquiring new words at a slightly faster rate than the average for children at the same stage of language development, who are acquiring English as their L1. The main difference is that, while this development on average takes place between ages 2;5 and 3;3 in English-speaking children, it took place between mean ages 6;1 and 6;9 in these L2 learners. This is quite an astonishing development, considering that the input to which these children have been exposed is still very limited compared to that of children acquiring their native language. The results thus clearly indicate that there is no inherent problem in the early-start foreign language classroom per se preventing it from being successful, at least not with respect to vocabulary development.

The nature of group vocabulary differences

It is worth looking at group differences for cognates and non-cognates separately, since there may be differences in how the two categories are acquired. Because of the PPVT™-4 discontinuation rule, where for each set of 12 words, testing stops if the participant makes eight or more errors, a few children were tested only on the 12 first words of the test in each session. These 12 words which all participants encountered are listed in Table 5 below, followed by the percentages of children in each group who answered correctly for each word in the post-test, as well as the results of a Mann-Whitney U test comparing the two groups' responses for each word. Words that sound similar in the two languages (cognates) are given in bold.

Table 5: Percentages of correct answers and Mann-Whitney U, Z, and p for between-groups comparisons of number of correct answers for cognate and non-cognate words in the bilingually-based and the native language-based groups

Word	BB	NB	Mann-Whitney U	Z	Sig. (1-tailed)
cat	100%	100%	341	0,000	.500
apple	100%	93.1%	310	-1,695	.045
balloon	100%	89.7%	321	-,902	.184
hand	100%	93.1%	310	-1,695	.045
airplane	32.3%	24.1%	324	-,386	.350
bird	32.3%	27.6%	335	-,139	.445
tree	96.8%	44.8%	183,5	-3,516	.000
table	90.3%	10.3%	134,5	-4,304	.000
drinking	96.8%	62.1%	208	-3,311	.001
frog	61.3%	55.2%	313,5	-,576	.283
money	67.7%	44.8%	249,5	-1,924	.027
umbrella	29%	6.9%	273	-1,747	.041

BB= Bilingually-based
NB= Native language-based

The words which are successfully identified by virtually all children in both groups are *cat*, *apple*, *balloon*, and *hand*, all of which are phonologically similar to their Norwegian counterparts *katt*, *eple*, *ballong*, and *hand*. However, the bilingually-based group scores slightly higher also on these words; for *apple and hand*, the difference is significant. Furthermore, the words *tree* and *drinking* are recognized by virtually all the children in the bilingually-based group, and the difference between the two groups here is significant. These words also sound relatively similar to their Norwegian counterparts *tre* and *drikke*.

However, the children in the bilingually-based group outperform their native language-based group peers also on non-cognates. The percentage of children who correctly identify the words *airplane* and *bird*, whose Norwegian equivalents are *fly* and *fugl* respectively, is slightly higher in the bilingually-based group than in the native language-based group, although the difference is not significant, while the differences for the words *money* (Norwegian *penger*) and *umbrella* (Norwegian *paraply*) are significant, and the word *table* (Norwegian *bord*) is the one with by far

the biggest difference in scores. While only 10% of the children in the native language-based group correctly identify this word, it is successfully identified by 90% of the children in the bilingually-based group. Although the number of items is too low to draw firm conclusions, we have an indication that the advantage in the bilingually-based group holds both for cognates and non-cognates.

Discussion

We see from the above results that English teaching in the native language-based group has had no significant impact on English receptive vocabulary. In other words, the 20+ hours out of the 138 hours of compulsory English teaching for grades 1-4 which this school is spending in the first grade have not had any measurable effect. We interpret this to mean that the L2 input received through this method of English teaching does not reach the critical threshold needed by children at this age for vocabulary development to take place. Children in both groups had acquired some English vocabulary prior to starting school, possibly through various sources such as computer games, music, TV, and movies. However, this vocabulary was very small for both groups, and included a number of cognates with Norwegian; word learning may have been incidental. The native language-based group's lack of English vocabulary development in the course of 8 months indicates that English exposure outside of school for young children in Norway is not sufficient for systematic acquisition. This further supports Murphy's (2010) argument that spending more time on the L2 in the classroom is especially important for learners who do not have extensive exposure to the target language outside of school.

We have also seen that the advantage in the bilingually-based group holds both for cognates and for non-cognates. Cognate and non-cognate acquisition may be slightly different processes in SLA. For example, Tonzar et al. (2009) show that cognates are acquired more easily than non-cognates both for English and German in Italian learners. Gascoigne (2001) proposes that cognates are in fact retrieved differently from other words in the L2, and that the representations in the two languages in the mental lexicon are partly overlapping. Aukrust (2007) argues that observed differences in whether vocabulary size in the two languages is correlated in bilingual children may be the result of how closely related the languages in question are, and consequently how many cognates there are that can be more or less transferred from one language to the other. Norwegian and English are both Germanic languages, and thus relatively closely related.

Although a great portion of the English vocabulary is of Romance origin, basic words are more often Germanic, and a possible hypothesis could have been that the improved performance of the bilingually-based group is mainly a result of cognate comprehension. However, we saw in table 5 that the children in the bilingually-based group seem to outperform the native language-based group both on words which are cognates in Norwegian and English and on words which are not.

Another question is what the exact problem is for acquisition in the native language-based group. There are two alternative explanations for their lack of measurable development on the PPVT™-4 in the course of 8 months. The first possible explanation is that the vocabulary items tested were not frequent enough in this group's input, while the second is that the words tested in the PPVT™-4 were not present in the input at all. The difference is in whether the input in the native language-based group is best described as generally impoverished, or whether it is just naturally more specialized due to being more limited. The early words in the PPVT™-4 are those expected to be familiar to very young American children, and it is not obvious that these are the same as those emphasized in early Norwegian English classrooms. Since the more limited input of the native language-based group necessarily includes a smaller number of words, it is possible that the children in this group have not happened to come across many of the words of the PPVT™-4, even if they may have acquired other words. This could give them an unfair disadvantage in the test. However, a look at the teaching materials and the activities reported for the native language-based group indicates that the vocabulary in the PPVT™-4 has been used in the classroom. Out of the (very few) words to be found in the native language-based group's work book *Junior Scoop 1-2*, several can be found in the early sets of the PPVT™-4, such as *tree, bird,* and *balloon*. Another area of vocabulary which teachers in this group specifically mentioned practicing was body parts, an area also present early on in the PPVT™-4 in words such as *hand* and *neck*. It is obviously impossible to establish whether the children have encountered all the words tested early in the PPVT™-4. Still, there is no reason to believe that the words in the PPVT™-4 are thematically different from those used in the native language-based classroom, and that this has created an unfair advantage for the bilingually-based group in the test.

Since the aim of the present study was to investigate the effect of an increase in English input which actually felt manageable and natural to the teachers in the bilingually-based classroom, the

researcher did not interfere with what happened in the classroom, and the structure and nature of input was therefore not carefully controlled. This lack of control potentially raises questions about whether it is the exposure per se or specific characteristics of it which have brought about the effect. For example, one of the teachers in the bilingually-based school was in fact a native speaker of English, although she was also completely fluent in Norwegian and taught all other subjects in this language. It is of course conceivable that this teacher's nativeness in itself is what led to increased acquisition in the bilingually-based group. However, this would mean assuming that native speakers are always better teachers of L2s or that language can only successfully be acquired from native-speaker input, which goes against research findings both on L1 development (e.g., J. L. Singleton & Newport, 2004) and on second/foreign language teaching (cf. Moussu & Llurda, 2008). The main benefits of the teacher's nativeness, i.e., language proficiency and the confidence to use English extensively, can both be trained also in non-native teachers. It is precisely the conclusion of this paper that teachers should be trained in this.

Another question is whether the native language-based group really is representative of normal Norwegian schools, or whether the lack of acquisition in this group is a result of "poor" teaching. However, as already mentioned, pupils from this school have been previously found to perform above average in national tests in English. Whereas results in these tests may have come from classes taught by teachers other than those in the present study, it is highly unlikely that the school standards of English instruction have dramatically dropped. Furthermore, as with the bilingually-based group, the native language-based teachers knew that their pupils would be tested after 8 months, and were naturally eager for them to do well. If anything, it is likely that they spent more time on English than they would have in a normal year. Finally, and significantly, there is nothing in what the native language-based teachers report that deviates from the stated norms of the curriculum. As with many early-start foreign language programs, nothing in the plans for early English teaching in Norway focuses on extensive input for vocabulary acquisition.

Conclusion and suggestions for further research

The overall conclusion of the present study is that there is nothing inherent in the classroom situation which prevents successful L2 acquisition in young learners, and that vocabulary can be

acquired at a fast rate in an early-start foreign language program. Furthermore, the study indicates that although such acquisition critically depends on input, exposure to the target language need not be unrealistically massive for acquisition to take place.

The PPVT™-4 only investigates receptive vocabulary, and tells us nothing about the productive vocabulary of the children in the study. However, we do know that the two are related, and that receptive vocabulary is important for comprehension, which in turn means that a larger receptive vocabulary allows more advanced input to be processed and understood. In this sense, receptive vocabulary can be assumed to be a predictor for further language acquisition.

A natural next step is to further examine whether such an increase in exposure to the target language has a long-term effect beyond the first year of school, and whether it is also evident in areas other than vocabulary comprehension. Furthermore, more research is needed concerning exactly what kind of input is necessary, including what proficiency level teachers must have attained and whether native input from sources other than the teacher, especially media (i.e., audio and video) can fruitfully be exploited to increase input in early-start second language classrooms.

Conflict of Interest Statement:
The research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Authors and Contributors:
AD has had main responsibility for the project, including research design, data collection, analysis and interpretation, and drafting and revising the paper. MDV has contributed substantially to the conception and design of the research, and to critical revision of the paper for important intellectual content. Both authors have final approvement of the version to be published and agree to be accountable for all aspects of the work.

Acknowledgement:
We acknowledge the support of the Faculty of Humanities, Norwegian University of Science and Technology, from whom the first author received a grant to conduct this research.

References

Abello-Contesse, C., Chacón-Beltrán, R., López-Jiménez, M. D., & Torreblancaa-López, M. M. (2006). Introduction and overview. In C. Abello-Contesse, R. Chacón-Beltrán, M. D. López-Jiménez & M. M. Torreblancaa-López (Eds.), *Age in L2 acquisition and teaching* (pp. 7-27). Bern: Peter Lang.

Aukrust, V. G. (2007). Young children acquiring second language vocabulary in preschool group-time: Does amount, diversity, and discourse complexity of teacher talk matter? *Journal of Research in Childhood Education, 22*(1), 17-37. doi: 10.1080/02568540709594610

Berman, R. A. (2007). Developing linguistic knowledge and language use across adolescence. In E. Hoff & M. Shatz (Eds.), *Blackwell handbook of language development* (pp. 347-367). Malden, MA: Blackwell.

Bley-Vroman, R. (1989). What is the logical problem of foreign language learning? In S. Gass & J. Schachter (Eds.), *Linguistic Perspectives on Second Language Acquisition* (pp. 41-68). Cambridge, MA: Cambridge University Press.

Bloom, P. (2000). *How children learn the meanings of words*. Cambridge, MA: MIT Press.

Bloom, P. (2004). Myths of Word Learning. In D. G. Hall & S. R. Waxman (Eds.), *Weaving a lexicon* (pp. 205-224). Cambridge, MA: MIT Press.

Bruskeland, P. A., & Ranke, C. T. (2005). *Junior Scoop 1-2* (Vol. 1-2). Oslo: Samlaget.

Burstall, C. (1975). French in the primary school: The British experiment. *Canadian Modern Language Review, 31*, 388-402.

Cenoz, J. (2003). The influence of age on the acquisition of English: General proficiency, attitudes and code-mixing. In M. García Mayo & M. García Lecumberri (Eds.), *Age and the acquisition of English as a foreign language* (pp. 77-93). Clevedon: Multilingual Matters.

Childers, J. B., & Tomasello, M. (2002). Two-year-olds learn novel nouns, verbs, and conventional actions from massed or distributed exposures. *Developmental Psychology, 36*(6), 967-978. doi: 10.1037/0012-1649.38.6.967

Clark, E. V. (1993). *The lexicon in acquisition*. Cambridge: Cambridge University Press.

Colledge, E., Bishop, D. V., Koeppen-Schomerus, G., Price, T. S., Happé, F. G., Eley, T. C., . . . Plomin, R. (2002). The structure of language abilities at 4 years: A twin study. *Developmental Psychology, 38*(5), 749-757. doi: 10.1037//0012-1649.38.5.749

Dale, P. S., Harlaar, N., Haworth, C. M., & Plomin, R. (2010). Two by two: A twin study of second-language acquisition. *Psychological Science, 21*(5), 635-640. doi: 10.1177/0956797610368060

DeKeyser, R. M. (2000). The robustness of Critical Period effects in second language acquisition. *Studies in Second Language Acquisition, 22*(4), 499-533.

DeKeyser, R. M., & Larson-Hall, J. (2005). What does the critical period really mean? In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism: psycholinguistic approaches* (pp. 88-108). Oxford: Oxford University Press.

Dunn, L. M., & Dunn, D. M. (2007a). Peabody Picture Vocabulary Test. Minneapolis, MN: Pearson Education, Inc.

Dunn, L. M., & Dunn, D. M. (2007b). *Peabody Picture Vocabulary Test Manual* (4 ed.). Minneapolis, MN: Pearson Education, Inc.

Felix, S. W. (1985). More evidence on competing cognitive systems. *Second Language Research, 1*(1), 47-72.

Foyn, C., Vulchanova, M., Nilsen, R. A., & Sigmundsson, H. (under revision). Links between phonological memory, first language competence and second language competence in 10 year-old children. *Learning and Individual Differences*.

García Lecumberri, M. L., & Gallardo, F. (2003). English FL sounds in school learners of different ages. In M. d. P. García Mayo & M. L. García Lecumberri (Eds.), *Age and the acquisition of English as a foreign language* (pp. 115-135). Clevedon: Multilingual Matters.

García Mayo, M. d. P. (2003). Age, length of exposure and grammaticality judgements in the acquisition of English as a foreign language. In M. L. García Lecumberri & M. d. P. García Mayo (Eds.), *Age and the acquisition of English as a foreign language* (pp. 94-114). Clevedon: Multilingual Matters.

Gascoigne, C. (2001). Lexical and conceptual representations in more- and less-skilled bilinguals: The role of cognates. *Foreign Language Annals, 34*(5), 446-452. doi: 10.1111/j.1944-9720.2001.tb02084.x

Gass, S. M. (2003). Input and interaction. In C. J. Doughty & M. H. Long (Eds.), *The Handbook of Second Language Acquisition* (pp. 224-255). Oxford: Blackwell.

Gathercole, S. E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics, 27*(4), 513-613. doi: 10.1017/S0142716406060383

Goldschneider, J. M., & DeKeyser, R. M. (2001). Explaining the "Natural Order of L2 Morpheme Acquisition" in English: A meta-analysis of multiple determinants. *Language Learning, 51*(1), 1-50. doi: 10.1111/1467-9922.00147

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

Hayiou-Thomas, M. E., Dale, P. S., & Plomin, R. (2012). The etiology of variation in language skills changes with development: A longitudinal twin study of language from 2 to 12 years. *Developmental Science, 15*(2), 233-249. doi: 10.1111/j.1467-7687.2011.01119.x

Hoff, E., & Naigles, L. (2002). How children use input to acquire a lexicon. *Child Development*(73), 418-433. doi: 10.1111/1467-8624.00415

Holmstrand, L. S. E. (1982). *English in the elementary school: Theoretical and empirical aspects of the early teaching of English as a foreign language*. Uppsala: Acta Universitatis Upsaliensis.

Hyltenstam, K. (1992). Non-native features of near-native speakers: On the ultimate attainment of childhood L2 learners. In R. J. Harris (Ed.), *Cognitive Processing in Bilinguals* (pp. 351-368). Amsterdam: Elsevier Science.

Hyltenstam, K., & Abrahamsson, N. (2003). Maturational constraints in SLA. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 539-588). Oxford: Blackwell.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology, 21*(1), 60-99.

Kersten, S. (2010). *The mental lexicon and vocabulary learning: Implications for the foreign language classroom*. Tübingen: Narr.

Larsen-Freeman, D. (1975). The acquisition of grammatical morphemes by adult ESL students. *TESOL Quarterly, 9*, 409-419.

Larson-Hall, J. (2008). Weighing the benefits of studying a foreign language at a younger starting age in a minimal input situation. *Second language research, 24*(1), 35-63. doi: 10.1177/0267658307082981

Lasagabaster, D., & Doiz, A. (2003). Maturational constraints on foreign-language written production. In M. L. García Lecumberri & M. d. P. García Mayo (Eds.), *Age and the acquisition of English as a foreign language* (pp. 136-160). Clevedon: Multilingual Matters.

Lightbown, P. M. (2000). Classroom SLA research and second language teaching. *Applied Linguistics, 21*(4), 431-462. doi: 10.1093/applin/21.4.431

Lightbown, P. M., Halter, R., White, J., & Horst, M. (2002). Comprehension-based learning: The limits of "Do It Yourself". *The Canadian modern language review, 58*(3), 427-464. doi: 10.3138/cmlr.58.3.427

MacWhinney, B. (2005). A unified model of language acquisition. In A. M. B. de Groot (Ed.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 49-67). Cary, NC: Oxford University Press.

Moussu, L., & Llurda, E. (2008). Non-native English-speaking English language teachers: History and research. *Language Teaching, 41*(3), 315-348. doi: 10.1017/S0261444808005028

Muñoz, C. (2001). Factores escolares e individuales en el aprendizaje formal de un idioma extranjero In S. P. Cesteros & V. S. García (Eds.), *Estudios de Lingüística. Anexo 1: Tendencias y Líneas de Investigación en Adquisición de Segundas Lenguas* (pp. 249–270). Alicante: Universidad de Alicante.

Muñoz, C. (2006). The BAF Project: Research on the effects of age on foreign language acquisition. In C. Abello-Contesse, R. Chacón-Beltrán, M. D. López-Jiménez & M. M. Torreblancaa-López (Eds.), *Age in L2 acquisition and teaching* (pp. 81-92). Bern: Peter Lang.

Murphy, V. A. (2010). The relationship between age of learning and type of linguistic exposure in children learning a second language. In E. Macaro (Ed.), *Continuum companion to second language acquisition* (pp. 158-178). London/New York, NY: Continuum International Publishing Group.

Nagy, W. E., & Herman, P. A. (1987). Breadth and depth of vocabulary knowledge: Implications for acquisition and instruction. In M. G. McKeown & M. E. Curtis (Eds.), *The nature of vocabulary acquisition* (pp. 19-36). Hillsdale, NJ: Erlbaum.

Newport, E. L. E. (1990). Maturational Constraints on Language Learning. *Cognitive science, 14*(1), 11-28.

Nikolov, M. (2009). The age factor in context. In M. Nikolov (Ed.), *The age factor and early language learning* (pp. 1-38). Berlin: Mouton de Gruyter.

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Statistical Learning in a Natural Language by 8-Month-Old Infants. *Child Development, 80*(3), 674-685.

Pinker, S. (1994). *The language instinct: How the mind creates language*. New York, NY: Harper Perennial Modern Classics.

Ruiz-González, G. (2006). Age effects on single phoneme perception. In C. Abello-Contesse, R. Chacón-Beltrán, M. D. López-Jiménez & M. M. Torreblancaa-López (Eds.), *Age in L2 acquisition and teaching* (pp. 155-173). Bern: Peter Lang.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language, 35*(4), 606-621. doi: 10.1006/jmla.1996.0032

Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental language learning: listening (and learning) out of the corner of your ear. *Psychological Science, 8*(2), 101-105. doi: 10.1111/j.1467-9280.1997.tb00690.x

Shintani, N. (2011). A comparative study of the effects of input-based and production-based instruction on vocabulary acquisition by young EFL learners. *Language Teaching Research, 15*(2), 137-158. doi: 10.1177/1362168810388692

Singleton, D. (1999). *Exploring the second language mental lexicon*. Cambridge: Cambridge University Press.

Singleton, D., & Ryan, L. (2004). *Language acquisition: The age factor*. Clevedon: Multilingual Matters.

Singleton, J. L., & Newport, E. L. (2004). When learners surpass their models: The acquisition of American Sign Language from inconsistent input. *Cognitive Psychology, 49*(4), 370–407. doi: 10.1016/j.cogpsych.2004.05.001

Sparks, R. L., Patton, J., Ganschow, L., & Humbach, N. (2009). Long-term relationships among early first language skills, second language aptitude, second language affect, and later second language proficiency. *Applied Psycholinguistics, 30*(4), 725-755. doi: 10.1017/S0142716409990099

Stern, H. H. (1983). *Fundamental concepts of language teaching*. Oxford: Oxford University Press.

Tonzar, C., Lotto, L., & Job, R. (2009). L2 vocabulary acquisition in children: Effects of learning method and cognate status. *Language Learning, 59*(3), 623-646. doi: 10.1111/j.1467-9922.2009.00519.x

Trønder-Avisa. (2007, March 15 2012). Bekymret for engelskfaget. *Trønder-Avisa*. Retrieved 11/11, 2013, from hhttp://www.t-a.no/nyheter/article176154.ece#.UoEQ2OKmb30

Utdanningsdirektoratet. (2006). Knowledge Promotion: Curriculum in English: Utdanningsdirektoratet, www.udir.no.

Utdanningsdirektoratet. (2007). Ansvaret for fordeling av skoledager utover året: Utdanningsdirektoratet, www.udir.no.

Vulchanova, M., Vulchanov, V., Sarzhanova, D., & Eshuis, H. (2012). The role of input in early bilingual lexical development. *Lingue e linguaggio*(2), 181-198. doi: 10.1418/38785

Wode, H. (1981). *Learning a second language: An integrated view of language acquisition*. Tuebingen: Narr.

Paper II

Sentence repetition as a measure of early L2 competence

Anne Dahl

Abstract

Childhood bilingualism and second language (L2) learning are increasingly common phenomena, creating an increasing need for valid and accurate tests of L2 competence in young learners. This paper evaluates sentence repetition as a measure of the acquisition of English as an L2 in young learners in Norway. The evaluation addresses three questions, namely whether sentence repetition is a reliable and valid measure of L2 competence in early learners, what the effect is of sentence length in such a test, and what specific language competencies can be tested through sentence repetition. Conclusions are that sentence repetition does indeed test L2 competence, but that sentence length and competence levels determine which specific skills are measured. It is argued that sentence repetition involves both phoneme perception and production as well as lexical knowledge and grammatical processing.

1. Introduction

Multilingualism is a complex phenomenon, and language competence in multilingual children may easily be misestimated if assessment is inaccurate, performed by non-experts, or if inadequate test materials are used that lack validity or reliability. It is thus important to develop accurate tools for assessing L2 competence in different populations of children in schools, e.g., those whose home language differs from the language of instruction, or early foreign language learners. Testing very young learners is a challenge, partly because of the academic nature of many traditional language tests, i.e., the involvement of literacy, and a focus on explicit rather than implicit knowledge (cf. Erlam, 2006). Young L2 learners may depend more on implicit and less on explicit knowledge than do older learners (Muñoz, 2006). Comprehension tests are commonly used, but designing such tests to validly and reliably test language beyond lexical knowledge can be difficult, because much comprehension may be based on semantic information and context (Ferreira, Bailey, & Ferraro, 2002; Izumi, 2003; VanPatten, 2012). Oral production tests, in contrast, cannot adequately control what is being tested. Such tests may be difficult to interpret because they allow for avoidance strategies, because it is hard to anticipate what

structures may be elicited, and because individual differences such as shyness may be confounding factors (Erlam, 2006; Van Moere, 2012).

Sentence repetition avoids many of these problems, as it is not confounded by communication strategies and thus creates less measurement noise than communicative production tests, while still testing implicit language knowledge (Erlam, 2006; Van Moere, 2012).

2. Sentence repetition as a measure of language competence

Repetition has been widely used as a language test in a variety of contexts. For example, non-word repetition has been used since the 1990s to measure phonological working memory and language aptitude (cf. Roy & Chiat, 2004). Sentence repetition has also been used for assessing early first language (L1) grammatical development (e.g., Brown & Fraser, 1963; Devescovi & Caselli, 2007), grammatical processing in atypical populations (e.g., Berry, 1976; Rondal, Lambert, & Sohier, 1981; Small, Kemper, & Lyons, 2000), and for testing pronunciation in foreign languages, e.g., in atypical children (Fabbro, Alberti, Gagliardi, & Borgatti, 2002).

In second language acquisition (SLA), sentence repetition has often been used to test precise grammatical factors (Erlam, 2006). Whether it is a valid measure of L2 competence has been debated, the argument being that at least with sentences short enough to be remembered as "acoustic images", it provides a measure of repetition ability rather than of language competence. Therefore, if subjects are merely "parroting" what they hear, they may not actually be processing and reproducing language (Jessop, Suzuki, & Tomita, 2007; Vinther, 2002). However, a number of studies have found convergence between sentence repetition and other tests of L2 competence (Flynn, 1986; Munnich, Flynn, & Martohardjono, 1994; Perkins, Brutten, & Angelis, 1986; Van Moere, 2012).

There is no consensus as to how long a sentence must be to avoid simple parroting, but a limit of approximately eight syllables for rote repetition has been suggested (Bley-Vroman & Chaudron, 1994; Erlam, 2006; Fujiki & Brinton, 1983; Perkins et al., 1986; Smith, 1970). Alternatively, there may not be an absolute threshold for sentence length across different competence levels (Bley-Vroman & Chaudron, 1994). For example, it is possible that sentence length is connected to time, not only to the number of elements, and one factor which may influence the appropriate sentence length is the learner's speed of processing. Thus, sentence

repetition may be a measure of automatization (Ellis, 1994; Hulstijn & De Graaff, 1994; Van Moere, 2012). Working memory is used to maintain information for further processing by constantly refreshing it through (overt or silent) articulation in the phonological loop, of which the capacity is approximately three seconds' worth of material (e.g., Baddeley, Thomson, & Buchanan, 1975; Van Moere, 2012). This means that not only processing speed but also speed of (re)production will be relevant, and faster speakers may be able to repeat more items than slower speakers. Furthermore, repetition may depend on the quality of lexical representations and how they are accessed, e.g., the extent to which L2 words are associated with their L1 translation equivalents, which may influence speed of retrieval (cf. de Groot, 2002; Kroll & Tokowicz, 2005; Pavlenko, 2009). Access, and thus reproduction, may also be slowed down by phonologically "fuzzy" or imprecise lexical representations. Such incomplete representations may be typical in individuals with language and reading disorders, but may also be relevant for processing in L2 (cf. Bishop & Snowling, 2004; Brady, 1997; Darcy, Daidone, & Kojima, 2013; Kamhi, 2008; Li & Zhao, 2014; Sebastian-Gallés, Rodríguez-Fornells, Diego-Balaguer, & Díaz, 2006). It follows that the appropriate sentence length may vary with proficiency level (Vinther, 2002), since representations, automatization, and speed of articulation develop as language is acquired.

Appropriately designed sentence repetition tests clearly require more than just rote repetition. Still, it is not obvious which language abilities are in fact involved, and errors in repetition can in principle stem from problems in either comprehension or production (Van Moere, 2012). The relationship between performance on sentence repetition and spontaneous production is not clear; while Larsen-Freeman and Long (1991) argue that there such a relationship, other studies have found that sentence repetition does not predict spontaneous speech (Connell & Myles-Zitzer, 1982). Repetition may be a test of *performance* (Munnich et al., 1994), but it clearly does not simply test natural production abilities (cf. Bley-Vroman & Chaudron, 1994; Vinther, 2002).

Not all aspects involved in free production are equally relevant in sentence repetition. If we take Levelt's (1989) model of speech production as a starting point, we can assume fundamentally different involvement of the Conceptualizer in sentence repetition compared to normal speech production. Free production entails conceiving an intention, and selecting and ordering information for expressing the intention. In contrast, for repetition, the intention and the

message come from outside and must be reconstructed rather than constructed. The Formulator in Levelt's model requires the same information for repetition as for free production, but the way in which this information is accessed is presumably different in repetition. For example, although lemmas must be accessed in the mental lexicon, they will have been primed by the linguistic stimuli, which likely facilitates the process of lexical retrieval. It is thus possible that lexical items which are best described as receptive or part of "passive vocabulary" (Nation, 2001) are available for production in sentence repetition. Similarly, the grammar of the utterance, provided that it has been processed and comprehended, will have been primed, and may therefore be more readily available for production in sentence repetition than in normal production (e.g., Pickering & Ferreira, 2008). In this respect, sentence repetition is similar to other elicited production tasks, e.g., the use of model sentences where participants are asked to produce analogous but not identical sentences (cf. Helland, 2008; Helland & Kaasa, 2005; Thornton, 1996). Again, performance on a sentence repetition task is closely connected to comprehension in this respect. Finally, the Articulator in Levelt's model can be argued to be where free production and repetition are the most similar, since at the actual output stage, sentence repetition as well as normal production entails the translation of the verbal message into the physical articulation of the utterance. However, there may be priming effects also here (cf. Trofimovich, 2005).

Sentence repetition can thus be assumed to involve cognitive processes such as processing the sentence, reconstructing it based on an individual's grammar, and reproducing it (e.g., Erlam, 2006; Jessop et al., 2007; Munnich et al., 1994; Vinther, 2002). It is likely, therefore, that repetition is a conservative measure of comprehension, i.e., that L2 learners may comprehend more than they can repeat. On the other hand, if they are able to repeat a sentence, we have strong evidence that it has also been comprehended (Naiman, 1974; Vinther, 2002). Thus, we can assume that sentence repetition is a measure of grammatical competence, and the finding that lexical words are imitated with greater accuracy than grammatical morphemes (Connell & Myles-Zitzer, 1982) indicates that it at least taps sensitivity to morphology.

The reconstructive nature of sentence repetition is also evidenced by the fact that subjects will often spontaneously correct ungrammatical sentences in repetition tasks (e.g., Hamayan, Saegert, & Larudee, 1977; Keeney & Wolfe, 1972; Markman, Spilka, & Tucker, 1975; Munnich et al., 1994). However, Bley-Vroman and Chaudron (1994) argue that there is a "narrow band of sensitivity" for sentence length in repetition tests, particularly between short sentences for which

high accuracy is expected, and very long sentences, where accuracy is expected to be low. Furthermore, they argue that such tests are relatively insensitive to high-level syntax and semantics, for instance binding relationships and quantifier scope. This may make sentence repetition particularly well suited to testing early L2 learners.

A final concern when it comes to repetition tests is that of uniformity of stimuli. The difficulty levels of items may vary if spoken by different people, and pre-recorded stimuli may be the best way to ensure uniformity (Eisenbeiss, 2010; Van Moere, 2012). However, Munnich et al. (1994) empirically tested both oral and taped stimuli in a sentence repetition task, and found no significant differences between the two.

3. The study

The present study assesses the use of a sentence repetition test in young, early L2 learners of English in Norwegian schools, both in an immersion setting and in traditional foreign language classrooms. Three main research questions are posed, namely: 1) Does such a test discriminate between competence levels even at very low levels?, 2) What is the effect of sentence length for performance on such a test?, and 3) Which specific language competences are tested in sentence repetition? Since each research question requires different analyses of the data, each question is discussed separately immediately following the relevant results, while the overall conclusion sums up the main findings from all three questions.

3.1. Participants

The participant sample comprised a total of 82 children in Norwegian first grade (or equivalent[1], mean age 6;9), with four different levels of English competence: A control group (CG) of 15 native (bilingual or monolingual) speakers of English, all attending international schools in Norway; an immersion group (IG) of 7 non-native speakers of English also attending international schools; a high-input group (HG) of 31 native Norwegian speakers in a Norwegian state school who had attended English classes for a year with a specific focus on L2 input; and finally a low-input group (LG) of 29 native Norwegian speakers in a Norwegian state school who had received normal English instruction for a year with no specific focus on input. All

[1] Some participants attended international schools with grade systems different from the standard Norwegian system. However, all children in this study were of the same age, i.e., turned six in the calendar year when they had started their current grade. For all L2 learners, it was their first year of school. Some children in the English native-speaker control group were in their second year of school.

participants were recruited through their schools, and all parents provided written consent for their child to participate in the study.

3.2. Control measures

A battery of control measures were administered in addition to the repetition test. All groups except the CG were tested on a number of control measures. Visio-spatial working memory was tested using a memory game, where participants were asked to memorize sets of picture cards, starting with a set of two pairs and incrementally increased until the final set of six pairs. The participant was given a fixed time to memorize the pattern of the cards before they were turned upside down and the child was asked to find the pairs in as few attempts as possible. In addition, the three L2 groups completed the Matrices subscale of the Kaufmann Brief Intelligence Test, second edition (K-Bit 2, Kaufman & Kaufman, 2004a) as a measure of non-verbal intelligence, and a measure of verbal ability (verbal comprehension, reasoning, and vocabulary knowledge, cf. Kaufman & Kaufman, 2004b) through the Riddles section of the K-Bit 2 translated into Norwegian. Norwegian vocabulary was tested using a translated version of the Peabody Picture Vocabulary Test, Fourth Edition (PPVT-IV, Dunn & Dunn, 2007a), form A.

All groups but the CG were tested for English receptive vocabulary with the PPVT-IV, forms A and B[2]. Furthermore, all four groups were tested for English sentence comprehension using a computerized test. Here, the child was asked to select from a set of four pictures the one matching the stimulus for each of 30 sentences. Sentences were read alternately by a female native speaker of British English and one of American English.

3.2.1. Results of control measures

Results from the non-English measures for all three non-native groups were compared to rule out extralinguistic differences. One-way ANOVAs[3] found no significant difference between groups for visio-spatial memory [$F(2, 64)=.76$, $p=.471$], for non-verbal intelligence [$F(2, 64)=.34$,

[2] The Norwegian version of the PPVT-IV form A was administered at the beginning of the school year, approximately eight months prior to the administration of the repetition test. Thus, there should be no effect of language of instruction on this test between the IG, attending an English-medium school, on the one hand, and the HG and the LG, attending Norwegian-medium schools, on the other. All other measures were administered at the same time as the sentence repetition test, at the end of the school year. This means that eight months had passed between the administration of the Norwegian and the English versions of the PPVT-IV form A.

[3] Because of the small size of the IG (n=7), analyses using parametric tests reported for this group were checked with corresponding non-parametric tests whenever possible. These tests confirmed the findings of the parametric tests, and are only reported in a few cases where they are particularly relevant.

p=.710], nor for verbal ability [F(2, 64)=1.94, p=.152]. For Norwegian vocabulary, however, a significant difference was found between groups [F(2, 64)=6.27, p=.003, r_{effect}=.40]. Post hoc tests with Hochberg's GT2 procedure showed that the difference was between the IG and the two other groups (HG and LG). This effect is probably best explained by the fact that all children in the HG and the LG were monolingual prior to English exposure, while most children in the IG were multilingual; a slightly smaller vocabulary in Norwegian may not be unexpected for a multilingual child (cf. Bialystok, 2009). The conclusion is that for the three groups of non-native English speakers, there is no reason to expect any important cognitive between-group differences which could cause different performance on the sentence repetition test.

To establish that the four groups of participants represent four English competence levels, the English measures were compared across groups. English vocabulary was not tested for the CG, because the PPVT-IV has been normed for native speakers (Dunn & Dunn, 2007b). For the three other groups, the combined score across both forms is used as the vocabulary score.

Through norming within the PPVT-IV, the significance of the difference between scores can be estimated, allowing us to evaluate the score of the IG against the test's reference group, matched for the age of the CG in the present study to which the other English test scores are compared. The expected mean raw score of 109 for the reference group translates into a growth scale value (GSV) of 147, while the mean score of 80 for the IG translates into a GSV of 128, i.e., a difference of 19 points. According to the PPVT-IV manual (Dunn & Dunn, 2007a), a GSV difference of 8 points represents a significant difference at this age. It should be noted that this estimation is intended for within-subject changes in scores over time, not for comparing different groups. Still, with a difference in scores more than twice what represents a significant change in scores, we can fairly safely conclude that performance in the IG was significantly below what would be expected for native speakers, although we do not know how the specific CG in the present study would compare. Next, a one-way ANOVA with planned comparisons was conducted for the results of the three non-native groups, and results show an overall effect of group [F(2, 64)= 26.7, p<.01, r_{effect}=.67]. Planned comparisons found that the IG scored significantly higher than the HG [t(64)= 5.74 p<.01, r_{effect}=.58], who in turn scored significantly higher than the LG [t(64)=2.6, p=.006, r_{effect}=.31].

On the sentence comprehension test, a one-way ANOVA with planned comparisons was conducted for all four groups. In this analysis, Levene's test for equality of

variance was significant, probably due to ceiling effects in the CG and the IG, whose means were both close to the maximum possible score. Thus, values not assuming equality of variance are reported. There was an overall effect of group [Welch's F(3, 33.0)= 157.089, p<.01, r_{effect}=.86]. For the planned comparisons, the CG and the IG were significantly different[4] [t(13.5)=2.6, p=.011, r_{effect}=.58], as were the IG and the HG [t(35.2)=13.04, p<.01, r_{effect}=.91]. Finally, the HG and the LG were compared, and again the difference was significant [t(53.8)=4.2, p<.01, r_{effect}=.50].

We thus seem to have robust evidence that the four groups in question did in fact represent four different English competence levels, although we do not know what scores the CG would have obtained on the PPVT-IV. We do, however, know that all L2 groups scored below what we would expect for native speakers on this measure. This is important for the interpretation of the results of the sentence repetition test.

3.3. Repetition test

The repetition test (Appendix) consisted of 17 sentences of various length and complexity, ranging from 2 to 11 words, and from 2 to 13 syllables. Each sentence was read aloud by the researcher (a proficient L2 speaker of American English), who instructed the child to "repeat after me" and to "mimic as closely as you can what I am saying". All instruction and encouragement was given in English for the CG and in Norwegian for the three other groups. Before the test started, a warm-up session with single words and three sentences made sure that the child understood what to do. Recording started before the warm-up session, so that by the time the actual test started, children were less aware of the recorder.

During the test session, the child was given as long as he or she needed to repeat the sentence. If needed, the researcher repeated the sentence once. If the child still failed to respond, the sentence was repeated one more time, and the child was also encouraged, for example, by being told that "it is all right if you only remember one or two words, or maybe even some of the sounds that I made – just try to repeat what you can remember". If the child still did not respond, the item was skipped and the next sentence was read. The entire session was recorded on an Olympus Digital Voice Recorder VN-2100PC.

[4] This finding is surprising given the similarity of the group means and the assumed ceiling effects in both groups. Because of the small sample size particularly in the IG, a non-parametric Mann-Whitney U test was run, which also found the difference between the CG and the IG to be significant [U (21) = 21, Z = -2.3, p = .049].

Test time, including instructions, was normally less than five minutes, and most children were able to complete the test without major encouragement[5]. A total of thirteen children failed to respond to one or more sentences even after repetition and encouragement. All of these children were in the two lowest competence groups, and eight out of 12 were in the LG, indicating that such inability to respond was a reflection of low L2 competence rather than of problems mimicking. The problematic sentences were the longest ones; with the exception of one child missing sentence 8 (four syllables), all non-responses occurred with sentences eight syllables or over; once for sentence 5 (eight syllables), five times for sentence 15 (10 syllables), and four times for sentence 16 (13 syllables).

Two independent raters, both university students of English linguistics, assessed the recordings. One rater was a native speaker of American English; the other was a bilingual native speaker of Norwegian and Dutch and a proficient L2 speaker of American English. The raters were trained on the scoring procedure, which awarded one point per morpheme produced, including points for certain zero morphemes, yielding a maximum total score of 142 (see Appendix for details). The raters were told specifically to disregard pronunciation as long as errors or non-native pronunciations did not hinder comprehension. The raters received no information about the participants' expected language levels nor about the four different competence groups. All repetitions, instructions and encouragement had been edited out of the recording before the raters heard it, and each participant's recording was given a random code unrelated to competence group.

4. Results and discussion

4.1. Does the test discriminate between competence levels?

The first research question was whether sentence repetition can be used to discriminate between language competence levels in populations such as those included here. In order to evaluate this, the test's reliability and validity as a language measure were assessed.

[5] In one case, the researcher failed to administer a sentence to a participant in the CG. This was sentence 9, on which all other participants in the CG and the IG obtained perfect scores. The participant in question scored normally for the CG on all other sentences, and there was thus no reason to assume she would not have received a full score on this sentence had it been administered. She was therefore awarded the full score of 7 for this sentence.

4.1.1. Results

Internal consistency was measured at .896 with Cronbach's Alpha, indicating that reliability was good. No individual item seems to have heavily biased the test; Cronbach's Alpha If Item Deleted ranged from .876 to .898. However, the value of Cronbach's Alpha did vary between groups; these figures are reported in Table 1.

Table 1: Cronbach's Alpha overall and for each individual participant group

Group	Cronbach's Alpha	Cronbach's Alpha If Item Deleted (minimum)	Cronbach's Alpha If Item Deleted (maximum)
Overall	.896	.876	.898
CG	.412	.316	.461
IG	.780	.723	.799
HG	.786	.756	.790
LG	.883	.865	.887

Table 1 shows good internal consistency in all groups except the CG. No individual item stands out as very different from the others, since deleting one item would not significantly change the value of Cronbach's Alpha. The fact that Cronbach's Alpha is relatively low for the CG is not surprising. After all, all normally developing children at the age of six should be able to repeat simple sentences in their native language, and ceiling effects are thus to be expected in this group, so that the occasional loss of a point on an item is non-systematic. This is also evident in Table 2 below, which shows very little variance in the CG.

Also the reliability of the scoring procedure was checked. Interrater reliability for the two raters was assessed using Intraclass Correlation Coefficient (ICC). For all groups taken together, ICC was .953, which is considered very good. Reliability varied, but was high across competence levels with no trend of improvement with higher or lower competence: In the CG (n=15), ICC=.851, in the IG (n=7), ICC =.905, in the HG (n=31), ICC=.796, and in LG (n=29), ICC=.900. Importantly, interrater reliability is also good for the LG, which is necessary for the test to be a promising measure of language competence at early stages of acquisition. Since inter-rater reliability was good, the mean score of the two raters was used as the test score for each child.

Sentence repetition as a measure

Validity was first investigated by checking that the four competence groups perform the way that we would expect if the test is a measure of language competence. The descriptive statistics in Table 2 show that this is the case; the CG performs best of the groups, and there is a linear relationship between group mean and expected competence level. Furthermore, standard deviations increase with lower mean scores, indicating higher variability with lower competence, which is to be expected; this difference holds also between the HG and the LG, where this is not a result of ceiling effects. We see ceiling effects in the the CG, since the maximum possible score of 142 is obtained by two subjects, and the mean is as high as 138.20. Also the highest IG score is close to ceiling, only 2.75 points from the maximum score. For the two lower competence groups, however, scores are much lower and no participant performed near ceiling.

Table 2: Descriptive statistics for the repetition test for each individual participant groups

Group	N	Range	Minimum	Maximum	Mean	Std. Deviation	Variance
CG	15	11.50	130.50	142.00	138.20	3.44	11.86
IG	7	19.25	120.50	139.75	132.64	6.53	42.67
HG	31	59.00	58.00	117.00	93.92	13.88	192.59
LG	29	68.75	45.75	114.50	85.07	20.08	403.12

A one-way ANOVA comparing group means was run with planned comparisons comparing the CG to the IG, the IG to the HG, and the HG to LG. Levene's test for equal variances was significant; values not assuming equal variances are therefore reported. As predicted, there was an overall effect for group [Welch's $F(3, 25.08)= 139.080$, $p<.001$, $r_{effect}=.82$]. The comparisons also revealed significant differences between the CG and the IG [$t(7.6)= 2.12$ $p=.034$, $r_{effect}=.61$], between the IG and the HG [$t(20.3)= 11.04$, $p<.001$, $r_{effect}=.93$], and between the HG and LG [$t(49.4) =1.97$, $p= .027$, $r_{effect}=.27$]. As illustrated in Figures 1 – 3, the pattern of differences between groups on the two other English measures, i.e., vocabulary and sentence comprehension, are similar to differences on sentence repetition, although absolute scores for the three measures are on different scales and not directly comparable. This similar pattern is an indication that the repetition test validly measures language competence.

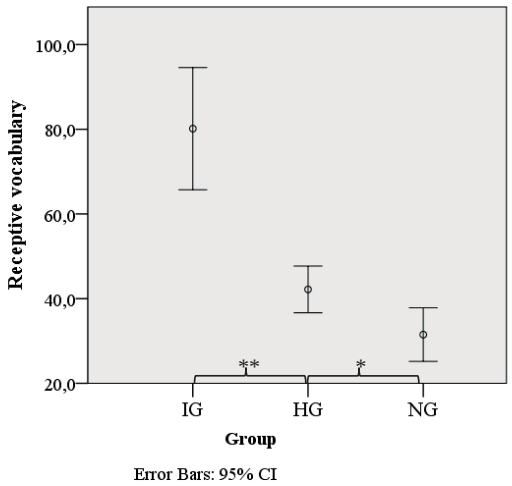


Figure 1: Means for all L2 participant groups on receptive vocabulary (not tested in CG)
*=pairwise comparison p<.05
**= pairwise comparison p<.01

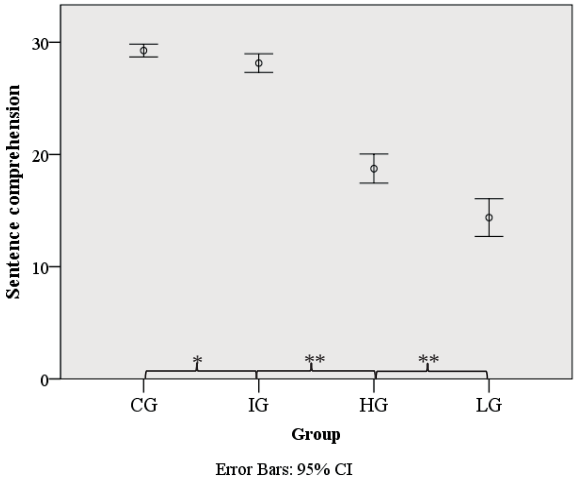


Figure 2: Means for all participant groups on sentence comprehension
*=pairwise comparison p<.05
**= pairwise comparison p<.01

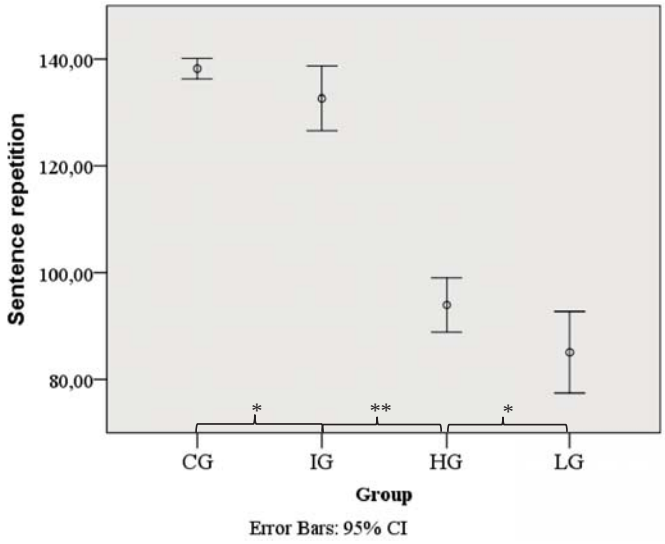


Figure 3: Means for all participant groups on sentence repetition
*=pairwise comparison p<.05, **= pairwise comparison p<.01

Sentence repetition as a measure

For further validity testing, correlation analyses were run for scores in the HG and the LG on the repetition test and all control measures, i.e., English vocabulary, English sentence comprehension, visio-spatial memory, non-verbal intelligence, verbal ability, and Norwegian vocabulary. These analyses were intended to ascertain that there were no strong correlations between sentence repetition and non-L2 measures which may indicate that the repetition test taps abilities not related to L2 competence. The CG and the IG were not included since the former was not tested on non-English measures, and since multilingualism in the latter might have caused different relationships than in the two groups where Norwegian was the only L1. Correlations are summarized in Table 3 below.

Table 3: Pearson correlation between English measures and all other measures in the HG and the LG, with bootstrap based on 1000 samples

		Sentence comprehension	English vocabulary	Norwegian vocabulary	Verbal ability	Non-verbal intelligence	Visio-spatial memory
Sentence repetition	*r*	.506**	.325**	.521**	.285*	.191	.120
	95% BCa CI	.286, .669	.094, .528	.302, .676	.036, .491	-.011, .357	-.102, .340
Sentence comprehension	*r*		.451**	.379**	.038	.164	.120
	95% BCa CI		.193, .658	.113, .603	-.203, .252	-.092, .424	-.157, .410
English vocabulary	*r*			.426**	.447**	.403**	.071
	95% BCa CI			.200, .620	.230, .633	.189, .626	-.179, .314

*Correlation is significant at the .05 level
** Correlation is significant at the .01 level

Again, results indicate that the sentence repetition test is a valid language measure. There is a quite strong correlation between sentence repetition and English sentence comprehension, which is expected if both involve language processing. Sentence repetition is also correlated with English vocabulary, with Norwegian (L1) vocabulary and with verbal ability. A correlation with L1 measures is to be expected if the repetition test validly assesses English competence, since English vocabulary and sentence comprehension are also correlated with Norwegian vocabulary, and English vocabulary is in fact more strongly correlated than sentence repetition with verbal ability. In other words, there is a general relationship between L1 and L2 measures. There is,

however, no correlation between sentence repetition and non-linguistic measures, i.e., visio-spatial memory or non-verbal intelligence[6].

Finally, results on the repetition test were checked for primacy and recency effects. Bley-Vroman and Chaudron (1994) argue that we can expect such effects in a repetition test if it involves serial recall rather than language processing. In serial recall, the last item is expected to be recalled the best, followed by the first item. Strong recency or primacy effects in this test could therefore indicate that memory is a stronger predictor for success than language proficiency.

A score was calculated for each individual word in the test, based on the percentage of children who were judged by both raters to have successfully repeated it. 100% correct in this context would thus entail that both raters agreed that all children had successfully repeated a word. Each word was then coded according to whether it was sentence initial, sentence final, or sentence medial, i.e., neither initial nor final. Sentence initial words had the highest mean score and the smallest variance, followed by sentence final words and sentence medial words. This is illustrated in Figure 4.

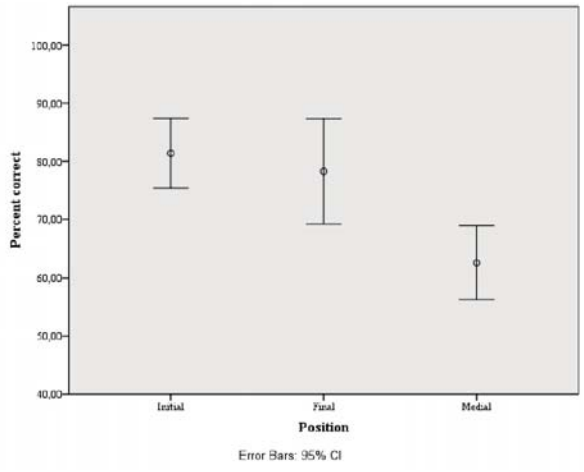


Figure 4: Percent correct for words by sentence position (initial, medial, or final) in all participant groups combined
(initial vs. final: p=non-significant, initial/final vs. medial: p<.001)

[6] In fact, English vocabulary is the only measure correlated with non-verbal intelligence, indicating that the relationship between L2 and other abilities is complex.

A one-way ANOVA was performed comparing sentence initial words, sentence medial words, and sentence final words. Levene's test for homogeneity of variance was significant, and thus values not assuming homogeneity are reported. The overall ANOVA found an effect of sentence position [Welch's F(2, 39.1)= 10.25, p=.001, r_{effect}=.40]. Planned comparisons were carried out first on sentence initial/final words versus sentence medial words, and then for sentence initial versus sentence final words. The first contrast found that the difference between non-medial and medial words was indeed significant [t(76.36)=4.24, p<.001, r_{effect}=.92]. The difference between initial and final words, however, was not significant [t(27.81)=.608, p<.548].

4.1.2. Discussion

It is clear from the above results that internal consistency and interrater reliability in the test was good, and that the test validly measures language competence, since performance on the test for each group corresponds to performance on other L2 measures. We do see primacy and recency effects, which could be an indication that the test measures serial recall rather than language. However, a lower score on medial words is expected also if language processing is measured, since the initial and final elements of a sentence are more salient than the middle. The fact that recency effects are not stronger than primacy effects indicates that the lower scores on medial words are not due to working memory alone. Thus, these effects are not really counterevidence to the overall conclusion that sentence repetition is a valid measure of L2 competence.

A shortcoming of the present study is the fact that no measure of phonological short-term or working memory, e.g., an L1 or non-word repetition test, was included. However, such a test would not necessarily have revealed important, new information. The fact that the present study found no correlation between visio-spatial working memory and sentence repetition makes a very strong correlation between phonological working memory and sentence repetition unlikely, since we can assume a connection between different memory components (Erlam, 2006, p. 467; Gathercole & Baddeley, 1993), and since there is a correlation between verbal and spatial working memory (cf. Kane et al., 2004). Furthermore, phonological working memory is known to facilitate L2 acquisition itself in areas such as vocabulary, grammar, and fluency, possibly especially at early stages; the phonological loop may even be the best candidate for a Language Acquisition Device in humans (e.g., Baddeley, 2014; Baddeley, Gathercole, & Papagno, 1998;

Gathercole, 2006; Martin & Ellis, 2012; Williams, 2012). Working memory may be involved in the process of comprehension specifically through the phonological loop (Gathercole & Baddeley, 1993; Van Moere, 2012), and it is likely that syntactic processing depends on short term or working memory in an L2 (Bley-Vroman & Chaudron, 1994; Williams, 2012). This means that if we were to find a correlation between sentence repetition and a phonological memory measure such as non-word repetition, this would not tell us much about causality; superior phonological working memory may be of particular help for repeating sentences regardless of L2 proficiency, but it may also be generally beneficial for L2 development, leading to high performance on valid L2 measures including sentence repetition. This study did, however, investigate the effect of sentence length on repetition, which may provide insights about memory processes in sentence repetition. This is addressed in the next section.

4.2. What is the effect of sentence length?

4.2.1. Results

The second research question was whether sentence length is important for whether and how sentence repetition tests language competence rather than the ability to repeat an acoustic image. In the present study, sentences ranged from two to 13 syllables. Five sentences were eight syllables or over, while the remaining 12 sentences were seven syllables or under. Following proposals for the threshold for mimicking, results were divided into scores on sentences of eight syllables and over, and scores on sentences of seven syllables and under. Correlation analysis using Pearson correlations with bootstrapping based on 1000 samples was run for scores on shorter and longer sentences for all participants together and for each competence group separately. These results are reported in Table 4.

Table 4: Pearson correlations between scores on sentences seven syllables and under, and sentences eight syllables and over, for all participant groups combined and separately

Group	*r*	Sig. (1-tailed)	95% BCa* CI
All (n=82)	.783	<.001	.715, .844
CG (n=15)	.292	.145	-.253, .591[a]
IG (n=7)	.795	.016	.146, 1.000
HG (n=31)	.711	<.001	.499, .846
LG (n=29)	.722	<.001	.578, .828

a. Based on 999 samples
* Bias corrected accelerated

A high correlation was found for all participants together. However, there is no significant correlation between short and long sentences in the CG, while the magnitude of correlations is similar in the other three groups. For the IG, Spearman's ρ was checked because of the low number of participants, and a significant correlation was again found ρ=.815, p=.013.

To investigate whether there was an interaction between sentence length and language competence level, a mixed 4x2 factorial analysis was conducted with group as the between-subjects factor and sentence length as the within-subjects factor. The analysis found a main effect for sentence length [$F(1,78)= 423.73$, $p<.001$, partial $\eta^2=.845$], as well as an interaction effect for group*sentence length [$F(3,78)= 83.42$, $p<.001$, partial $\eta^2=.762$]. The interaction is shown in Figure 5:

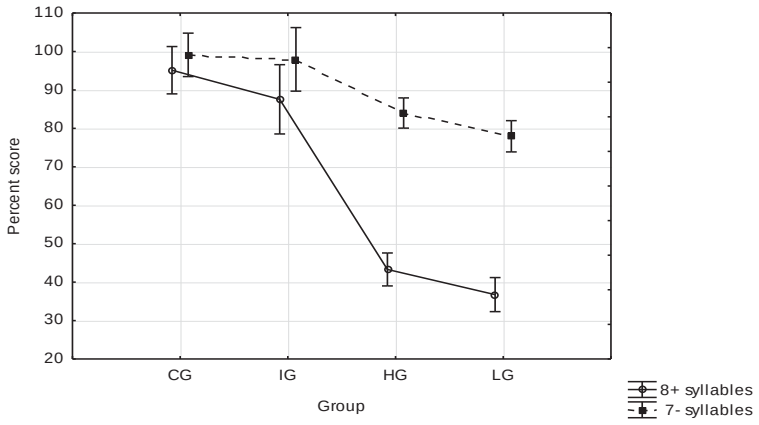


Figure 5: Mean percent correct on shorter and longer sentences for all participant groups

To find the source of the interaction effect, LSD post hoc tests were conducted. The tests revealed no significant differences between long and short sentences in the CG (p=.086) but a significant difference in the IG (p=.003), in the HG (p<.001), and in the LG (p<.001). The test furthermore revealed no significant difference between the CG and the IG on short (p=.823) nor on long (p=.153) sentences. However, the difference was significant between the IG and the HG on short (p=.005) and on long (p<.001) sentences, and also between the HG and the LG on short (p=.044) and on long (p=.030) sentences.

4.2.2. Discussion

The above results show that sentence length, predictably, does play a role in repetition. For the CG, there is no significant difference between shorter and longer sentences, but there is some evidence that the long sentences were challenging even for native speakers: A total of seven out of 15 participants in the CG made alterations or additions to one or more sentences, and with the exception of one participant who changed four sentences, all changes were made to the two longest sentences, i.e., sentences 15 and 16. These changes are summarized in Table 5.

Table 5: Alterations and additions made by CG participants

Sentence	**Changes made**
2. Horses run fast	Horses run **very** fast.
3. The baby drinks milk	The baby drinks milk **fast**.
15. The children's dog always sleeps on the floor	The children's dog **is always sleeping** on the floor .* The children's dog **sleeps always** on the floor.
16. The boy has been playing with his friends for an hour	The boy**s have** been playing with **their** friends for an hour. The boy has **played** with **the** friends **in** an hour. The **little** boy has been playing with his friends for an hour.

*Alteration was made by three children. All other alterations were made by one child each.

In the IG, we see clear ceiling effects in the short sentences, with a mean percent correct of 97.95 (SD 2.7), while performance is lower on longer sentences, with a mean percent of 87.56 (SD 7.5). In the HG and the LG, performance drops even more dramatically for the longer sentences, with mean percentages of 43.30 (SD 11.9) in the HG and 36.7 (SD 15.0) in the LG. However, also the short sentences seem to have been challenging since their mean percentages are only 84.00 (SD 9.6) and 77.94 (SD 15.3), respectively. These two groups' English

competence has been established as clearly below that of the CG and the IG on other language measures, and their low scores also on the repetition of short sentences thus indicate that even when sentences are short, language competence, i.e., the ability to decode, comprehend, and/or reproduce language, is measured.

If a sentence is too short to be a valid test of L2 development, it should display no difference in performance between native speakers and very low-competence L2 learners; after all, if a test merely requires parroting an acoustic image, language competence should play no role at all. However, t-tests with bootstrapping based on 1000 samples comparing the means of the two extreme groups, i.e., the CG and the LG, find a significant difference for each individual sentence in the test. This includes sentence 1, which has only two syllables, [t(28.0)=3.36, p=.001, r_{effect}=.54]. It also includes two sentences with three syllables, namely sentence 7, [t(37.8)=3.66, p=.001, r_{effect}=.51], and sentence 11, [t(29.7)=3.91, p<.001, r_{effect}=.58]. This is evidence that even with very short sentences, participants were not able to merely parrot an acoustic image. It is thus clear that when evaluating sentence repetition as a measure of L2 competence, the question is not how long sentences need to be in order to test language rather than parroting skills. Language abilities seem to be involved already with the shortest possible sentences.

On the other hand, there is no clear evidence of floor effects. For the two low-competence groups, i.e., the HG and the LG, there is, as already mentioned, a significant difference also on the long sentences. However, these scores are for overall accuracy on all morphemes, and it is possible that there would have been floor effects if what was evaluated was a specific structure in the more complex sentences. In this context it is also worth remembering that English morphology is fairly impoverished. Again, the question seems to be what sentence length is appropriate for testing which domain of language.

4.3. Which language competences are tested?

4.3.1. Results

The evidence that sentence repetition is in fact a valid measure of language competence is in line with conclusions that it is at least a reasonable measure of global proficiency (cf. Bley-Vroman & Chaudron, 1994, p. 247; Romeo, Gentile, & Bernhardt, 2008, p. 299). However, the question remains of exactly what aspects of language proficiency it can assess, i.e., the third

research question of the present study. The hypothesis was that sentence repetition requires more detailed grammatical analysis than sentence comprehension. The expected consequence would be that uninflected lexical words are repeated more successfully than are functional words and inflected lexical words (cf. Connell & Myles-Zitzer, 1982).

The analysis to investigate the effect of word type (functional, inflected lexical, or uninflected lexical) was carried out on sentence medial words only. This was done to rule out the effect of primacy and recency effects in the data which we saw in section 4.1.3 since out of the 17 sentences in the test, 14 started with a functional word, while all sentences ended in a lexical word. Because of ceiling effects in the CG and the IG (cf. section 4.1.1), only the HG and the LG were included in the analysis.

A 3 (word type: functional, uninflected, inflected) x 2 (group: HG, LG) mixed factorial ANOVA was conducted which found, predictably, a main effect of group [$F(1,47)=13.45$, $p=.0006$, partial $\eta 2 =0.22$, but no main effect of word type. An interaction effect was found of group*word type $F(2,47)=4.22$, $p=.021$, partial $\eta 2=0.15$. This interaction is illustrated in Figure 6.

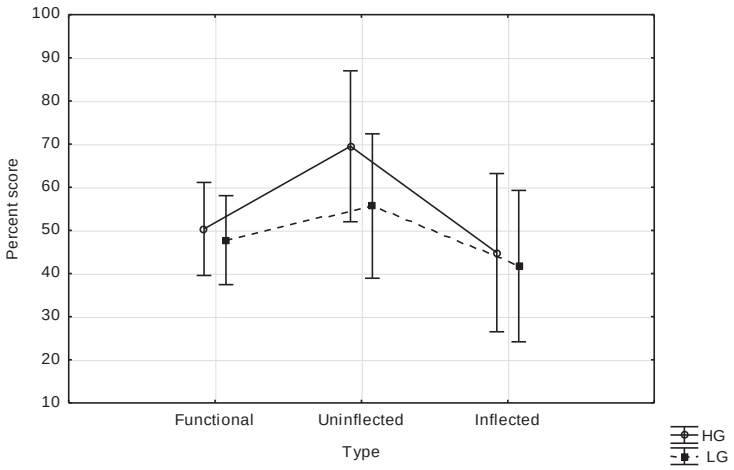


Figure 6: Percentage correct for word type (functional, uninflected lexical, or lexical) in HG and LG

Post hoc Tukey HSD tests found the source of the interaction effect to be that the groups differed significantly on uninflected lexical words (p=.002), but not on inflected lexical (p=.950) nor on functional words (p=.814). Planned contrasts revealed that in the HG there was a significant difference between lexical uninflected words on the one hand and lexical inflected/functional words on the other [t(47)=2.15, p=.036]. Planned contrasts in the LG found no corresponding significant difference [t(47)= 1.12, p=.268].

4.3.2. Discussion

The fact that there is an interaction effect of word type, and that uninflected lexical words are more accurately reproduced than functional words and inflected lexical words in the bilingually-based group, is evidence that sentence repetition is indeed a test of grammatical processing. It is possible that in this group, some sentences have only been semantically processed, and only "shallow" or "good enough" representations built (e.g., Clahsen & Felser, 2006; Ferreira, 2003; Ferreira et al., 2002; VanPatten, 2012), so that the participant may be able to repeat the lexical words, but without inflections, and without functional words.

The lack of effect of word type in the LG indicates that lexical words and not only grammatical morphemes are problematic. It is possible that perceptive skills in the L2 are involved; after all, for even a lexical word to be accurately reproduced, it needs to have been accurately perceived. It has already been demonstrated that the LG performed significantly below the CG even on the very shortest sentence, which consisted of only two words. A closer look at the errors in the LG on this sentence reveals two types of problem. Some children miss the plural -s on birds, but a more frequent problem is that the word "fly" is mispronounced as "fry" by a number of LG participants. This mistake may have a number of sources. It could be a simple pronunciation problem. However, it is just as likely that the problem is in perceiving the phoneme as in reproducing it. If children do not realize that the word in the stimulus sentence is "fly", not "fry", this may result from a combination of two different problems, namely a lack of vocabulary and low decoding skills. If a child is familiar with the word "fly", she would be able to use her L1 and/or world knowledge to expect that this word is more likely than "fry" to take "birds" as its subject. However, if "fly" is not a familiar word and if, in addition, the child's skills in decoding English speech are not well developed, "fly" may be misheard as "fry". Again, it is clear that language skills are being tested although it is not clear which skill – phoneme perception or production.

It is plausible that the language skills involved in repetition at very low levels are simply aspects of phonological awareness such as phoneme discrimination, i.e., segmenting the speech stream, and recognizing lexical items. However, these are language competences in themselves, which do not necessarily transfer across languages (e.g., Bialystok, McBride-Chang, & Luk, 2005). In other words, even short sentences are valid measures of an L2 skill although they may only be useful at very low competence levels.

In order to be sure which language skill is being tested in a sentence repetition test, carefully controlling other abilities is necessary. In the present study, controlling for participants' ability to produce the English /l/ phoneme in isolation or in other contexts might have helped discern whether the problem in the shortest sentence is one of production or of perception. Checking vocabulary knowledge could have established whether children actually knew the word "fly". If children knew the word and were able to pronounce its phoneme structure in a different context, we would have evidence that problems were in processes of comprehension such as phoneme perception and segmentation. Such controlling was outside the scope of this study. However, it seems that with the appropriate measures to control for ambiguity in results, sentence repetition can be a useful test also of specific domains of language, even at very low levels.

We thus have evidence for at least three aspects of proficiency involved in the sentence repetition test, namely language-specific phoneme perception, lexical knowledge, and grammatical processing. For children with low L2 competence, the first two may in fact be the most important for differences in performance. In the IG, we have little evidence to draw strong conclusions about the source of errors. The lower mean score in this group compared to the CG may have to do with automaticity of processing, i.e., that processing takes enough memory resources to cause "slips" and non-systematic errors. Such an account would predict more errors in long than in short sentences, in line with the results in section 4.2. However, the IG is very small and variable, and some children show ceiling effects while others do not. Thus, we cannot draw firm conclusions about the source of errors in the repetition test for this group.

5. Conclusion

The sentence repetition test discussed in this paper was found to be a reliable and valid measure of L2 English for children at approximately age six, with different levels of L2 competence. It is a practical measure which is easy to administer, which can be completed by

young children even at early stages of acquisition, and which discriminates well between proficiency levels. Sentence repetition clearly measures language competence, not just mimicking. An effect was found of sentence length, but this does not seem to be because shorter sentences can be parroted or remembered as "acoustic images", since language proficiency predicts performance also on short sentences. Low scores on short sentences in the low-proficiency groups may be a result of difficulties in production, in lexical knowledge, or in phonological awareness or phoneme perception. Appropriate sentence length in such a test will thus depend on the proficiency of the participants, and on exactly which competence we want to measure. The complexity of the skills involved in the task indicates that if it is used to test specific subfields of language, other factors such as pronunciation and vocabulary should be carefully controlled. However, sentence repetition may also be useful as part of a test battery, for example, to test "overall" or "global" competence.

The present study has a number of limitations. The the fact that phonological working memory was not controlled has already been discussed. The small size of the IG also prevents us from drawing strong conclusions about the test for their competence level. There is a substantial gap in competence levels between the IG and the HG, and the inclusion of a more intermediate-level group might have provided more knowledge about the effect of factors such as sentence length and word type. Finally, the languages included in the present test, i.e. L1 Norwegian and L2 English, are related and structurally similar. Evaluating such a test for more different language pairs may thus reveal new information.

Appendix

Repetition test and scoring guide

Trials

Pen
Fish
Eight
Shoulder
Pencil
Seven
I am a boy/girl.
You are three years old.
Is it raining outside?

Test sentences and scoring guide

Sentence	*Morphemes*	*Total*
1. Birds fly.	bird+s fly+Ø	4
2. Horses run fast.	Horse+s run+Ø fast	5
3. The baby drinks milk.	the baby+Ø drink+s milk+Ø	7
4. She is jumping.	she+x is+x jump+ing	6
5. He is eating a banana.	he+x is+x eat+ing a/ banana+Ø	9
6. She has washed the dog.	she+x has+x wash+ed the dog+Ø	9
7. They have played.	they+x have+x play+ed	6
8. He is not young.	he+x is+x not young	6
9. Cats do not fly.	cat+s do+Ø not fly+Ø	7
10. Is the elephant playing?	is+x the elephant+Ø play+ing	7
11. Does she sing?	do+es she+x sing+Ø	6
12. Do you like to read?	do+Ø you+x like+Ø to read+Ø	9
13. Are the children at school?	are+x the child+ren at school+Ø	8
14. The cat is playing with a ball.	the cat+Ø is+x play+ing with a/ ball+Ø	11
15. The children's dog always sleeps on the floor.	the child+ren+'s dog+Ø always sleep+s on the floor+Ø	13
16. The boy has been playing with his friends for an hour.	the boy+Ø has+x be+en play+ing with his+x friend+s for an/ hour+Ø	18
17. The house with the blue door looks nice.	the house+Ø with the blue door+Ø look+s nice	11
Total:		142

Ø = point for zero morpheme
x = 1 point for auxiliary/pronoun
/ = 0.5 point for indefinite article, 0.5 for correct form

References

Baddeley, A. D. (2014). *Essentials of human memory*. East Sussex: Psychology Press.

Baddeley, A. D., Gathercole, S. E., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review, 105*(1), 158-173. doi: 10.1037/0033-295X.105.1.158

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior, 14*(6), 575-589. doi: 10.1016/S0022-5371(75)80045-4

Berry, P. B. (1976). Elicited imitation of language: some ESNS population characteristics. *Language and speech, 19.1*, 350-363.

Bialystok, E. (2009). Bilingualism: The good, the bad, and the indifferent. *Bilingualism: Language and Cognition, 12*(1), 3-11. doi: 10.1017/S1366728908003477

Bialystok, E., McBride-Chang, C., & Luk, G. (2005). Bilingualism, language proficiency, and learning to read in two writing systems. *Journal of Educational Psychology, 97*(4), 580-590. doi: 10.1037/0022-0663.97.4.580

Bishop, D. V., & Snowling, M. J. (2004). Developmental dyslexia and specific language impairment: same or different? *Psychological Bulletin, 130*(6), 858-886. doi: 10.1037/0033-2909.130.6.858

Bley-Vroman, R., & Chaudron, C. (1994). Elicited imitation as a measure of second-language competence. In E. Tarone, S. M. Gass & A. D. Cohen (Eds.), *Research methodology in second-language acquisition* (pp. 245-261). Hillsdale, NJ: L. Erlbaum.

Brady, S. A. (1997). Ability to encode phonological representations: An underlying difficulty of poor readers. In B. A. Blachman (Ed.), *Foundations of reading acquisition and dyslexia: Implications for early intervention* (pp. 21-47). Mahwah, NJ: Lawrence Erlbaum.

Brown, R., & Fraser, C. (1963). The acquisition of syntax. In C. N. Cofer & B. S. Musgrave (Eds.), *Verbal behavior and learning: Problems and processes. Proceedings of the second conference sponsored by the Office of Naval Resarch and New York University.* (pp. 158-197). New York, NY: McGraw-Hill.

Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics, 27*(1), 3-42. doi: 10.1017/S0142716406060024

Connell, P. J., & Myles-Zitzer, C. (1982). An analysis of elicited imitation as a language evaluation procedure. *J Speech Hear Disord, 47*(4), 390-396. doi: 10.1044/jshd.4704.390

Darcy, I., Daidone, D., & Kojima, C. (2013). Asymmetric lexical access and fuzzy lexical representations in second language learners. *The Mental Lexicon, 8*(3), 372-420. doi: 10.1075/ml.8.3.06dar

de Groot, A. M. B. (2002). Lexical representation and lexical processing in the L2 user. In V. Cook (Ed.), *Portraits of the L2 user* (pp. 29-64). Clevedon: Multilingual Matters.

Devescovi, A., & Caselli, C. (2007). Sentence repetition as a measure of early grammatical development in Italian. *International Journal of Language & Communication Disorders, 42*(2), 187-208. doi: 10.1080/13682820601030686

Dunn, L. M., & Dunn, D. M. (2007a). Peabody Picture Vocabulary Test. Minneapolis, MN: Pearson Education, Inc.

Dunn, L. M., & Dunn, D. M. (2007b). *Peabody Picture Vocabulary Test Manual* (4 ed.). Minneapolis, MN: Pearson Education, Inc.

Eisenbeiss, S. (2010). Production methods in language acquisition research. In E. Blom & S. Unsworth (Eds.), *Experimental methods in language acquisition research* (pp. 11-34). Philadelphia, PA: John Benjamins.

Ellis, R. (1994). *The study of second language acquisition*. Oxford: Oxford University Press.

Erlam, R. (2006). Elicited imitation as a measure of L2 implicit knowledge: An empirical validation study. *Applied Linguistics, 27*(3), 464-491. doi: 10.1093/applin/aml001

Fabbro, F., Alberti, B., Gagliardi, C., & Borgatti, R. (2002). Differences in native and foreign language repetition tasks between subjects with William's and Down's syndromes. *Journal of Neurolinguistics, 15*(1), 1-10. doi: 10.1016/S0911-6044(00)00016-6

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology, 47*(2), 164-203. doi: 10.1016/S0010-0285(03)00005-7

Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science, 11*(1), 11-15. doi: 10.1111/1467-8721.00158

Flynn, S. (1986). Production vs. comprehension: Differences in underlying competences. *Studies in Second Language Acquisition, 8*, 135-164. doi: 10.1017/S0272263100006057

Fujiki, M., & Brinton, B. (1983). Sampling reliability in elicited imitation. *Journal of Speech and Hearing Disorders, 48*(1), 85-89. doi: 10.1044/jshd.4801.85

Gathercole, S. E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics, 27*(4), 513-613. doi: 10.1017/S0142716406060383

Gathercole, S. E., & Baddeley, A. D. (1993). *Working memory and language*. Hove: Erlbaum.

Hamayan, E., Saegert, J., & Larudee, P. (1977). Elicited imitation in second language learners. *Language and speech, 20*, 86-97. doi: 10.1177/002383097702000109

Helland, T. (2008). Second language assessment in dyslexia: Principles and practice. In J. Kormos & E. H. Kontra (Eds.), *Language learners with special needs: An international perspective* (pp. 63-85). Clevedon: Channel View Publications.

Helland, T., & Kaasa, R. (2005). Dyslexia in English as a second language. *Dyslexia, 11*(1), 41-60. doi: 10.1002/dys.286

Hulstijn, J., & De Graaff, R. (1994). Under what conditions does explicit knowledge of a second language facilitate the acquisition of implicit knowledge? A research proposal. *AILA Review, 11*(97-113).

Izumi, S. (2003). Comprehension and production processes in second language learning: In search of the psycholinguistic rationale of the Output Hypothesis. *Applied Linguistics, 24*(2), 168-196. doi: 10.1093/applin/24.2.168

Jessop, L., Suzuki, W., & Tomita, Y. (2007). Elicited imitation in second language acquisition research. *The Canadian Modern Language Review/La Revue canadienne des langues vivantes, 64*(1), 215-238. doi: 10.3138/cmlr.64.1.215

Kamhi, A. G. C., Hugh W. (2008). The language basis of reading: Implications for classification and treatment of children with reading disabilities. In K. G. Butler & E. R. Silliman (Eds.), *Speaking, reading, and writing in children with languge learning disabilities* (pp. 45-73). Mahwah, NJ: Lawrence Erlbaum.

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General, 133*(2), 189-217. doi: 10.1037/0096-3445.133.2.189

Kaufman, A. S., & Kaufman, N. L. (2004a). Kaufman Brief Intelligence Test. Minneapolis, MN: Pearson Education, Inc.

Kaufman, A. S., & Kaufman, N. L. (2004b). *Kaufman Brief Intelligence Test, second edition manual*. Minneapolis, MN: Education, Inc.

Keeney, T. J., & Wolfe, J. (1972). The acquisition of agreement in English. *Journal of Verbal Learning and Verbal Behavior, 11*(6), 698–705. doi: 10.1016/S0022-5371(72)80004-5

Kroll, J. F., & Tokowicz, N. (2005). Models of bilingual representation and processing: Looking back and to the future. In J. F. d. G. Kroll, Annette M. B. (Ed.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 531-553). Cary, NC, USA: Oxford University Press.

Larsen-Freeman, D., & Long, M. H. (1991). *An introduction to second language acquisition research*. New York, NY: Longman.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: The MIT Press.

Li, P., & Zhao, X. (2014). Connectionist bilingual representation. In R. R. A. Heredia, Jeanette (Ed.), *Foundations of bilingual memory* (pp. 63-82). New York, NY: Springer.

Markman, B. R., Spilka, I. V., & Tucker, G. R. (1975). The use of elicited imitation in search of an interim French grammar. *Language Learning, 75*, 31-41. doi: 10.1111/j.1467-1770.1975.tb00107.x

Martin, K. I., & Ellis, N. C. (2012). The roles of phonological short-term memory and working memory in L2 grammar and vocabulary learning. *Studies in Second Language Acquisition, 34*(3), 379-413. doi: 10.1017/S0272263112000125

Munnich, E., Flynn, S., & Martohardjono, G. (1994). Elicited imitation and grammaticality judgment tasks: What they measure and how they relate to each other. In E. Tarone, S. M. Gass & A. D. Cohen (Eds.), *Research methodology in second-language acquisition* (pp. 227-243). Hillsdale, N.J.: L. Erlbaum.

Muñoz, C. (2006). The BAF Project: Research on the effects of age on foreign language acquisition. In C. Abello-Contesse, R. Chacón-Beltrán, M. D. López-Jiménez & M. M. Torreblancaa-López (Eds.), *Age in L2 acquisition and teaching* (pp. 81-92). Bern: Peter Lang.

Naiman, N. (1974). The use of elicited imitation in second language acquisition research. *Working Papers on Bilingualism, 2*, 37.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Pavlenko, A. (2009). Conceptual representation in the bilingual lexicon and second language vocabulary learning. In A. Pavlenko (Ed.), *The bilingual mental lexicon: interdisciplinary approaches* (pp. 125-160). Bristol: Multilingual Matters.

Perkins, K., Brutten, S. R., & Angelis, P. J. (1986). Derivational complexity and item difficulty in a sentence repetition task. *Language Learning, 36*(2), 125-141. doi: 10.1111/j.1467-1770.1986.tb00375.x

Pickering, M. J., & Ferreira, V. S. (2008). Structural priming: A critical review. *Psychological Bulletin, 134*(3), 427-459. doi: 10.1037/0033-2909.134.3.427

Romeo, K., Gentile, L., & Bernhardt, E. (2008). Sentence repetition and story retelling as indicators of language proficiency in young bilingual children. In Y. Kim, V. J. Risko, D. L. Compton, D. K. Dickinson, M. K. Hundley, R. T. Jiménez, K. M. Leander & D. W. Rowe (Eds.), *57th Yearbook of the National Reading Conference* (pp. 298-310). Oak Creek, WI: National Reading Conference.

Rondal, J. A., Lambert, J. L., & Sohier, C. (1981). Elicited verbal and nonverbal imitation in Down's syndrome and other mentally retarded children: A replication and extension of Berry. *Language and speech, 24*(3), 245-254. doi: 10.1177/002383098102400304

Roy, P., & Chiat, S. (2004). A prosodically controlled word and nonword repetition task for 2- to 4-year-olds: Evidence from typically developing children. *Journal of Speech, Language, and Hearing Research, 47*(1), 223-234. doi: 10.1044/1058-0360(2002/009)

Sebastian-Gallés, N., Rodríguez-Fornells, A., Diego-Balaguer, R. d., & Díaz, B. (2006). First- and second-language phonological representations in the mental lexicon. *Journal of Cognitive Neuroscience, 18*(8), 1277-1291. doi: 10.1162/jocn.2006.18.8.1277

Small, J. A., Kemper, S., & Lyons, K. (2000). Sentence repetition and processing resources in Alzheimer's disease. *Brain and Language, 75*(2), 232-258. doi: 10.1006/brln.2000.2355

Smith, C. S. (1970). An experimental approach to children's linguistic competence. In J. R. Hayes (Ed.), *Cognition and the development of language* (pp. 109-135). New York, NY: Wiley.

Thornton, R. (1996). Elicited production. In D. McDaniel, C. McKee & H. S. Cairns (Eds.), *Methods of assessing children's syntax* (pp. 77-102). Cambridge, MA: MIT Press.

Trofimovich, P. (2005). Implicit memory support for bilingual speech processing: An investigation of auditory word priming in English-Spanish bilinguals. In J. Cohen, K. T. McAlister, K. Rolstad & J. MacSwan (Eds.), *Proceedings of the 4th International Symposium on Bilingualism* (pp. 2257-2274). Somerville, MA: Cascadilla Press.

Van Moere, A. (2012). A psycholinguistic approach to oral language assessment. *Language Testing, 29*(3), 325-344. doi: 10.1177/0265532211424478

VanPatten, B. (2012). Input processing. In A. Mackey & S. M. Gass (Eds.), *The Routledge handbook of second language acquisition* (pp. 268-281). London: Routledge.

Vinther, T. (2002). Elicited imitation: A brief overview. *International Journal of Applied Linguistics, 12*(1), 54-73. doi: 10.1111/1473-4192.00024

Williams, J. N. (2012). Working memory and SLA. In S. M. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 427-441). London: Routledge.

Paper III

Input and language competence in early-start foreign language classrooms

Anne Dahl

Abstract

The aim of the present study was to investigate the effect of target language exposure on second language competence in early-start English classrooms in Norway. Participants were 60 children. In one group (n=29, mean age 6; 1), English was taught in the standard manner for Norwegian English classrooms, with English mainly as the object but not medium of instruction. In the other group (n=31, mean age 6;1), teachers focused on providing as much English input as possible within the framework of the normal curriculum, i.e., with English as the medium and not only object of instruction. L2 proficiency after the first year of school was measured through three different tests. Receptive vocabulary was tested through the Peabody Picture Vocabulary Test, fourth edition. Sentence comprehension and sentence repetition were tested through tests developed specifically for the project. Results indicate that input in such English classrooms can be sufficient for acquisition to take place, as t-tests found a difference between the two groups on all three measures. The strongest effect was for sentence comprehension. Multiple regression found that group membership (i.e., volume of input) predicts sentence comprehension better than any background factor tested. The conclusion is that even limited input in the classroom can lead to implicit acquisition in young learners, and that exposure is especially beneficial for language processes such as speech segmentation and lexical retrieval, possibly enabling learners to build "good enough" representations. This type of processing is argued to be a prerequisite for further acquisition.

1. Introduction

Increased needs for multilingualism and studies indicating that a young age of acquisition (AoA) is beneficial for ultimate attainment in language have led many countries to lower starting ages for foreign language (FL) in schools. However, a number of studies indicate that such early-start FL programs are not useful for successful acquisition. In the present paper, the effect of increased input in an early FL classroom is investigated for a group of young L1 speakers of Norwegian acquiring English. L2 competence was measured with three tests of English, namely

receptive vocabulary, sentence comprehension, and sentence repetition. The study investigates whether input in such an early FL classroom can be substantial enough for an effect to be found on all these language measures, and if so, which measure shows the greatest effect of increased input. The final question is what this can tell us about the nature of early FL competence. The paper is organized as follows: Section 2 outlines the theoretical background for the study. Section 3 describes the study, including methods and participants. Section 4 presents the results of a number of background measures, as well as of the three measures of English competence. In section 5, the findings are discussed. Section 6 provides the conclusion.

2. Background

2.1. Second language acquisition in the classroom

Processes and outcomes in language acquisition may differ greatly depending on context and circumstances. In particular, there are a number of observable differences between children's acquisition of their native language and the subsequent acquisition of additional languages at later ages. First language (L1) acquisition normally leads to perfect ultimate attainment, at least in terms of grammatical competence, although lexical knowledge may vary. Perfect grammatical competence, however, is not the norm for second languages (L2s), which are subject to individual differences and rarely, if ever, reach native-like levels. Also circumstances of acquisition differ. L1 acquisition is wholly naturalistic, while L2 acquisition often incorporates elements of instruction. Finally, while L1 acquisition is an unconscious process, L2 learners often make use of conscious learning strategies. Many accounts of the differences between L1 and L2 acquisition focus on the distinction introduced by Krashen (e.g., 1981b; 1985) of the implicit process of acquisition on the one hand, and explicit learning on the other. Krashen's claim is that only acquisition, which can make no use of explicit instruction, is the basis for automatic language use, i.e., "real" language competence.

In many accounts of L1/L2 differences, AoA is central, the idea being that language acquisition proceeds more naturally and with more ease in young than in older learners. Penfield and Roberts (1959) and Lenneberg (1967) proposed a biologically determined Critical Period for language acquisition, after which native-like competence cannot be achieved based on exposure alone. Later work (e.g., Bley-Vroman, 1988, 1989, 1990), hypothesized that there is a

fundamental difference between acquisition in children and adults. However, children may also be L2 learners, and studies have failed to find evidence of a sharp decline in language acquisition abilities at a certain age. It has been proposed that the period may be better described as *sensitive* (cf. Long, 1990), or that age differences in language acquisition are best explained by affective and/or cognitive factors, for instance, that affective factors in adults impede input-based acquisition (Krashen, Long, & Scarcella, 1982; Rosansky, 1975).

It may even be that there is no specific period during childhood when language acquisition is particularly easy or automatic, but that our natural language acquisition abilities start declining already at birth (Birdsong, 2006; Hyltenstam & Abrahamsson, 2003; Newport, 1988, 1990). If this is the case, an early AoA is still beneficial for ultimate attainment, but the difference is gradual, not fundamental.

There are other reasons to doubt the idea of a "fundamental difference" between L1 and L2 acquisition. For example, while L2s are subject to influence from L1 (cf. Gass & Selinker, 1992; Odlin, 1989), recent research shows that influence may be bidirectional, and that L1 is also permeable to crosslinguistic influence in various areas including phonology, morphosyntax, and the lexicon (Cook, 2003; Pavlenko, 2000). Brown and Gullberg (2008), for example, found bidirectional crosslinguistic influence for manner of motion encoding in Japanese L1 speakers with only intermediate knowledge of English. Furthermore, although there is agreement that the languages of bilinguals are separate systems, there is interaction between them even when children grow up bilingually, for example, in code switching; the linguistic competence of all bilinguals is different from that of monolinguals (Grosjean, 1989; Myers-Scotton, 2005).

It also seems that many of the mechanisms underlying L1 and L2 acquisition and use are essentially the same. For example, all acquisition relies on primary linguistic data, or input (Krashen, 1981b, 1985; Schwartz, 1993). Also statistical learning is likely to be employed in all types of acquisition (Saffran, Newport, & Aslin, 1996; Saffran, Newport, Aslin, Tunick, & Barrueco, 1997), as are mechanisms such as contingency, saliency, and cue competition (Ellis, 2006). In MacWhinney's (2005a, 2005b) Unified Competition Model, sentence comprehension in L1 and L2 alike is based on cue competition (e.g., word order or morphology) for interpreting thematic roles. Acquiring an L2 may partly entail adjusting cue weight from L1 to L2 appropriate settings.

Thus, bilingualism may be described as a continuous rather than categorical variable. The starting point, at least in theory, may be the perfectly balanced bilingual who has acquired both languages with equal exposure from birth. The other end of the continuum may be a speaker who has grown up monolingually and has learned an L2 to a rudimentary level as an adult. Between these two extremes, we can expect various competence levels and contexts of acquisition, and definitions of exactly what counts as bilingual competence may vary.

In this view of bilingualism, age is still an important factor, and a younger AoA is beneficial for ultimate attainment. However, many studies of instructed settings, such as FL classrooms, have found no benefit of starting early; later learners seem to perform as well as younger ones (e.g., Burstall, 1975; Cenoz, 2003; García Lecumberri & Gallardo, 2003; García Mayo & García Lecumberri, 2003; Holmstrand, 1982; Lasagabaster & Doiz, 2003; Muñoz, 2006). It is possible that classroom language learning is too different from naturalistic acquisition for age effects to be relevant, or the findings may reflect the conspiracy of other factors; it may, for example, be that the classroom is more suited to an adult learning style (cf. Muñoz, 2008).

While the basis for L1 acquisition and naturalistic or immersion L2 acquisition is target language exposure, most FL students are also exposed to explicit instruction, the role of which is debated. An intuitive reason to be skeptical of the role of instruction is that it plays no role in L1 acquisition, the most successful type of language acquisition. Empirically, studies investigating the effect of instruction, such as Felix (1981), have found largely invariable orders of acquisition for instructed and non-instructed learners. Theoretically, Krashen's (e.g., 1981a) distinction between learning and acquisition dismisses the role of explicit instruction. In the 1980s, these factors all contributed to communicative approaches, abandoning explicit instruction in favor of language classrooms which imitated naturalistic settings (Lightbown, 2000).

However, such approaches ran into problems. First, the superiority of naturalistic acquisition may depend on more massive exposure to the target language than the FL classroom can possibly provide (cf. Lightbown, 2000; Muñoz, 2006). Second, studies such as Pienemann (1984, 1989) found that although instructed learners generally develop along the same path as naturalistic learners, instruction may speed up this progression. Explicit learning may thus facilitate acquisition, and may to some extent make up for limited input in instructed settings (e.g., Hulstijn, 2002). Studies also indicate that naturalistic learners tend to fossilize short of grammatical accuracy, for instance, in morphology (cf. Ortega, 2009). The most successful L2

learners seem to be those with ample access to input combined with some explicit instruction (Lightbown, 2000; Lightbown & Spada, 1990). The pendulum has thus swung from strictly communicative approaches to a stronger focus on instruction. Still, this does not mean that exposure to the target language can be abandoned as the prime focus of any language learning context. Explicit instruction may be beneficial, but input is absolutely crucial for acquisition. This is probably especially the case for young learners, who most likely excel at implicit acquisition compared to explicit learning (cf. Muñoz, 2006). The question is whether input can ever be substantial enough in a FL setting for acquisition to take place. Dahl and Vulchanova (2014) investigated vocabulary development in the two groups of children reported in the present paper. Findings were that the children exposed to standard Norwegian English teaching did not show significant receptive vocabulary development in the course of their first year of school. However, for the other group, who were exposed to a moderately increased volume of English input compared to standard classroom norms, there was a significant increase in receptive vocabulary. This group's development was found on average to be slightly faster than that of vocabulary-matched, younger L1 acquirers of English in the reference group for the Peabody Picture Vocabulary Test, fourth edition (PPVT-4).

The present study investigates the benefits of this increased target language input not only on receptive vocabulary but also on sentence comprehension and sentence repetition, in order to gain a better picture of what benefits we can expect of increased exposure in the early FL classroom.

2.2. Early second language proficiency

Many models distinguish between two components involved in language use, namely lexical knowledge, which is typically described as structured and is at least partly subserved by declarative memory, and a procedural or combinatorial system for language processing (Clahsen, 1999; Clahsen & Felser, 2006; Pinker, 1999; Ullman, 2001a, 2001b). In naturalistic acquisition, both components develop based on target language exposure, but the declarative aspect of lexical knowledge entails that it can probably also be learned via intentional learning (cf. Schwartz, 1993).

Language processing entails the segmentation of speech into "meaningful chunks" (Clahsen & Felser, 2006) and retrieving the relevant lexical items, i.e., what Carroll (2004) calls

"hearing words ". There is evidence that bilinguals transfer segmentation and lexical access strategies between languages. Cutler, Mehler, Norris, and Segui (1992) found that French-English bilinguals employed either stress-based or syllable-based segmentation procedures in both their languages, and that choice of procedure depended on language dominance. Carroll (2004) suggests that transfer of L1 segmentation strategies may explain why learners have problems parsing the L2. Sebastian-Galles and Diaz (2012) argue that acquiring L2 speech perception entails the reweighting of cues for phoneme discrimination, but that since perception is hierarchical, successful phoneme perception does not guarantee the appropriate lexical representation.

In addition to the complex process of perceiving words, comprehension depends on assigning semantic and syntactic roles for working out the meaning of each sentence (cf. Gibson & Pearlmutter, 1998; Kroll & Dussias, 2004). This process may use a variety of cues, such as word order, morphology, or animacy, depending on language. Again, L2 learners may rely on different information than L1 speakers for role assignment, at least initially. Information and strategies may transfer from the L1 (cf. MacWhinney, 2005a, 2005b; Stringer, 2008). McDonald (1987), for example, found evidence of L1 transfer of role assignment strategies in dative constructions in English learners of Dutch. Other studies have found that L2 learners rely on semantic over syntactic information, for instance, for ambiguity resolution in relative clauses (Clahsen & Felser, 2006; Felser, Roberts, Marinis, & Gross, 2003; Marinis, Roberts, Felser, & Clahsen, 2005; Papadopoulou & Clahsen, 2003). VanPatten (2004, 2012) proposes a number of principles for input processing which L2 learners may initially apply, including processing content words first, and processing the first NP by default as the subject of the sentence.

Crucially, language comprehension need not always entail detailed grammatical analysis, and a range of studies have found evidence that language comprehension in L2 may often be based on lexical-semantic, pragmatic, and non-linguistic knowledge. Based on studies of ambiguity resolution in relative clause attachment and the processing of syntactic (filler-gap) dependencies, Clahsen and Felser (2006) propose the Shallow Structure Hypothesis, suggesting that L2 users' syntactic representations during comprehension are less detailed than those of native speakers. VanPatten (2012) explains misinterpretations of sentences such as passives in L2 in a similar manner, suggesting that strategies for efficient semantic interpretation lead to "good enough for now" processing, where syntactic roles are assigned based on lexical semantics. This

kind of processing may not be idiosyncratic to L2; studies of garden path sentences indicate that also L1 speakers may sometimes rely on "good enough" representations (Ferreira, 2003; Ferreira, Bailey, & Ferraro, 2002). For relative clause disambiguation, Gertken (2013) found evidence of slightly different shallow processing in L1 and L2 speakers.

While shallow processing in L2 learners may not always have observable consequences for comprehension, it will normally result in non-target forms in production. It has been suggested that development of L2 grammatical complexity and accuracy depends on pushed output, form-focused learning, or explicit instruction (Collentine, 2004; Ellis, 2008; Izumi, 2003; Lightbown, 1992, 2000; Lightbown, Halter, White, & Horst, 2002; Swain, 1985, 1995; Swain & Lapkin, 1995).

The present study investigates whether we can expect benefits from increased input in a child FL classroom, and if so, in what areas of language competence. If enough input for naturalistic acquisition can be provided in such settings, benefits may potentially show up in lexical knowledge, abilities in segmentation and lexical retrieval allowing "good enough" representations, and/or in grammatical processing. While evidence of "good enough" processing has been found for adult L2 learners, there are, to my knowledge, no studies of this phenomenon in child SLA. However, it may be that exactly "good enough" comprehension is the aspect which benefits the most from increased target language exposure in young L2 learners. This would follow if an early AoA is particularly beneficial for procedural knowledge, and if naturalistic exposure is less beneficial for accuracy than for fluency.

3. The study

In the present study, two groups of children in their first year in two Norwegian state schools participated. One group (n=29) is referred to as the native-language based group, the other (n=31) as the bilingually-based group. All participants were monolingual speakers of Norwegian just starting to learn English. Both schools were situated in similar, relatively affluent and ethnically homogeneous neighborhoods in the same town. Students from the two schools had previously achieved similar scores on national tests for English, at the national average (bilingually-based group's school) or slightly above it (native-language based group's school). In each school, three different classes and class teachers participated in the project.

English classes in both schools reflected the lack of literacy skills in the children, who were just learning to read in Norwegian. The main difference in the two schools was the role that the English language itself played in the classroom. In one school, teachers taught English as mandated by the Norwegian curriculum (Utdanningsdirektoratet, 2006), with formal English class of about 30 minutes per week, plus some discussion during morning meetings of English words, for topics such as the weather and the days of the week. These discussions used some routine interactions in English, but the meetings themselves were conducted in Norwegian, as was classroom management and instructions during English class. This is the norm in Norwegian first grade classrooms. Total time spent on English per week was approximately 45 minutes. This group used a commonly used English workbook (Bruskeland & Ranke, 2005), which contains simple activities, such as matching pairs, coloring, and rhymes. Associated with each activity is a set of teacher instructions in Norwegian specifying how communication is to take place in English. For example, for an activity where the students are to color a total of 11 pictures of animals, instructions tell the teacher to say "A blue fish. What color is the fish? ". The students are expected to answer "Blue", whereupon the teacher says "Color the fish blue". This is repeated for all 11 animals. This dialogue is an illustration of the English communication taking place in this classroom, and indeed in most English classrooms in Norwegian first grades. This group is referred to as native-language based, since Norwegian was the language of communication in the classroom, and English merely the object of instruction, with very limited communicative use.

In the other school, no changes were made to the English curriculum or time spent on English lessons, but instead of activities conducted in Norwegian with only routine instruction in English as described above, teachers used English for real communication during English class, instead of or in addition to Norwegian. Furthermore, English classes focused on input-heavy activities such as the teacher reading aloud or talking about objects or pictures; no workbook was used. Input was thus not only increased but also more naturalistic in nature than the routine input of the native-language based group. As in the native-language based group, English words were discussed during morning meetings, but these meetings were conducted more or less in English on the part of the teacher, with repetitions in Norwegian where necessary. Also simple classroom management was sometimes carried out in English or bilingually in subjects other than English. This amounted to approximately 70 minutes per week spent focusing on English. It is worth mentioning that one teacher in this school was a native speaker of (British) English, although she

was also completely fluent and taught all other subjects in Norwegian. This group is referred to as bilingually-based, since both the L1 and the L2 were used for communication. However, it is important to point out that Norwegian was the language of instruction in all subjects except English.

One teacher in each school was responsible for reporting to the researcher on time spent on English and on activities and materials used. During the test periods at the beginning and the end of the school year, reports were frequent and relatively informal. In the middle of the spring term, teachers in both groups formally reported on the same three questions (time, activities, and materials) in writing. Consistent reports from both schools throughout the year confirm the patterns described above. Children in the two groups were tested on English receptive vocabulary, sentence comprehension, and sentence repetition after one year of English classes.

It is worth pointing out that the participants' L1, Norwegian, and the target language, English, are relatively closely related languages. They are fairly similar both structurally and in terms of the lexicon, at least for basic vocabulary which tends to be of Germanic origin in English, and such similarity may facilitate acquisition (Lindgren & Muñoz, 2012; Van der Slik, 2010). Furthermore, English is encountered in a range of contexts in Norway, at least for older (literate) children and adults, and this may allow for incidental learning outside of the language classroom. For very young children, however, such exposure to English is limited, and cannot be expected to lead to substantial learning (cf. Dahl & Vulchanova, 2014; Lefever, 2012 ; Lindgren & Muñoz, 2012; Unsworth, Persson, Prins, & De Bot, 2014).

3.1. Research questions

The three research questions were 1) Is it possible that input in an early FL classroom can be substantial enough for an effect to be found on all three language measures? 2) If so, which measure shows the greatest effect of increased input? 3) What can this tell us about the nature of early FL competence?

3.2. Participants

Participants were 60 primary school students, 29 (15 boys, 14 girls) in the native-language based group and 31 (17 boys, 14 girls) in the bilingually-based group. Mean age at the time of pre-testing was 6;1 and mean time between pre- and post-tests was eight months in both groups.

All participants were monolingual speakers of Norwegian, without any known disability or condition which might influence their language acquisition. They lived in similar neighborhoods in the same Norwegian town, and their exposure to English outside of school can thus be assumed to have been similar. In addition to providing written consent for their child to participate, parents completed questionnaires at the beginning of the school year. These included information about the child's hours of exposure to English through music, computer games, and media prior to starting school as well as about weeks spent outside of Scandinavia. Statistics for these measures are presented in the results section.

Each child was tested twice. A pre-test session of control measures was conducted early in the school year (≈September), while English competence was measured in a post-test session late in the year (≈May). Each child was tested alone in a quiet room, with only the researcher and sometimes an assistant present. Each test session lasted for approximately one hour. For practical reasons, only sentence repetition, which could not be scored directly, was audiotaped.

3.3. Control measures

Since we know that L1 skills, working memory, and non-verbal intelligence may influence language acquisition (Colledge et al., 2002; Dale, Harlaar, Haworth, & Plomin, 2010; Foyn, Vulchanova, Nilsen, & Sigmundsson, under revision; Gathercole, 2006; Hayiou-Thomas, Dale, & Plomin, 2012; Sparks, Patton, Ganschow, & Humbach, 2009), control measures were administered to make sure that the groups were not significantly different on such factors. Norwegian receptive vocabulary was tested through a translated version of the PPVT-4, Form A. Verbal and non-verbal intelligence was tested using the Matrices section and a translated version of the Riddles section of the Kaufman Brief Intelligence Test, Second Edition (K-Bit 2). For working memory, the measure was a memory game where the participant memorized sets of pairs of picture cards. After a fixed amount of time, the cards were turned upside-down and the task was to find the pairs in as few attempts as possible. All children were tested on initial English vocabulary through the PPVT-4, Form B, and on English sentence comprehension using an online test (Elllo, 2013a, 2013b).

3.4. Post-test measures of English competence

The post-test of receptive vocabulary consisted of forms A and B of the PPVT-4, which is normed for native speakers of English aged 2;6–90. The subject hears a word, and is asked to select from a set of four pictures the one which matches the word. Words come in consecutive sets of 12. For each set, the subject must respond correctly to at least 5 words in order to continue on to the next set. Words get increasingly difficult between sets throughout the test, and when testing is discontinued, it is unlikely that the subject will know any words in subsequent sets. The two forms are highly correlated for English native speakers. However, for L2 learners, there is no norming, and results may be less reliable, since cognate status with the L1 may complicate the assumed difficulty levels. If more cognates in a later set makes it easier than the previous set, the difference between making 7 errors in a set and be allowed to move on, and making 8 and be discontinued, would lead to a larger difference in scores than when the test is used for L1. For this reason, both forms of the test were administered and vocabulary scores reported are the means of both forms. This procedure reduced the number of outliers from three and seven for forms A and B respectively, to only two, one in each participant group. This indicates that children did not score highly by chance on both sets, and that the means contributed to normalizing scores to realistically reflect each child's competence.

The sentence comprehension test was produced for this study, and consisted of 30 sentences, each accompanied by a set of four pictures from which the subject was asked to select the corresponding one (see Appendix 1). The test used basic, frequent vocabulary expected to be acquired early. Identifying the matching image thus minimally entailed segmenting the sentence and retrieving items from the mental lexicon. For this test, further analysis could probably be fairly simple. There were 28 declarative sentences, and two questions; one yes/no question and one wh-question. All sentences were relatively short, ranging from four to eight words and from five to ten syllables. One sentence was ditransitive, but no other sentence contained more than two verbal arguments, sometimes with one additional adverbial. No VP contained more than one auxiliary, and no syntactic element consisted of more than three words, with the exception of two sentences where the PP "in this picture" modified the subject noun.

At least two elements had to be understood in each sentence for the correct picture to be identified. Only one picture completely corresponded to the sentence, while two pictures showed elements corresponding to part of the sentence, and one picture was completely unrelated. The

crucial information might be in different syntactic elements, such as subject and direct object or adverbial, or both pieces of information might be in the same phrase, for example, in a noun and its adjectival modifier. For example, for the question "Can you see the yellow flower?" there was, in addition to a picture of a yellow flower, also a purple flower and a yellow rabbit. Thus, the child had to understand both the adjective and the noun in order to identify the correct picture. The last picture was of a completely unrelated gray cat. For the sentence "This elephant is playing with a ball" there was, in addition to the target picture and the unrelated distractor, also a picture of an elephant without a ball, and a yellow emoticon playing with a ball. The test was computerized, and sentences were read alternately by a native speaker of American English and a native speaker of British English; both speakers were female.

It is likely that correct answers in this test could be achieved without detailed syntactic analysis. Comprehension did not hinge on role assignment, since no pictures in a set differed only in thematic relationship, for instance, in who was agent or patient. The test sentences also did not contain ambiguities where the correct (grammatical) reading was semantically more implausible than another reading, and thus there was no reason for conflict between semantic and syntactic cues. The reason is that English and Norwegian are very similar in the relevant respects. There is a difference in that Norwegian has no subject-verb agreement, whereas English does, but the main cue for syntactic function in both languages is word order. For the sentence types in question, word order is generally SVO for both languages[1]. Thus, there is no way to tell what strategies participants used for role assignment. Whether L2 learners universally interpret the first noun of the sentence as the subject (cf. VanPatten, 2012), or processing strategies are transferred from the L1 (cf. MacWhinney, 2005a; MacWhinney, 2005b), there is no reason to assume that early Norwegian learners of English should have a particular problem in assigning roles in simple English sentences.

We thus have no way of deciding whether successful sentence comprehension was based on only shallow or "good enough" representations. However, we do know that to reach the point where such representations are possible, the participant would need to segment the speech stream and access the relevant lexical items. Segmentation may be facilitated by the fact that both

[1]Norwegian is a V2 language, with SVO word order in subject-initial clauses. All the declarative sentences in this test were subject-initial, and word order was thus the same as it would have been in Norwegian. Word order in the two questions would also be the same.

Norwegian and English are stress-timed languages, but phonology is different both in phoneme inventory, phonetic realizations, and phonotactics. We can thus assume that the sentence comprehension test minimally measures the participants' ability to "hear words" in the L2 (e.g., Carroll, 2004).

Another test type was employed to investigate more detailed grammatical processing, namely a sentence repetition test consisting of 17 sentences ranging from two to 13 syllables and of various complexity (see Appendix 2). Again, sentences were not ambiguous, nor structurally very different from their Norwegian translations. However, the test evaluated reproduction of morphology. Since sentence repetition is assumed to require the processing, reconstruction and reproduction of the stimulus sentence (cf. Jessop, Suzuki, & Tomita, 2007), it can measure grammatical processing in L2 even when the L1 and the L2 are closely related and syntactically similar. Sentence repetition also requires production (Van Moere, 2012), and errors may be related not only to comprehension. However, many of the differences between comprehension and (elicited) production can be said to be exactly in processing load. As long as accuracy in pronunciation is not evaluated, sentence repetition can thus be assumed to be a measure of language processing, although errors may reflect problems in processing related to either comprehension, reconstruction, or reproduction specifically. The present test was evaluated and found to be reliable and valid for both longer and shorter sentences (Dahl, n.d.).

The sentences were read by the researcher who is a proficient L2 speaker of American English. The session was recorded, and the child's responses were scored by two independent raters, both university students of English linguistics. One was a monolingual native speaker of American English, the other a bilingual native speaker of Norwegian and Dutch, also proficient in American English. The raters were instructed to award one point per correctly repeated morpheme (including zero morphemes; see Appendix 2). Importantly, instructions were not to penalize non-native pronunciation as long as the morpheme was recognizable. Interrater reliability was assessed using Intraclass Correlation Coefficient (ICC). ICC was .796 and .900 in the bilingually-based and the native-language based groups, respectively, which is considered good.

4. Results

In the following, results of all tests administered in the study are presented. First, descriptive and inferential statistics are reported for the control measures, followed by descriptive and inferential statistics for the three post-test measures of English. The latter answer the two first research questions, namely whether there is a significant difference between groups on all three measures, and which measure shows the greatest effect of input. Since the greatest effect is found for sentence comprehension, a multiple regression analysis is reported next comparing the effect of input on this measure to potentially influential background factors.

4.1. Control measures

Table 1 shows descriptive statistics for all control measures, including tests of initial English comprehension.

Table 1: Descriptive statistics, control measures.

	Mean		Range		Std. Deviation	
	Native based	Bilingually based	Native based	Bilingually based	Native based	Bilingually based
Hours of media etc. exposure	1.19	1.75	0-7	0-11	1.64	2.50
Weeks outside Scandinavia	6.38	4.23	0-24	0-14	6.80	3.63
Norwegian vocabulary	113.79	119.90	85–145	97–157	14.56	14.78
K-Bit Riddles	16.10	16.58	10–26	11–24	4.72	3.95
K-Bit Matrices	17.69	18.74	11–38	12–32	5.25	4.61
Memory game attempts	47.97	46.29	40–62	40–59	5.94	4.56
Pre-test English vocabulary	23.72	25.39	3–61	2–56	13.55	11.38
Pre-test English sentences	5.24	5.65	2–10	3–8	1.92	1.54

Multivariate Analysis of Variance (MANOVA) was carried out for all control measures to check for differences between groups on the control measures. Because of concerns about the normality of the distribution of data[2], bootstrapping was performed based on 1000 samples to obtain robust significance levels. The MANOVA showed no overall effect of group on the control measures, $F(8, 51) = 1.466$, $p = .193$.

[2] Kolmogorov-Smirnov tests found deviations from the normal distribution for one or both groups on a number of measures.

There is an indication of some, albeit very limited, English comprehension in the pre-test. Both groups' scores on English vocabulary and sentence comprehension are low enough that they can probably be explained mainly by cognate comprehension, possibly combined with some incidental learning from English encountered outside of the classroom.

4.2. Post-test English

Descriptive statistics for the three English measures in the post-test are provided in Table 2.

Table 2: Descriptive statistics, post-test English measures.

	Mean		Range		Std. Deviation	
	Native based	Bilingually based	Native based	Bilingually based	Native based	Bilingually based
Vocabulary	31.50	42.18	11.50 – 82.50	20.50 – 79.00	16.65	15.04
Sentence comprehension	14.38	18.74	6 – 27	11 – 25	4.40	3.54
Sentence repetition	85.07	93.92	45.75 – 114.50	58.00 – 117.00	20.08	13.88

In order to investigate the effect of target language exposure on all measures of English competence, a MANOVA was carried out for all three tests. Bootstrapping was again performed based on 1000 samples. The analysis showed that there was an overall effect of group on the measures combined, $F(3, 56) = 6.2$, $p<.01$. More interesting for our purposes is the difference between the two groups on each measure separately, which was investigated by separate t-tests. Again, bootstrapping was performed based on 1000 samples. The results are reported in Table 3. Levene's Test for Equality of Variances was applied, and where appropriate, values not assuming equality of variances are reported.

Table 3: Between-group t-tests, post-test English measures. Confidence intervals and standard errors based on 1000 bootstrap samples.

			Bootstrapped				
	t	df	*p* (one-tailed)	std. error	BCa 95% Confidence Interval		effect size *r*
					Lower	Upper	
Vocabulary	2.6	58	.007	4.28	2.55	18.60	.32
Sentence comprehension	4.2	58	.001	0.99	2.34	6.23	.48
Sentence repetition	2.0	49.4	.025	4.34	0.46	17.45	.27

Note: Equal variances assumed for Vocabulary and Sentence Comprehension. Equal variances not assumed for Sentence repetition.

As Table 3 indicates, there was a significant difference between the two groups on all three English measures. The effect size is smallest for sentence repetition, larger for vocabulary, and clearly largest for sentence comprehension.

Since the greatest difference between the two groups was found for sentence comprehension, the meaningfulness of this effect was investigated by checking how well sentence comprehension score is predicted by group membership compared to other potentially influential factors. Among the control measures, exploratory correlation analyses[3] showed that only two were significantly correlated with English sentence comprehension, namely Norwegian vocabulary and pre-test English vocabulary. Multiple regression was used to assess the effect of these two measures when the effect of Group (bilingually-based or native-language based) had been accounted for. The analysis was run with forced entry, with Group as the only variable in block 1, and Norwegian vocabulary and initial English vocabulary in block 2. Because of some concern about inequality of variances, bootstrapping was again performed.

[3] Pearson correlation with bootstrap based on 1000 samples.

Table 4: Linear model of predictors of sentence comprehension. Confidence intervals and standard errors based on 1000 bootstrap samples.

	b	SE B	Beta	*p*
Step 1				
Constant	10.017	1.640		*p* = .001
BCa 95% Confidence Interval	7.015 – 13.056			
Group	4.363	0.990	.487	*p* =.001
BCa 95% Confidence Interval	2.199 – 6.307			
Step 2				
Constant	1.975	4.198		*p* =.656
BCa 95% Confidence Interval	-6.263 – 10.310			
Group	3.866	0.984	.431	*p* =.002
BCa 95% Confidence Interval	1.874 – 5.890			
Norwegian vocabulary	0.055	0.041	.180	*p* =.192
BCa 95% Confidence Interval	-0.024 – 0.143			
Initial English vocabulary	0.098	0.045	.269	*p* =.036
BCa 95% Confidence Interval	0.012 – 0.178			

Note. $R^2 = 0.24$ for Step 1 ($p<.001$); $\Delta R^2 = .14$ for Step 2 ($p=.002$).

We see that group membership (i.e., volume of target language exposure) accounts for 24% of the variance in sentence comprehension scores, while the two variables added in step 2 combined account for only approximately 14%. Out of the two variables added in Step 2, only initial English vocabulary significantly contributes to the model, while the contribution of Norwegian vocabulary is non-significant.

5. Discussion

The results section provided answers to the first two research questions. Firstly, input in the bilingually-based group seems to have been substantial enough to have an effect on all three language measures in the course of one year of school. This is an important finding, since it indicates that input in FL classrooms is not necessarily too impoverished for naturalistic acquisition in young learners. The bilingually-based group's L2 exposure was much less substantial than what one can expect in naturalistic or immersion settings, but still seems to have been sufficient for some acquisition to take place. As Muñoz (2006) points out, implicit learning

is a slow process, and more massive input might have led to more rapid acquisition. However, the fact that the difference between groups is not zero indicates that children in the bilingually-based group were able to make use of such learning even with the modest input to which they were exposed. Teachers from both groups reported similar focus on explicit vocabulary learning and on students' production; the difference was that the bilingually-based group was exposed to language in context. Yet, there is a difference between the two groups also for receptive vocabulary. This is not surprising, since increased exposure will have resulted in an increase in the quality of input, and thus in an increase both in number and frequency of lexical items encountered, both of which are likely to be important for acquisition.

Secondly, the greatest effect was found on sentence comprehension. This effect was substantial, and stood out from the more similar effect sizes of receptive vocabulary and sentence repetition. This is consistent with Carroll's (2004) proposal that the very first stages of SLA consist of acquiring segmentation skills, and leads us to the third research question, namely what the results can tell us about early child FL learning.

The greater effect on sentence comprehension than on receptive vocabulary may reflect a greater benefit of a young AoA on aspects of language which depend solely on the procedural system. For example, Ullman (2001b) proposes that younger language learners may rely predominantly on procedural memory, while declarative memory improves with age. The results may thus indicate a benefit of increased exposure on the ability to perceive and access lexical items in particular. The weaker effect on sentence repetition may be an indication of limited development of morphology, for example, verb inflection, which is a cue to role assignment in English. This benefit on a measure which requires lexical processing in particular may thus indicate that the advantage of the bilingually-based group is in particular in shallow processing or the ability to build "good enough" representations (cf. Clahsen & Felser, 2006; Ferreira et al., 2002; VanPatten, 2012). The effects of such shallow processing cannot be directly observed in the comprehension test, since the sentences can be accurately comprehended whether role assignment is based on transfer from the L1, on universally assigning the role of subject to the first NP, or on L2 appropriate strategies. However, the difference on sentence repetition is small enough that it can probably be explained by the advantage of the bilingually-based group in hearing and reproducing lexical words, an assumption supported by very similar differences between the groups on repetition of shorter and longer sentences. This supports the hypothesis

that the advantage in the bilingually-based group on sentence comprehension is a result of shallow processing. It is likely that younger L2 learners are slower than older learners in particular in the acquisition of morphosyntax (Krashen et al., 1982). Grammatical development may also depend on more exposure, or possibly even explicit instruction or pushed output (Ellis, 2008; Izumi, 2003; Lightbown, 1992, 2000; Lightbown et al., 2002; Swain, 1985, 1995; Swain & Lapkin, 1995).

A detailed look at performance on individual sentences in the comprehension test does not reveal group differences connected to specific sentence types. The bilingually-based group outperformed the native-language based group on 26 out of 30 sentences. For three sentences (sentences 2, 29, and 30), the mean score in the bilingually-based group is 1, meaning that all children responded correctly. No sentence achieved this score in the native-language based group. It is not apparent what these three sentences have in common which makes them easy for the bilingually-based group. They range in length from seven to 10 syllables, and have either SVA or SVO structure. They thus share both length and structure with other sentences in the test. Similarly, there are no obvious similarities between the four sentences on which the native-language based group's mean is higher than that of the bilingually-based group (sentences 8, 10, 13 and 15). Three of them are among the more complex sentences; in sentence 8, the subject contains the PP "in this picture", sentence 10 is a question, and sentence 15 contains existential *there*. However, there was another sentence in the test of each of these types, on which the bilingually-based group outperformed the native-language based group. Sentence 13, where performance is only slightly higher in the native-language based group, is among the shortest sentences in the test. It contains the auxiliary *has*, but so does another sentence where performance is higher in the bilingually-based group. It is therefore difficult to find syntactic explanations for group performance.

For the four sentences where the native-language based group performed better, a theoretically interesting question is whether this is a result of a particularly high score in this group or of a particularly low score in the bilingually-based group. However, again, there is no clear pattern. For sentences 8 and 10, the mean score in the native-language based group is .45 and .48, respectively, indicating that almost half of the participants responded correctly. In the bilingually-based group, the scores for the same sentences are relatively low, with means of .19 and .35, respectively. For sentence 13, a little over half of the participants in each group

responded correctly, with means of .59 for the native-language based and .58 for the bilingually-based group. Finally, for sentence 15 performance was high in both groups, with a mean of .90 in the native-language based group, and .71 in the bilingually-based group. Consequently, we can draw no conclusions about why the native-language based group performed better on these particular sentences.

There are no other obvious factors which influence comprehension. There do not seem to be any systematic differences in performance between odd-numbered sentences, which were read by the American speaker, and even-numbered sentences, which were read by the British speaker. This is particularly relevant since the native speaker teacher in the bilingually-based group was British. Furthermore, the native-language based group's lower performance results both from selecting pictures corresponding to only one word in the sentence more often than the bilingually-based group, and from more often making completely wrong choices where no word corresponded to the picture. Children in the native-language based group selected sentences with one corresponding word on average for approximately 13 out of the 30 sentences, while the bilingually-based group did so on average for approximately 11 sentences. The native-language based group selected the completely wrong picture on average for almost 3 out of 30 sentences, while the bilingually-based group made such errors on average for less than one sentence.

The advantage of the bilingually-based group on sentence comprehension thus seems to be general, and we can assume that it has come about based on comprehensible input (e.g., Krashen, 1981a; 1985). Even if grammatical detail has not been processed, input must have been processed as *intake* (Corder, 1967), i.e., must have been *comprehended* in the sense of Gass (1997), since only such input leads to acquisition.

Izumi (2003) suggests that one reason for why learners at low proficiency levels are unable to attend to grammatical form in comprehension and production alike may be that lexical retrieval is taking up too many cognitive resources. The improved comprehension in the bilingually-based group may thus free up resources to attend to form in the next stage. Comprehension may also be a motivator for further acquisition. As Segalowitz and Trofimovich (2012, p. 187) put it, "the acquisition of L2 processing skills along the cognitive dimension [...] is likely to increase a person's motivation to engage in L2 contact compared to a person whose efficiency of L2 cognitive processing remains low". Improved processing may thus lead to "[a] number of self-reinforcing loops [...], all leading to increased L2 processing skills" (Segalowitz &

Trofimovich, 2012, p. 187). The better comprehension of the bilingually-based group after only eight months as compared to the native-language based group may thus have an effect on subsequent acquisition, at least as long as the volume of exposure is maintained.

It is worth pointing out that input in the bilingually-based group in the study was not only more substantial and more varied than that of the native-language based group, but that it was also largely provided by a native speaker of English. This may have contributed to the results. There is, however, little reason to assume that input must be provided by a native speaker to be beneficial; Unsworth et al. (2014), for example, found an effect of teacher competence on acquisition in a group of young Dutch learners, but no effect of nativeness. No analysis of teacher competence was conducted in the present study, but the presence of a native speaker as well as the fact that also the other teachers in the bilingually-based group agreed to participate in the project, i.e., in providing more input, means that it is not unlikely that higher proficiency in this group's teachers was important for the students' development. It is also likely that reliance on the textbook and on routine interactions in English in the native-language based group was precisely a result of lower English proficiency in this group's teachers. Teachers' competence, fluency, and self-confidence in the target language can be expected to be necessary for increased input in the classroom to be effective.

6. Conclusion

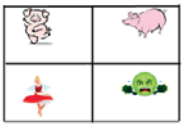
The results of the present study indicate that increased exposure within the limits of a normal FL curriculum can lead to acquisition in young learners during one year of school, at least when the L1 and the target language are similar. It thus seems that there is no absolute threshold, unattainable in the FL classroom, for how much input young learners need for implicit acquisition, and it is likely that any increase in target language exposure is beneficial. The greatest benefit of increased input was found in sentence comprehension, which may in turn be beneficial for further acquisition. The magnitude of the effect of input on sentence comprehension indicates that the difference between the two groups is not only theoretically significant but also practically relevant. Early FL learning is increasingly common, and it is essential that we know what constitutes effective learning in such classrooms. It seems that even in a FL setting, and even when input cannot be very substantial, it can lead to successful acquisition of important language competence in young learners.

There are limitations to the present study, especially in that it investigated learners from only two schools, with a relatively low number of participants. It could also be argued that the presence of a native speaker teacher in one of the conditions is a limitation. It is clear that more research is needed on the effect of input in early FL, including more groups of learners and different teachers, and more longitudinal study. Also, the similarity of the participants' L1 and L2 in the present study prevents us from drawing conclusions about exactly what the improved sentence comprehension in the bilingually-based group means, and studies of more language pairs are therefore needed.

Appendix 1

Sentence comprehension test

Trial 1: The chair is green



Trial 2: The pig is dancing

1. The girl is eating an apple

2. The boy is out in the rain

3. There are trees outside the house

4. The man is wearing a white shirt

5. This boy has a nice bird

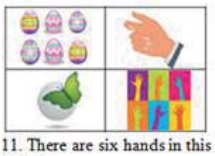
6. The little girl is jumping

7. The car in this picture is red

8. The people in this picture are happy

9. The boy is very sad

10. Where is the black shoe?

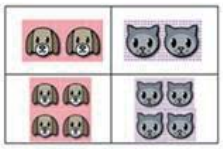


11. There are six hands in this picture

12. The girl really likes to sing

13. She has washed the dog

14. Daddy is reading to the boy



15. There are four cats in this picture

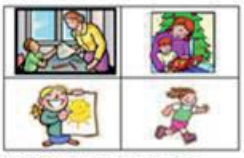
16. This funny rabbit has pink ears.

17. Can you see the yellow flower?

18. This elephant is playing with a ball.

19. The teacher is very old.

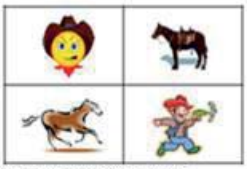
20. The mother is feeding the baby.



21. The boy has drawn his mother a picture.

22. The cow jumped over the moon.

23. Three children are playing.



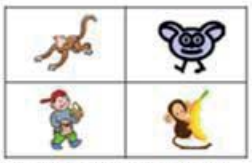
24. The horse is running.

25. She is holding the baby.

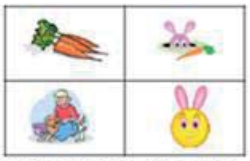
26. He is sleeping under a tree.

27. They are playing in the water.

28. The girl is holding the cat.



29. The monkey has a banana.



30. The rabbit is looking at the carrot.

Appendix 2

Repetition test and scoring guide

Trials

Pen
Fish
Eight
Shoulder
Pencil
Seven
I am a boy/girl.
You are three years old.
Is it raining outside?

Test sentences and scoring guide

Sentence	*Morphemes*	*Total*
1. Birds fly.	bird+s fly+Ø	4
2. Horses run fast.	Horse+s run+Ø fast	5
3. The baby drinks milk.	the baby+Ø drink+s milk+Ø	7
4. She is jumping.	she+x is+x jump+ing	6
5. He is eating a banana.	he+x is+x eat+ing a/ banana+Ø	9
6. She has washed the dog.	she+x has+x wash+ed the dog+Ø	9
7. They have played.	they+x have+x play+ed	6
8. He is not young.	he+x is+x not young	6
9. Cats do not fly.	cat+s do+Ø not fly+Ø	7
10. Is the elephant playing?	is+x the elephant+Ø play+ing	7
11. Does she sing?	do+es she+x sing+Ø	6
12. Do you like to read?	do+Ø you+x like+Ø to read+Ø	9
13. Are the children at school?	are+x the child+ren at school+Ø	8
14. The cat is playing with a ball.	the cat+Ø is+x play+ing with a/ ball+Ø	11
15. The children's dog always sleeps on the floor.	the child+ren+'s dog+Ø always sleep+s on the floor+Ø	13
16. The boy has been playing with his friends for an hour.	the boy+Ø has+x be+en play+ing with his+x friend+s for an/ hour+Ø	18
17. The house with the blue door looks nice.	the house+Ø with the blue door+Ø look+s nice	11
Total:		142

Ø = point for zero morpheme
x = 1 point for auxiliary/pronoun
/ = 0.5 point for indefinite article, 0.5 for correct form

References

Birdsong, D. (2006). Age and second language acquisition and processing: A selective overview. *Language Learning, 56*(S1), 9-49. doi: 10.1111/j.1467-9922.2006.00353.x

Bley-Vroman, R. (1988). The fundamental character of foreign language learning. In W. Rutherford & M. S. Smith (Eds.), *Grammar and second language teaching* (pp. 19-30). Rowley, MA: Newbury House.

Bley-Vroman, R. (1989). What is the logical problem of foreign language learning? In S. Gass & J. Schachter (Eds.), *Linguistic Perspectives on Second Language Acquisition* (pp. 41-68). Cambridge, MA: Cambridge University Press.

Bley-Vroman, R. (1990). The logical problem of foreign language learning. *Linguistic analysis, 20*(1-2), 3-49.

Brown, A., & Gullberg, M. (2008). Bidirectional crosslinguistic influence in L1-L2 encoding of manner in speech and gesture: A study of Japanese speakers of English. *Studies in Second Language Acquisition, 30*(2), 225-251. doi: 10.1017/S0272263108080327

Bruskeland, P. A., & Ranke, C. T. (2005). *Junior Scoop 1-2*. Oslo: Samlaget.

Burstall, C. (1975). French in the primary school: The British experiment. *Canadian Modern Language Review, 31*, 388-402.

Carroll, S. E. (2004). Segmentation: Learning how to 'hear words' in the L2 speech stream. *Transactions of the Philological Society, 102*(2), 227-254. doi: 10.1111/j.0079-1636.2004.00136.x

Cenoz, J. (2003). The influence of age on the acquisition of English: General proficiency, attitudes and code-mixing. In M. García Mayo & M. García Lecumberri (Eds.), *Age and the acquisition of English as a foreign language* (pp. 77-93). Clevedon: Multilingual Matters.

Clahsen, H. (1999). Lexical entries and rules of language: A multidisciplinary study of German inflection. *Behavioral and Brain Sciences, 22*(6), 991-1013; discussion 1014-1060.

Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics, 27*(1), 3-42. doi: 10.1017/S0142716406060024

Colledge, E., Bishop, D. V., Koeppen-Schomerus, G., Price, T. S., Happé, F. G., Eley, T. C., . . . Plomin, R. (2002). The structure of language abilities at 4 years: A twin study. *Developmental Psychology, 38*(5), 749-757. doi: 10.1037//0012-1649.38.5.749

Collentine, J. (2004). The effects of learning contexts on morphosyntactic and lexical development. *Studies in Second Language Acquisition, 26*(2), 227-248. doi: 10.1017/S0272263104262040

Cook, V. (2003). *Effects of the second language on the first*. Clevedon: Multilingual Matters.

Corder, P. (1967). The significance of learners' errors. *International Review of Applied Linguistics, 5*, 161-169. doi: 10.1515/iral.1967.5.1-4.161

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1992). The monolingual nature of speech segmentation by bilinguals. *Cognitive Psychology, 24*(3), 381-410. doi: 10.1016/0010-0285(92)90012-Q

Dahl, A. (n.d.). *Sentence repetition as a measure of early L2 development*. Trondheim, Norway.

Dahl, A., & Vulchanova, M. (2014). Naturalistic acquisition in an early language classroom. *Frontiers in Psychology. Special issue. Learning a non-native language in a naturalistic environment: Insights from behavioural and neuroimaging research., 5*, 1-9.

Dale, P. S., Harlaar, N., Haworth, C. M., & Plomin, R. (2010). Two by two: A twin study of second-language acquisition. *Psychological Science, 21*(5), 635-640. doi: 10.1177/0956797610368060

Ellis, N. C. (2006). Selective attention and transfer phenomena in L2 acquisition: Contingency, cue competition, salience, interference, overshadowing, blocking, and perceptual learning. *Applied Linguistics, 27*(2), 164-194. doi: 10.1093/applin/aml015

Ellis, N. C. (2008). The dynamics of second language emergence: Cycles of language use, language change, and language acquisition. *The Modern Language Journal, 92*(2), 232-249. doi: 10.1111/j.1540-4781.2008.00716.x

Elllo. (2013a). Listening Games: Fruit. Retrieved 08.29.13, 2013, from http://www.elllo.org/english/Games/G008-Fruit.html

Elllo. (2013b). Listening Games: Pets. Retrieved 08.29.13, 2013, from http://www.elllo.org/english/Games/G039-Pets.html

Felix, S. W. (1981). The effect of formal instruction on second language acquisition. *Language Learning, 31*(1), 87-112. doi: 10.1111/j.1467-1770.1981.tb01374.x

Felser, C., Roberts, L., Marinis, T., & Gross, R. (2003). The processing of ambiguous sentences by first and second language learners of English. *Applied Psycholinguistics, 24*(3), 453-489. doi: 10.1017.S0142716403000237

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology, 47*(2), 164-203. doi: 10.1016/S0010-0285(03)00005-7

Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science, 11*(1), 11-15. doi: 10.1111/1467-8721.00158

Foyn, C., Vulchanova, M., Nilsen, R. A., & Sigmundsson, H. (under revision). Links between phonological memory, first language competence and second language competence in 10 year-old children. *Learning and Individual Differences*.

García Lecumberri, M. L., & Gallardo, F. (2003). English FL sounds in school learners of different ages. In M. d. P. García Mayo & M. L. García Lecumberri (Eds.), *Age and the acquisition of English as a foreign language* (pp. 115-135). Clevedon: Multilingual Matters.

García Mayo, M. d. P., & García Lecumberri, M. L. (Eds.). (2003). *Age and the acquisition of English as a foreign language*. Clevedon: Multilingual Matters.

Gass, S. M. (1997). *Input, interaction, and the second language learner*. Mahwah, NJ: Lawrence Erlbaum Associates.

Gass, S. M., & Selinker, L. (1992). *Language transfer in language learning*. Amsterdam: J. Benjamins.

Gathercole, S. E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics, 27*(4), 513-613. doi: 10.1017/S0142716406060383

Gertken, L. M. (2013). *Priming of relative clause attachment during comprehension of French as a first and second language.* (PhD thesis), University of Texas at Austin, Austin, TX.

Gibson, E., & Pearlmutter, N. J. (1998). Constraints on sentence comprehension. *Trends in Cognitive Sciences, 2*(7), 262-268. doi: 10.1016/S1364-6613(98)01187-5

Grosjean, F. (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language*(36), 3-15. doi: 10.1016/0093-934X(89)90048-5

Hayiou-Thomas, M. E., Dale, P. S., & Plomin, R. (2012). The etiology of variation in language skills changes with development: A longitudinal twin study of language from 2 to 12 years. *Developmental Science, 15*(2), 233-249. doi: 10.1111/j.1467-7687.2011.01119.x

Holmstrand, L. S. E. (1982). *English in the elementary school: Theoretical and empirical aspects of the early teaching of English as a foreign language*. Uppsala: Acta Universitatis Upsaliensis.

Hulstijn, J. (2002). Towards a unified account of the representation, processing and acquisition of second language knowledge. *Second Language Research, 18*(3), 193-223. doi: 10.1191/0267658302sr207oa

Hyltenstam, K., & Abrahamsson, N. (2003). Maturational constraints in SLA. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 539-588). Oxford: Blackwell.

Izumi, S. (2003). Comprehension and production processes in second language learning: In search of the psycholinguistic rationale of the Output Hypothesis. *Applied Linguistics, 24*(2), 168-196. doi: 10.1093/applin/24.2.168

Jessop, L., Suzuki, W., & Tomita, Y. (2007). Elicited imitation in second language acquisition research. *The Canadian Modern Language Review/La Revue canadienne des langues vivantes, 64*(1), 215-238. doi: 10.3138/cmlr.64.1.215

Krashen, S. D. (1981a). Aptitude and attitude in relation to second language acquisition and learning. In K. C. Diller (Ed.), *Individual differences and universals in language learning aptitude* (pp. 155-175). Rowley, MA: Newbury House.

Krashen, S. D. (1981b). *Second language acquisition and second language learning*. Oxford: Pergamon Press.

Krashen, S. D. (1985). *The Input Hypothesis: Issues and implications*. New York, NY: Longman.

Krashen, S. D., Long, M. H., & Scarcella, R. C. (1982). *Child-adult differences in second language acquisition*. Rowley, MA: Newbury House.

Kroll, J. F., & Dussias, P. E. (2004). The comprehension of words and sentences in two languages. In T. K. Bhatia & W. C. Ritchie (Eds.), *The handbook of bilingualism* (pp. 169-200). Malden, MA: Blackwell.

Lasagabaster, D., & Doiz, A. (2003). Maturational constraints on foreign-language written production. In M. L. García Lecumberri & M. d. P. García Mayo (Eds.), *Age and the acquisition of English as a foreign language* (pp. 136-160). Clevedon: Multilingual Matters.

Lefever, S. (2012). Incidental foreign language learning in young children. In A. Hasselgreen, I. Drew & B. Sørheim (Eds.), *The young language learner : research-based insights into teaching and learning* (pp. 87-99). Bergen: Fagbokforlaget.

Lenneberg, E. H. (1967). *Biological foundations of language*. New York, NY: Wiley.

Lightbown, P. M. (1992). Can they do it themselves? A comprehension-based ESL course for young children. In R. J. Courchene, J. St John, C. Thérien & J. I. Glidden (Eds.), *Comprehension-based second language teaching* (pp. 353-370). Ottawa: University of Ottawa Press.

Lightbown, P. M. (2000). Classroom SLA research and second language teaching. *Applied Linguistics, 21*(4), 431-462. doi: 10.1093/applin/21.4.431

Lightbown, P. M., Halter, R., White, J., & Horst, M. (2002). Comprehension-based learning: The limits of "Do It Yourself". *The Canadian modern language review, 58*(3), 427-464. doi: 10.3138/cmlr.58.3.427

Lightbown, P. M., & Spada, N. (1990). Focus-on-Form and corrective feedback in communicative language teaching. *Studies in Second Language Acquisition, 12*(04), 429-448. doi: 10.1017/S0272263100009517

Lindgren, E., & Muñoz, C. (2012). The influence of exposure, parents, and linguistic distance on young European learners' foreign language comprehension. *International Journal of Multilingualism, 10*(1), 105-129. doi: 10.1080/14790718.2012.679275

Long, M. H. (1990). Maturational constraints on language development. *Studies in Second Language Acquisition, 12*(3), 251-285. doi: 10.1017/S0272263100009165

MacWhinney, B. (2005a). Extending the Competition Model. *International Journal of Bilingualism, 9*(1), 69-84. doi: 10.1177/13670069050090010501

MacWhinney, B. (2005b). A unified model of language acquisition. In A. M. B. de Groot (Ed.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 49-67). Cary, NC: Oxford University Press.

Marinis, T., Roberts, L., Felser, C., & Clahsen, H. (2005). Gaps in second language sentence processing. *Studies in Second Language Acquisition, 27*(1), 53-78. doi: 10.1017/S0272263105050035

McDonald, J. L. (1987). Sentence interpretation in bilingual speakers of English and Dutch. *Applied Psycholinguistics, 8*(4), 379-413.

Muñoz, C. (2006). The BAF Project: Research on the effects of age on foreign language acquisition. In C. Abello-Contesse, R. Chacón-Beltrán, M. D. López-Jiménez & M. M. Torreblancaa-López (Eds.), *Age in L2 acquisition and teaching* (pp. 81-92). Bern: Peter Lang.

Muñoz, C. (2008). Symmetries and asymmetries of age effects in naturalistic and instructed L2 learning. *Applied Linguistics, 29*(4), 578-596. doi: 10.1093/applin/amm056

Myers-Scotton, C. (2005). Supporting a differential access hypothesis: Code switching and other contact data. In A. M. B. de Groot (Ed.), *Handbook of Bilingualism: Psycholinguistic Approaches* (pp. 326-348). New York, NY: Oxford University Press.

Newport, E. L. (1988). Constraints on learning and their role in language acquisition: Studies of the acquisition of American Sign Language. *Language Sciences, 10*(1), 147-172. doi: 10.1016/0388-0001(88)90010-1

Newport, E. L. (1990). Maturational constraints on language learning. *Cognitive science, 14*(1), 11-28. doi: 10.1016/0364-0213(90)90024-Q

Odlin, T. (1989). *Language transfer: Cross-linguistic influence in language learning*. Cambridge: Cambridge University Press.

Ortega, L. (2009). *Understanding second language acquisition*. London: Hodder Education.

Papadopoulou, D., & Clahsen, H. (2003). Parsing strategies in L1 and L2 sentence processing: A study of relative clause attachment in Greek. *Studies in Second Language Acquisition, 25*(4), 501-528. doi: 10.1017/S0272263103000214

Pavlenko, A. (2000). L2 influence on L1 in late bilingualism. *Issues in Applied Linguistics, 11*(2), 175-205.

Penfield, W., & Roberts, L. (1959). *Speech and brain-mechanisms*. Princeton, NJ: Princeton University Press.

Pienemann, M. (1984). Psychological constraints on the teachability of languages. *Studies in Second Language Acquisition, 6*(02), 186-214. doi: 10.1017/S0272263100005015

Pienemann, M. (1989). Is language teachable? Psycholinguistic experiments and hypotheses. *Applied Linguistics, 10*(1), 52-79. doi: 10.1093/applin/10.1.52

Pinker, S. (1999). *Words and rules: The ingredients of language*. London: Weidenfeld & Nicolson.

Rosansky, E. J. (1975). The Critical Period for the acquisition of language: Some cognitive developmental considerations. *Working Papers on Bilingualism*(6), 1-12.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language, 35*(4), 606-621. doi: 10.1006/jmla.1996.0032

Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental language learning: listening (and learning) out of the corner of your ear. *Psychological Science, 8*(2), 101-105. doi: 10.1111/j.1467-9280.1997.tb00690.x

Schwartz, B. D. (1993). On explicit and negative data effecting and affecting competence and linguistic behavior. *Studies in Second Language Acquisition, 15*(2), 147-163. doi: 10.1017/S0272263100011931

Sebastian-Galles, N., & Diaz, B. (2012). First and second language speech perception: Graded learning. *Language Learning, 62*(S2), 131-147. doi: 10.1111/j.1467-9922.2012.00709.x

Segalowitz, N., & Trofimovich, P. (2012). Second language processing. In A. Mackey & S. M. Gass (Eds.), *The Routledge Handbook of Second Language Acquisition* (pp. 179-192). London: Routledge.

Sparks, R. L., Patton, J., Ganschow, L., & Humbach, N. (2009). Long-term relationships among early first language skills, second language aptitude, second language affect, and later second language proficiency. *Applied Psycholinguistics, 30*(4), 725-755. doi: 10.1017/S0142716409990099

Stringer, D. (2008). What else transfers? In R. Slabakova, J. Rothman, P. Kempchinsky & E. Gavruseva (Eds.), *Proceedings of the 9th Generative Approaches to Second Language Acquisition Conference (GASLA 2007)* (pp. 233-241). Somerville, MA.: Cascadilla Proceedings Project.

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. M. Gass & C. G. Madden (Eds.), *Input in second language acquisition* (pp. 235-253). Rowley, MA: Newbury House.

Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seildlhofer (Eds.), *Principle & practice in applied linguistics: studies in honour of H.G. Widdowson* (pp. 125-144). Oxford: Oxford University Press.

Swain, M., & Lapkin, S. (1995). Problems in output and the cognitive processes they generate: A step towards second language learning. *Applied Linguistics, 16*, 371-391. doi: 10.1093/applin/16.3.371

Ullman, M. T. (2001a). The declarative/procedural model of lexicon and grammar. *Journal of Psycholinguistic Research, 30*(1), 37-69. doi: 10.1023/A:1005204207369

Ullman, M. T. (2001b). The neural basis of lexicon and grammar in first and second language: The declarative/procedural model. *Bilingualism: Language and Cognition*,(4), 17. doi: 10.1017/S1366728901000220

Unsworth, S., Persson, L., Prins, T., & De Bot, K. (2014). An investigation of factors affecting early foreign language learning in the Netherlands. *Applied Linguistics*, 1-23. doi: 10.1093/applin/amt052

Utdanningsdirektoratet. (2006). Knowledge Promotion: Curriculum in English. www.udir.no.

Van der Slik, F. W. P. (2010). Acquisition of Dutch as a second language. *Studies in Second Language Acquisition, 32*(03), 401-432. doi: 10.1017/S0272263110000021

Van Moere, A. (2012). A psycholinguistic approach to oral language assessment. *Language Testing, 29*(3), 325-344. doi: 10.1177/0265532211424478

VanPatten, B. (2004). Input processing in second language acquisition. In B. VanPatten (Ed.), *Processing instruction: Theory, research, and commentary* (pp. 5-31). Mahwah, NJ: L. Erlbaum Associates.

VanPatten, B. (2012). Input processing. In A. Mackey & S. M. Gass (Eds.), *The Routledge handbook of second language acquisition* (pp. 268-281). London: Routledge.

Paper IV

Individual differences and different outcomes in young children starting school in a new language

Anne Dahl

Abstract

This paper discusses the development of L2 English in an immersion school setting in a small group of children in Norway. Seven children starting grade 1 in international schools with little or no knowledge of the language of instruction (English) were followed over one year of school, and their English competence and development were measured using the Peabody Picture Vocabulary Test for receptive vocabulary, a sentence comprehension test, and a sentence repetition test. The results show that acquisition can be very rapid in such a context, but that there is great inter- and intra-participant variability, highlighting the need to discuss bilingual children in schools as individuals rather than as a uniform group.

1. Introduction

Child bilingualism is a common, but complex, phenomenon. While the past four decades have seen a dramatic increase in research on the topic, we still need more knowledge about factors such as age of onset (AoA) and volume of language input, and their impact on language competence for children who speak more than one language. The present paper focuses on a specific group of child language learners, namely a group of six-year-olds starting first grade in international schools in Norway with little or no knowledge of English, which is the language of instruction.

2. Background

With migration, increasing numbers of children with home languages other than the language of instruction are starting school all over the world. Linguists generally agree that these children's multilingualism is not cause for concern, since the knowledge and use of more than one language is very common and a reality for at least half of the world's population (Gathercole, Thomas, & Hughes, 2008; Grosjean, 2010; Wei, 2000), and the human brain seems to be well equipped to learn multiple languages from childhood (Meisel, 2004; Ortega, 2014). Knowing more than one language has obvious advantages for communication, but there are also cognitive

benefits to bilingualism (Adesope, Lavin, Thompson, & Ungerleider, 2010; Bialystok, 2009, 2011; Craik & Bialystok, 2005; Lauchlan, Parisi, & Fadda, 2013; Luk, De Sa, & Bialystok, 2011; Serratrice, 2012).

At the same time, there is debate about vocabulary size in bilingual children, and some studies report smaller average vocabularies in each language than those of monolinguals (Bialystok, 2009; Bialystok, Craik, Green, & Gollan, 2009; Scheele, Leseman, & Mayo, 2010; Serratrice, 2012). This could potentially be a disadvantage for children with home languages other than the language of instruction when they start school, in that they may be unable to fully follow the content of classes and fall behind. There is concern especially when it comes to learning to read, since literacy partly depends on vocabulary knowledge, especially at early stages of literacy (cf. Cummins, 1979; Dickinson, Flushman, & Freiberg, 2009; Lee, 2011; Lervåg & Aukrust, 2010; Melby-Lervåg & Lervåg, 2013; Ricketts, Nation, & Bishop, 2007). In particular, it can be argued that catching up with native-speaking peers at school is hard because the target is moving; after all, native speaking children's language skills are also developing, not least in terms of literacy (Cummins, 2007).

However, bilinguals cannot be expected to simply have competence in each language identical to the competence of monolingual native speakers of the same languages (Grosjean, 1989; Meisel, 2004). If it is indeed the case that the vocabularies of bilinguals in each language tend to be smaller than those of monolinguals, this is not surprising, given that they may use their languages for different purposes, and it is likely that the total conceptual vocabulary of bilinguals is on average larger than for monolinguals (Bialystok, Luk, Peets, & Yang, 2010; Gathercole et al., 2008; Grosjean, 1997; Oller, Pearson, & Cobo-Lewis, 2007; Serratrice, 2012). For reading, it is not clear that vocabulary size in each language is more relevant than total vocabulary, since the relationship between vocabulary and reading may in part have to do with larger vocabularies providing more words upon which to base phonological analysis (Aukrust, 2008; Gorman, 2012). Furthermore, there may be positive transfer from one language to the other in bilinguals when it comes to other skills which facilitate literacy, such as phonological awareness or understanding sound-symbol correspondences, at least when the two languages use similar script (cf. Bialystok, Luk, & Kwan, 2005; Bialystok, McBride-Chang, & Luk, 2005; Cardenas-Hagan, Carlson, & Pollard-Durodola, 2007).

Individual differences and different outcomes

The term "bilingual" is used to refer to a heterogeneous group of individuals with varying degrees of competence in each language (cf. Bialystok, McBride-Chang, et al., 2005; Butler, 2012; Grosjean, 2010; Meisel, 2004; Wei, 2000). It cannot be assumed, for example, that simultaneous and successive bilinguals are directly comparable, since factors such as age of acquisition and volume of target language exposure will influence language proficiency. Furthermore, although average vocabulary scores for bilinguals may possibly be lower in each language, great overlap has been found between monolinguals and bilinguals, and with sufficient exposure to both languages, vocabulary size need not suffer with bilingualism (e.g., Bialystok et al., 2010; De Houwer, Bornstein, & Putnick, 2013; Pearson, Fernandez, & Oller, 1993; Vulchanova, Vulchanov, Sarzhanova, & Eshuis, 2012). It follows that concerns about language proficiency for bilinguals in school are more warranted for children whose input in the language of instruction has been low, i.e., children with other home languages who have also not had sufficient exposure to the language of instruction in contexts outside of the home.

When children start school with low proficiency in the language of instruction, a concern may be that it is impossible for them to attain the level of their native-speaking peers. After all, native competence in a language by definition requires that it is acquired from very early childhood (cf. Cook, 1999), and observed age effects in language acquisition have been suggested to result from a critical or sensitive period for acquisition (DeKeyser, 2000; Johnson & Newport, 1989; Lenneberg, 1967; Penfield & Roberts, 1959). From this perspective, children who start school in a language in which they have low proficiency might always be at a disadvantage when it comes to learning. However, these concerns hinge on certain assumptions. For example, they assume a fundamental difference between native and non-native proficiency, and that native-like skills can only be acquired with an age of onset prior to school age. They also assume that the difference encompasses all areas of language acquisition, or at least those crucial to academic learning through the language. Neither assumption is necessarily true. For example, there is no clear agreement that there is a critical or sensitive period for language acquisition, and certainly not that it ends before school age (cf. Bialystok, 2002; Hyltenstam & Abrahamsson, 2003). Furthermore, it has been suggested that age effects may influence different areas of language differently, i.e., that there exist a number of smaller sensitive periods for different aspects of language (Locke, 1997; Long, 1990; Meisel, 2009; Ruben, 1997; Seliger, 1978).

Significantly, for vocabulary there is little evidence of a critical period. Although many studies discuss a so-called "vocabulary spurt" in early language development, young children do not learn vocabulary at a faster rate than older ones; vocabulary learning happens at its fastest throughout a person's late childhood and adolescence, largely due to school and reading (e.g., Bloom, 2000; Bloom, 2004; Verhoeven, van Leeuwe, & Vermeer, 2011). Thus, there is no reason to think that vocabulary acquisition in a second language – certainly in a context of child SLA – is necessarily flawed, or that acquiring the vocabulary necessary to follow classes cannot happen after starting school.

Language comprehension does not only depend on vocabulary size, but also on processes such as speech perception and segmentation, lexical access, and mapping the input onto grammatical structure (Carroll, 2004; Gibson & Pearlmutter, 1998; Kroll & Dussias, 2004). We know that receptive language skills can be learned well along with academic content in an immersion setting (e.g., Genesee, 2004; Wesche, 2002; Wode, 1999). For example, so-called Content and Language Integrated Learning (CLIL) programs seem to provide contexts for successful learning (see e.g., Pérez-Cañado, 2012 for an overview). Thus, we should expect that academic content can be learned well in a non-native language, and that language can in fact be learned along with academic content in a school setting. It is also worth noting that bilingual cognitive advantages have been found not only for early bilinguals, but also for children who have become bilingual through immersion programs, and that these effects increase with increased experience and proficiency in the L2 (Bialystok & Barac, 2012; Bialystok, Peets, & Moreno, 2014; Nicolay & Poncelet, 2013).

The idea that the language of instruction may be learned along with academic content may be new to many communities with monolingual traditions. However, it is an idea familiar, for example, to international schools. In such schools, the language of instruction is not that of the community in which the school is situated; typical examples are English-language schools in many European countries. Parents' motivation for enrolling their children in such schools may vary. The child may speak English as an L1, but children with multilingual or international backgrounds also typically attend international schools. This results in highly multilingual and heterogeneous classrooms where native or near-native competence in the language of instruction cannot be assumed in all students from day one (e.g., Sears, 1998).

The debate about whether a home language other than that of instruction is a handicap when starting school reflects the need for more research on such situations. The children under discussion are clearly not a homogenous group, and studies of competence and development in the language of instruction need to take this into account. In this context, it is useful to investigate the language development of children starting international schools with very low competence in the language of instruction, since they are not exposed to that language in the local community. Thus, we can assume that their language acquisition largely takes place in the classroom.

3. The study

This paper reports on a study of a group of children starting international schools in Norway at age six, with no or virtually no knowledge of the language of instruction. The study focuses on their L2 vocabulary development over one year of school, as well as their sentence comprehension and sentence repetition abilities at the end of the year.

3.1. Participants

The participants were seven children starting international schools in Norway where English is the language of instruction. Participants 1-6 all started at the same school, participants 1-3 in one class with the same teacher, and participants 4-6 in another. Participant 7 started at a different school as the only student in grade 1 who did not speak English. The parents of all children provided informed written consent for their child to participate in the study.

Participant 1 was male, aged 6;7 at the time of pre-testing and 7;3 at the time of the post-test. His home language was Norwegian, and he was monolingual prior to starting school. However, he had lived outside of Norway, with a total of about 1;8 years of stays in Russia, the US and Australia before the age of three.

Participant 2 was female, aged 5;8 at the time of pre-testing and 6;4 at the time of the post-test. She was from a multilingual home; her parents reported that they spoke Polish and Italian to the children at home, but that the child and her eight-year old brother spoke Norwegian amongst themselves. Although all three languages were reported as native languages, she was reported to be Norwegian-dominant. The parents also reported that they spoke English to each other but not to the children, and that the child did not speak English herself when she started school. This child was born in South Africa but had lived in Norway since age one.

Participant 3 was female, aged 5;10 at the time of pre-testing and 6;6 at the time of the post-test. This girl's home language was Swedish, which is a language very similar to Norwegian, but she was born and raised in Norway and was reported to be a relatively balanced Swedish/Norwegian bilingual when she started school.

Participant 4 was male, aged 6;4 at the time of pre-testing and 7;0 at the time of the post-test. He was a German/Norwegian bilingual, reported by his parents to be relatively balanced, but with German as his home language. He was born in Germany but had lived in Norway since age one.

Participant 5 was female, aged 5;10 at the time of pre-testing and 6;6 at the time of the post-test. She was a Swedish/Norwegian bilingual, with Swedish as her home language and dominant language when she started school. She was born in Norway but had lived in Poland for 18 months between the ages of 0;8 and 2;3.

Participant 6 was female, aged 5;8 at the time of pre-testing and 6;4 at the time of the post-test. Her mother spoke Norwegian to her, and her father French, but she was strongly Norwegian dominant with mostly only receptive knowledge of French. Her father's native language was Portuguese, of which she also understood a little.

Participant 7 was male, aged 6;1 at the time of pre-testing and 6;9 at the time of the post-test. His home language was Bulgarian, but he was born and raised in Norway and was reported to be a balanced Norwegian/Bulgarian bilingual. This child was the only student in his class with no initial knowledge of English, and the only one in the test group to attend English as an Additional Language (EAL) classes when he started school; he attended such classes for 2 hours per week for the first four months of the school year.

3.2. Test sessions

Each child in the test group was tested twice. Pre-tests took place in the first week of September in the fall term, and post-tests in the first week of May in the following spring term. Time between testing was thus approximately eight months for each child. Testing took place in a quiet room at the child's school, with only the participant and the researcher present. Each test session lasted for approximately one hour.

3.3. Control measures

All children were tested for Norwegian vocabulary using a translated version of the Peabody Picture Vocabulary Test, Fourth edition (PPVT-IV) form A. Non-verbal intelligence was tested with the Matrices section of the Kaufman Brief Intelligence Test, Second edition (KBit-2), while verbal ability (verbal comprehension, reasoning, and vocabulary knowledge, cf. Kaufman & Kaufman, 2004) was tested using a Norwegian translation of the Riddles section of the same test. Visio-spatial memory was tested using a game where the participant was given sets of picture cards in a given pattern to memorize before the cards were flipped upside down and the child was asked to find the pairs. When the results of these control measures are presented, the test group is compared to a control group of 60 monolingual Norwegian children[1].

3.4. English measures

Three English measures were used to study the participants' L2 development and their competence at the end of the first year of school. Receptive vocabulary was tested using the PPVT-IV. In the pre-test session, only form B was administered, while the post-test used both forms A and B. In this test, the participant hears a series of words, and for each word is asked to select from a set of four pictures the one matching the word.

Sentence comprehension was tested both in the pre-test and in the post-test through computerized tests where the participant heard auditory stimuli and was asked to select the matching picture from a set. In the fall, an online test was used (Elllo, 2013a, 2013b), where each linguistic stimulus consisted of several sentences, with three alternative pictures for each stimulus. There were 12 such picture sets, and the maximum possible score was thus 12. Linguistic stimuli were spoken by a male native speaker of American English. For the post-test, a comprehension test was produced specifically for the project of which this study is a part. Each linguistic stimulus in this test consisted of one sentence only, and each sentence was accompanied by sets of four alternative pictures (see Appendix 1). There were 30 picture sets, and the maximum possible score was thus also 30. Only one picture in each set corresponded completely to the sentence, while two other pictures contained elements mentioned in the sentence but without full correspondence. For example, for the sentence "Daddy is reading to the boy," only one picture showed a man reading to a boy. However, one picture showed a man

[1] These children were participants in another, related study. They were all L1 Norwegian speakers who had just started learning English as a foreign language in regular Norwegian state schools.

reading with no boy present, and another showed a boy who was reading. The fourth picture in each set was completely unrelated to the sentence; in the present example, this picture was of a girl and a dog. Sentences in this test were read alternately by two female speakers, one native speaker of British English and one native speaker of American English.

The final English measure, which was only administered in the post-test session, was a test of sentence repetition, where the researcher, a proficient L2 speaker of American English, read a total of 17 sentences of varying length and complexity, which the participant was asked to repeat (see Appendix 2). This test was intended to evaluate grammatical processing, and in particular sensitivity to morphology, although with such tests, also aspects of productive competence are assessed (cf. Dahl, n.d.; Jessop, Suzuki, & Tomita, 2007; Van Moere, 2012). The test was recorded, and scored by two independent raters, one native speaker of American English and one balanced Norwegian/Dutch bilingual who was also proficient in American English. Both raters were university students of English linguistics. They were instructed to award one point per correctly repeated morpheme, including certain zero morphemes (see Appendix 2), and were specifically told to disregard pronunciation as long as the morphemes were comprehensible. The maximum possible score on this test was 142. The test was evaluated and found to be a reliable and valid measure of language competence, with good internal consistency and interrater reliability (see Dahl, n.d. for details). The scores reported in the present paper are the means of the two raters' scores unless otherwise specified.

When results are presented, PPVT-IV results are compared to the averages of the reference group of American native speakers reported in the test manual (Dunn & Dunn, 2007). Results on the sentence comprehension and repetition tests are compared to an age-matched control group of 15 native English-speaking students at international schools in Norway. The criterion for inclusion in this control group was that English was one of the child's main languages, for example, a home language or the language of the community where the child had grown up and gone to pre-school. According to parents' reports, 10 children in this group were bilingual, with Norwegian being the most commonly spoken language in addition to English, while five children were reported to be monolingual English speakers. However, knowing that they lived in Norway and that Norwegian was taught as an L2 in their school, all children in this control group had at least some multilingual experience.

4. Results

Group results

Descriptive statistics for the overall results of the seven children in the test group on all post-test measures of English are reported in Table 1 below, along with the means of the reference group for the PPVT-IV forms A and B, and the means of the native English-speaking control group for sentence comprehension and sentence repetition.

Table 1: Post-test group results, all English measures, with reference/control group means

	Test group minimum	Test group maximum	Test group mean	Test group SD	PPVT-IV reference group/ native speaker control group means
PPVT-IV A Vocabulary	58	89	77	12.10	112.5[a)]
PPVT-IV B Vocabulary	58	107	83.29	19.85	115 .0[a)]
Sentence comprehension	27	29	28.14	.90	29.27[b)]
Sentence repetition	120.50	139.75	132.64	6.53	138.22[b)]

[a)] Mean of PPVT-IV reference group matched for the age of the test group (6;8) (Dunn & Dunn, 2007)
[b)] Mean of the native speaker control group (n=15)

The above results show that in the post-test, receptive vocabulary, as measured by the PPVT-IV forms A and B, was on average much lower in the test group than in the reference group. The test manual only provides average scores and no individual data for the reference group, and statistical analysis is therefore impossible for these results. However, the manual allows us to convert raw scores into Growth Scale Values (GSVs) which can be used to check for significant differences. The test group mean of 77 on the PPVT-IV form A corresponds to a GSV of 126, while the reference group mean of 112.5 corresponds to a GSV of 148. According to the test manual (Dunn & Dunn, 2007), a GSV difference of 8 points or more is significant in the relevant age range, and the difference of 22 points between the test group and the reference group on the PPVT-IV form A is thus clearly significant. For form B, the test group mean of 83 translates into a GSV of 129 while the reference group mean of 115 corresponds to a GSV of 150; this difference of 21 points is also significant. Vocabulary comprehension in the test group after almost one year of school was thus, not surprisingly, significantly below what would be expected for native speakers.

In the comprehension test there seem to be ceiling effects in the test group. Although all participants scored between 27 and 29 out of the maximum possible score of 30, i.e., they all made at least one error, the mean is not much lower than that of the native English-speaking

control group. The sentences in this test were simple, and results cannot be taken to mean that the participants actually had native-like listening comprehension; a more challenging test may have revealed greater differences. Also, no participant in the test group actually obtained a perfect score on this measure, while the majority of the control group, nine out of 15, did. A non-parametric Mann-Whitney U test found the difference between the control group and the test group to be significant, U (21) = 21, Z = -2.3, p = .010, r_{effect}=-.49. Thus, what seems to be near-ceiling performance in the test group is still systematically below that of the control group. The reason for the significant difference found is likely to be exactly the fact that the majority in the control group obtained perfect scores, while no participant in the test group did. The very low, but consistent, number of errors in the test group may reflect performance problems rather than problems in underlying competence, i.e., that in real-time processing, children in this group were more prone to make occasional mistakes than were the native speakers. Such an account is supported by the lack of any clear pattern to their mistakes. Another indication is the fact that no participant in the test group at any time chose the picture in a set which had no elements corresponding to any part of the sentence; in other words, all their responses corresponded at least to some words in the stimulus.

In the repetition test, the native-speaking control group performed near ceiling with a mean of 138.22 out of a maximum of 142, while the test group showed greater variability as well as a lower mean of 132.64. Three out of the seven children in the test group scored below the minimum score in the control group, which was 130.5. The difference between the means of the two groups was tested with a Mann-Whitney U test and found significant, U(21)=23.5, Z = -2.0, p = .020, r_{effect} =-.43.

To look for intra-subject consistency or variability, a correlation analysis was run with Spearman's Rho investigating the relationship between results on the four post-test measures of English competence (two forms of the PPVT-IV, sentence comprehension, and sentence repetition) within the test group. The results are reported in Table 2. The fact that only the two forms of the PPVT-IV are correlated is an indication that there was intra-subject variation on the different English measures.

Table 2: Test group correlations between all post-test English measures

		PPVT-IV A	PPVT-IV B	Sentence comprehension
PPVT-IV B	Spearman's ρ	.716*		
Sentence comprehension	Spearman's ρ	.144	.038	
Sentence repetition	Spearman's ρ	-.193	.100	.181

*Correlation is significant at the 0.05 level (1-tailed)

Another correlation analysis with Spearman's rho was conducted to check the influence of the factors tested by the background measures on L2 development, i.e., on differences in English scores from the pre- to the post-test. Two of the English measures could be meaningfully treated as a test-retest design where development from the pre-test to the post-test could be measured. For vocabulary, this was straightforward, since the PPVT-IV form B was administered both in the pre-test and in the post-test. For sentence comprehension, different tests were administered in the fall and the spring; they differed both in how many items they included (12 in the fall and 30 in the spring) and in number of alternative responses per item (three in the fall and four in the spring). However, scores from each test were transformed into standard scores by dividing each participant's percentage of correct responses by the number of alternative responses, allowing us to look at relative development in performance also on this measure. Table 3 shows the correlation between control measures and increase in vocabulary and sentence comprehension.

Table 3: Test group correlations between background measure scores and increase in scores on English vocabulary and sentence comprehension from pre-test to post-test

		No. voc.	K-Bit Mat.	K-Bit Rid.	Memo	No of L1s	Pre PPVTB	Pre sent. comp.	PPVTB increase
K-Bit Mat.	ρ	-.148							
K-Bit Rid.	ρ	.544	-.281						
Memo	ρ	-.815*	.000	-.243					
No of L1s	ρ	-.348	.020	.444	.374				
Pre PPVTB	ρ	.148	-.750*	.599	.179	.374			
Pre sent. comp.	ρ	-.552	.315	.000	.355	.576	.118		
PPVTB increase	ρ	.449	.595	.415	-.559	.129	-.450	.020	
Sent. comp. increase	ρ	.224	-.144	.113	.072	-.457	-.144	-.716*	.045

*Correlation is significant at the .01 level (one-tailed)

No. voc. = Norwegian vocabulary

K-Bit Mat. = Kaufmann Brief Intelligence Test, Matrices section

K-Bit Rid. = Kaufmann Brief Intelligence Test, Riddles section

Memo = Memory game

No of L1s = Number of native languages

Pre PPVTB = Pre-test PPVT-IV form B

Pre sent. comp. = Pre-test sentence comprehension

PPVTB increase = increase in PPVT-IV form B score from pre- to post-test.

Sent. comp. increase = increase in sentence comprehension score from pre- to post-test.

Table 3 provides little evidence of a relationship between the non-English background factors tested and success in English acquisition. There are only three significant correlations in the matrix, all negative; one between visio-spatial memory and Norwegian vocabulary, one between non-verbal intelligence and pre-test vocabulary, and one between pre-test score on the sentence comprehension test and development on sentence comprehension. While the two first negative correlations may be slightly surprising, only the third is relevant for language development during the test period. However, this negative correlation is hard to interpret, since the mean score on the sentence comprehension post-test was close to ceiling. The negative correlation might indicate that the participants whose sentence comprehension was the poorest in the pre-test were those whose comprehension improved the most. This is a plausible explanation since comprehension of connected speech is likely to develop fastest for children whose initial

comprehension is very low (see also Dahl, to appear). Those with better initial comprehension may not have needed an equally dramatic development in order to follow classes. However, a more challenging post-test may also have allowed those with better initial comprehension to display more development, and the negative correlation may be a result of ceiling effects for those with higher competence. In conclusion, the correlation matrix provides no clear evidence of what factors may influence L2 development in such an immersion setting.

The lack of correlation with the number of L1s means that we specifically have no evidence of an effect of bilingualism on L2 acquisition. Quantifying multilingualism is difficult, and this study included no measure of the participants' language competence in each language beyond their parents' superficial report of number of L1s and language dominance. However, an attempt was made to look for an effect of the number of L1s by adding it to the correlation matrix. For the figures reported in Table 3, the child who was reported to have mainly only receptive French and some Portuguese in addition to Norwegian was counted as having 1.5 L1s. An analysis was also run counting this child as having two L1s, but still, no significant correlation was found. Because of the low number of participants, and the fact that only one child was really monolingual at the start if the year, and only one trilingual, this lack of correlation does not warrant strong conclusions. However, it is an indication that a child's number of L1s does not necessarily have an impact on success when starting school in a new language. Still, it is clear that with a small and varied group such as this, correlation analysis is of limited value, and looking at individuals separately is necessary. This will be done in the next section.

Individual results

Table 4 shows the individual scores of each participant on each background and English measure, as well as gender, age, and language background. For the background and pre-test measures, means and SDs are provided for the control group of monolingual native speakers of Norwegian (n=60) for comparison. For post-test scores on the PPVT-IV forms A and B, descriptive statistics are provided for the reference group reported in the test manual (Dunn & Dunn, 2007) for test group's mean age. For the post-test measures of English sentence comprehension and repetition, the same descriptives are provided for the control group of English native speakers (n=15).

Table 4: Individual test group participants' scores on all background and English measures, with control/reference group means and SDs

Participant	1	2	3	4	5	6	7	NO ctrl	EN ctrl	Ref	Ctrl SD
Gender	m	f	f	m	f	f	m	*f29/m31*	*f7/m8*	-	-
Age	7;3	6;4	6;6	7;0	6;6	6;4	6;9	*6;10*	*6;6*	-	-
Languages	1	3	2	2	2	2*	2	*1*	*1-3*	-	-
No.voc.	105	103	105	89	71	105	112	*117*	-	-	*15*
K-Bit Rid.	11	15	16	11	11	13	15	*16*	-	-	*4*
K-Bit Mat.	24	16	15	13	25	12	21	*18*	-	-	*5*
Memo	52	42	50	48	40	44	57	*47*	-	-	*5*
Pre PPVTB	35	67	56	42	27	66	39	*25*	-	-	*12*
Pre sent. comp.	7	9	7	7	7	7	6	*5.5*	-	-	*1.5*
Post PPVTA	68	86	86	68	58	89	84	-	-	*112.5*	*6*
Post PPVTB	66	96	107	66	58	87	103	-	-	*115*	*8*
Post sent. comp.	28	27	29	27	29	29	28	-	*29.27*	-	*1*
Post sent. rep.	120.5	136.7	135.0	129.7	139.7	129.7	137.0	-	*138.22*	-	*3.5*

* One language receptive only, also some comprehension of a third language

NO ctrl= Norwegian control group (n=60)

EN ctrl = English control group (n=15)

Ref= PPVT-IV reference group

Ctrl SD = Native speaker control/PPVT-IV reference group's Standard Deviation

No. voc = Norwegian vocabulary

K-Bit Rid. = Kaufmann Brief Intelligence Test, Riddles section

K-Bit Mat. = Kaufmann Brief Intelligence Test, Matrices section

Memo= Memory game

Pre PPVTB = Pre-test PPVT-IV form B

Pre sent. comp. = Pre-test sentence comprehension

Post PPVTA = Post-test PPVT-IV form A

Post PPVTB = Post-test PPVT-IV form B

Post sent. comp. = Post-test sentence comprehension

Post sent. rep. = Post-test sentence repetition

For form B of the PPVT-IV, we can check L2 development for participants individually, comparing the difference in age equivalents in the pre-test in September to the post-test in May. This is done in Table 5, which shows each participant's chronological age at the time of the pre-

and the post-test, along with his or her raw scores on the PPVT-IV form B in both test sessions, the corresponding age equivalents, and finally the number of months to which this age equivalent development corresponds.

Table 5: Increase in age equivalents for receptive vocabulary from pre- to post-test for each test group participant, based on PPVT-IV form B

Participant	Age pre-test	Age post-test	Score pre-test	Score post-test	Age equivalent pre-test	Age equivalent post-test	Increase in age equivalents
1	6;7	7;3	35	66	2;10	4;1	15 months
2	5;8	6;4	67	96	4;2	5;9	19 months
3	5;10	6;6	56	107	3;9	6;5	32 months
4	6;4	7;0	42	66	3;2	4;1	11 months
5	5;10	6;6	27	58	2;6	3;10	16 months
6	5;8	6;4	66	87	4;1	5;3	14 months
7	6;1	6;9	39	103	3;1	6;2	37 months

In the following, the individual results of each participant are presented in detail and discussed. For the two forms of the PPVT-IV, several data points are given in addition to raw scores. From the raw score on each form, a standard score is calculated which tells us the relative value of the score compared to the age-matched reference group for this child's age; a standard score of 100 is the average. Then the percentile of the standard score is calculated, indicating the percentage of children in the age-matched reference group who scored at or below the participant's standard score. Finally, raw scores are translated into age equivalents, indicating at which age the raw score is the average score in the native-speaking reference group. All these estimates are readily available from the PPVT-IV manual (Dunn & Dunn, 2007). For each participant, all data from the PPVT-IV form B are presented first, since this test allows us to discuss the participant's development from the pre-test to the post-test in terms of age equivalents. Next, the same data are presented for the PPVT-IV form A, but without a discussion of development in age equivalents, since this form was not administered in the pre-test. Scores for the comprehension test and the repetition test are then presented; for each of these only a raw score is available since they were produced specifically for the present study and no norming is

available. However, each score is compared to the average and range of the study's English native-speaker control group.

Participant 1 obtained a score of 35 on the PPVT-IV form B in the pre-test. In the post-test, his raw score was 66. In terms of age equivalents, this means development from 2;10 to 4;1, or 15 months, in a time span of 8 months. It is clear that this participant's vocabulary comprehension has developed rapidly, but it is still well below the average for native-speaking age-matched peers. A raw score of 66 for his age corresponds to a standard score of 69, and to the 2nd percentile, meaning that it is expected to be below 98% of the native speaker population at his age. His PPVT-IV form A raw and standard scores are very similar, at 68 and 71, respectively. This means an age equivalent of 4;3, and also places him at the 2nd percentile.

This participant's score on the comprehension test was 28, only slightly below the control group mean of 29.27, while on the repetition test his score was 120.5, which is well below the control group mean of 138.22. This makes his repetition score the only one to stand out compared to the rest of the test group, being by far the lowest. His performance on vocabulary and sentence comprehension is not particularly low compared to the rest of the test group, and it is therefore worth investigating what errors he made on the repetition test. It turns out that this participant's problems in sentence repetition are found almost uniquely in grammatical endings and function words. He receives zero scores on many of these, and there is almost 100% agreement between the two raters on these errors. This indicates that the elements are being omitted entirely, and are not just unclear or mispronounced. The problem in this participant's sentence repetition thus probably has to do with grammatical processing, either in perceiving or in producing grammatical endings and function words.

Participant 2 had a raw score of 67 on the PPVT-IV form B in the pre-test, with an age equivalent of 4;2, and a raw score of 96, with an age equivalent of 5;9, in the post-test, which means development of an equivalent of 19 months. This puts her close to her chronological age of 6;4 in terms of age equivalents, and her standard score is 91. This score places her at the 27th percentile. Her PPVT-IV form A score was a little lower at 86, with an age equivalent of 5;4, and a standard score of of 87, which places her at the 19th percentile. On the sentence comprehension test, her score was 27, which is among the lowest scores in the group, while on the repetition test it was 136.75, which is close to the control group mean of 138.22. This participant's pre-test vocabulary score was the highest in the experimental group, possibly because her parents spoke

English among themselves at home, but her pre-test age equivalent still places her well below what would be expected from a native speaker. The overall results in the post-test, however, indicate that her performance at this point is within the normal range of native speakers on most measures albeit below average. The comprehension test stands out with a low score, however, indicating variability in her performance.

Participant 3's pre-test PPVT-IV form B raw score was 56, with an age equivalent of 3;9. Her post-test raw score was 107, with an age equivalent of 6;5, which means improvement of an equivalent of 32 months. This also puts her only one month behind her chronological age in terms of age equivalents, with a standard score of 98, which places her at the 45th percentile, i.e., very close to the mean. However, there is clearly variability in her performance, because her PPVT-IV form A score is only 86, with a standard score of 85, and an age equivalent of 5;4, and which places her at the 16th percentile. Her score on the comprehension test is 29, and on the repetition test it is 135. In sum, her vocabulary development in terms of age equivalents has been very rapid, although there is variability in her performance on the PPVT-IV, and performance on the other measures is also high and within the range of the native-speaker control group.

Participant 4 had a pre-test PPVT-IV form B raw score of 42 with an age equivalent of 3;2, and a raw score of 66 in the post-test, with an age equivalent of 4;1, i.e., development equivalent to 11 months. This is the smallest development seen in the test group in terms of age equivalents, and this participant is still far behind his native-speaking age-matched peers, with a standard score of 71, which places him at the 3rd percentile. However, his development is still faster than for language-matched younger native speakers, with 11 months in age equivalents in a chronological time span of eight months. His PPVT-IV form A score is very similar to that of form B, at 68, which has an age equivalent of 4;3, and a standard score of 72, which places him at the 3rd percentile. This child's score on the comprehension test is 27; this is one of the lowest scores in the test group. His score on the repetition test is 129.75, which means that he is one out of three participants to score below the range of the native-speaker control group, whose lowest score was 130.5. Unlike participant 1, who also received a low score on the repetition test, there is not great agreement between the two raters about errors on the repetition test for participant 4, and in fact, he received a score of 134.5 from one rater and of 125 from the other. To resolve such cases of discrepancy, the researcher had scored all participants' repetitions before seeing the independent raters' assessments. Thus, a third assessment existed which could be consulted. It

turns out that the researcher's assessment was very similar to that of the most restrictive independent rater, with a score of 123 for this participant. Both the researcher and the restrictive rater had penalized what they deemed to be clear mispronunciations in several cases where the less restrictive rater had made no comment. It thus seems like the reason for the different scores is disagreement as to whether pronunciation was just non-native or whether the word was incomprehensible; it is worth noting that the less restrictive rater was the English native speaker. This participant's low mean score may thus have been a result of unclear pronunciation, which might indicate a simple articulatory problem. However, it could also be a sign of poor comprehension leading to the child often trying to mimic sounds rather than actually producing language. Errors reported are found in both functional and lexical words. It is difficult to know exactly what this indicates, but combined with his low score on vocabulary and sentence comprehension, it may be concluded that his problems in the repetition test are likely to stem from generally low English proficiency.

Participant 5 had the lowest score on the PPVT-IV form B at the beginning of the year out of the test group, with a raw score of 27 and an age equivalent of only 2;6. In the post-test, her raw score was 58, with an age equivalent of 3;10. This is the lowest post-test score in the group, but still means improvement of an equivalent of 16 months. However, the standard score is only 72, which places her at the 3rd percentile. Her score on the PPVT-IV form A was also 58, for which the age equivalent is 3;9, and the standard score is 71, and which also corresponds to the 3rd percentile. However, this child's scores on the comprehension and repetition tests were high; 29 and 139.75, respectively. This again shows variability in performance across tasks.

Participant 6's raw score on vocabulary at the beginning of the year was 66, with an age equivalent of 4;1. Her raw score in the post-test is 87, which has an age equivalent of 5;3, corresponding to development in age equivalents of 14 months. The standard score is also 87, and this places her at the 19th percentile. Her PPVT-IV form A score was very similar at 89, which has an age equivalent of 5;6, and a standard score of 88, which places her at the 21st percentile. On the comprehension test, her score is 29, which is one of the highest scores in the test group, but on the repetition test, her score is only 129.75, which is below the range of the native-speaker control group. There is not total agreement between raters about which errors she made, but their total scores are similar, with 129 vs. 130.5. Again, this might indicate a pronunciation problem, or it could have to do with general language competence. This participant's relatively high scores

on vocabulary and comprehension indicate that the problem is specifically related to repetition, possibly in the processing of grammatical detail.

Participant 7 had a raw score on the PPVT-IV form B at the pre-test of 39, which has an age equivalent of 3;1. His increase in the post-test is the greatest out of all the participants, and his post-test score of 103 is the second highest score in the group, with an age equivalent of 6;2 – only 7 months behind his chronological age. His development in age equivalents is thus 37 months in the course of eight months. The standard score of 92 places him at the 30th percentile. However, his PPVT-IV form A score is clearly lower than the score on form B, at 84, which also has a standard score of 84. This is still a high score compared to the rest of the group, but it has an age equivalent of only 5;3 and puts him at the 14th percentile, indicating that there is variability also in this participant's performance. His score on the comprehension test is 28, and on the repetition test it is 137. From a very low starting point, this child thus seems to have been a highly successful L2 acquirer by all accounts, although his performance is still variable. It is worth noting that this child had a different school experience to the other six; he was the only non-English speaker in his class at the outset, he started at a different school, and he was the only child to attend EAL classes. This may of course have been an important factor in his development, although it is important to note that these classes only took place for two hours a week for a total of about four months. It is also possible that this participant's language exposure at school was different from that of the others. The fact that he was the only non-English speaker in his class at the start of the year may have meant that more communication between classmates took place in English rather than in Norwegian also outside of the classroom, compared to the situation of the other participants, who each had two non-English speaking classmates.

5. Discussion

We see in the results of these seven children that L2 development in such an immersion setting is substantial. All the participants displayed vocabulary development as measured by the PPVT-IV faster than that of vocabulary-matched (younger) native speakers, in that their increase in age equivalents was invariably greater than their chronological age development. It thus seems like they may be on their way to catching up to the moving target of the native speaker (cf. Cummins, 1979), although more longitudinal studies as well as a more thorough investigation of different aspects of vocabulary knowledge would be needed to ascertain if and when this eventually happens. Still, these initial results are in line with arguments that vocabulary

development is faster during school years than for younger children (Bloom, 2004). They also indicate that language learning along with academic content can be effective.

It is also clear that there are great individual differences in L2 development over the school year, and it is not obvious from the results which factors predict success. There is no evidence that a child's number of L1s predicts development in a new language in a school immersion setting, and certainly not negatively; the only initially monolingual child in the study, participant 1, showed slightly less development than the only trilingual, participant 3, with an age equivalent of 15 vs. 19 months. Interestingly, these two children are also examples that multilingualism and home language status are not necessarily decisive for vocabulary in the majority L1, in that their scores on the Norwegian vocabulary test are only two points apart.

Furthermore, there is no evidence that language similarity is a benefit. All children knew Norwegian, which is a Germanic language relatively closely related to English, and all their other languages are also Indo-European and thus related to English. However, the child with the greatest development was participant 7, who in addition to Norwegian spoke Bulgarian, while the child with the smallest development, participant 4, spoke German and Norwegian. Bulgarian is structurally more similar to English and Norwegian than many other Slavic languages, e.g., in being largely analytic and lacking case morphology, but it is not closely related to the Germanic languages. German is much more closely related to both Norwegian and English, and can in particular be expected to share more cognates.

It is also clear that the same school situation can lead to different acquisition rates, since participants 1-3 and 4-6 attended the same classes, and development within these groups varied. However, the fastest-developing participant 7 was the only child in the test group to start at his school, in a class where all other children already spoke English fluently, either as an L1 or as an L2. His situation was also different from that of the other children in the study in that he initially attended EAL classes. It is possible that school factors including these classes were beneficial for this child's learning, although it is impossible to generalize from one participant. It is also important to note that EAL classes only took place for a limited time. Although they may certainly have contributed, attributing this child's success entirely to these classes would probably be overly optimistic as to the impact that special language classes can be expected to have.

Individual differences and different outcomes

The results also indicate that rapid development in the language of instruction is possible whether the starting point is extremely low or slightly higher. There is no obvious relationship between starting level and rate of development; the participant with the most rapid development (participant 7, 37 months) was among those with the lowest starting level, while the participant with the second most rapid development (participant 3, 32 months) had a relatively high starting level.

A further finding is that although all children in this study show great language development, there is intra-subject variability. Although the sentence repetition and especially the sentence comprehension measures were less sensitive than the PPVT-IV, it is clear that a few children scored differently on the former from what would be expected from their vocabulary scores. Furthermore, some children, notably those whose scores on the PPVT-IV form B are very high, score relatively differently on the two forms of the PPVT-IV, also indicating a variability that would be unexpected in native speakers.

All children's scores are still fairly low compared to native speaker norms. None of the scores on either form of the PPVT-IV are above the age-matched average of the reference group for their age, and no child scored consistently above the 25th percentile for both forms of the PPVT-IV. Comparison to the largely monolingual reference group of the PPVT-IV[2] may not be entirely fair if it is true that a lower score on this test is to be expected also in balanced bilinguals. However, with the exception of the two very high scores on form B (participants 3 and 7), all participants in the test group scored more than the ten points lower than the reference group average, which has been reported as a bilingual average (Bialystok, 2009). On the sentence comprehension and repetition tests, the group scored significantly lower than the control group of native English speakers, all of whom had some bilingual experience in that they spoke English but lived in Norway; ten out of 15 participants in the control group were reported by their parents to be bilingual to a greater or lesser degree. In other words, it is clear that all participants in the test group are still better described as child L2 learners than as bilinguals in terms of their English.

The findings in the present study may have relevance for other groups of children starting school in a new language, including minority language children starting school in a majority

[2] The PPVT-IV does not assume monolingualism in its reference group, but since this group consists of English speakers in the US, the majority can be assumed to be monolingual.

language. There are, however, some striking differences between the participants in the present study and the majority of such minority language children. One is the fact that the international schools in the present study were highly multilingual. While English was the language of instruction, it was not the native language of all students, and most children were bilingual at the outset. This will have influenced teaching styles and communication (cf. Sears, 1998). Another is socioeconomic status (SES). International schools are among the few private schools in Norway, and it is very likely that SES averages are higher in such a school than in society as a whole. All children in the present study came from high-SES families, and most parents had academic backgrounds. Many children starting school in a language which is not their L1 in other contexts, e.g., immigrants, may come from low-SES backgrounds. This is surely relevant, since we know that children from high-SES backgrounds are more likely to have larger and more academically relevant L1 vocabularies than low-SES children (cf. Cummins, 1979; Farrant & Zubrick, 2013; Oller et al., 2007). Significantly, however, differences in SES must not be confused with differences between mono- and bilingualism. If it is true that high-SES children perform better than low-SES children when starting school in a new language, this shows that it is not multilingualism per se but other factors which create problems for some groups of multilingual children in school, at least as far as learning the language of instruction is concerned (cf. Edwards, 2009).

A final difference between the children in the present study and most children starting school in a majority language is the availability of language input outside of the classroom. Although English is ambient in Norwegian society, it is not the community language; the language to which children are exposed in everyday life outside of the school context is Norwegian. The L2 development seen for English in this study, then, is largely in the absence of such community input, which is present for minority-language children, and which can be assumed to be an additional benefit (Vulchanova et al., 2012).

6. Conclusion

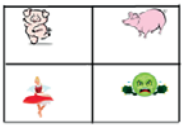
It is clear that multilingual children and children starting school in a language other than their L1 are not a homogeneous group and should not be treated as such. It is also clear that an L2 can be expected to be acquired rapidly in a school setting, but that one year of school certainly is not enough to catch up to the moving target of the (bilingual or monolingual) native speaker. Any studies of, or recommendations for, children with languages other than that of instruction must

take this variation into account, and not treat multilingual children in school settings as one uniform group. This also highlights the need for more research on different groups of multilingual children, focusing on different learning contexts and other factors which may influence language development.

Appendix 1

Sentence comprehension test

Trial 1: The chair is green



Trial 2: The pig is dancing

1. The girl is eating an apple

2. The boy is out in the rain

3. There are trees outside the house

4. The man is wearing a white shirt

5. This boy has a nice bird

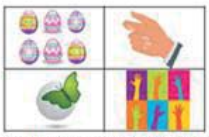
6. The little girl is jumping

7. The car in this picture is red

8. The people in this picture are happy

9. The boy is very sad

10. Where is the black shoe?

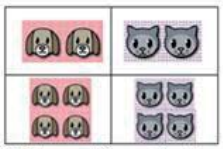


11. There are six hands in this picture

12. The girl really likes to sing

13. She has washed the dog

14. Daddy is reading to the boy



15. There are four cats in this picture

16. This funny rabbit has pink ears.

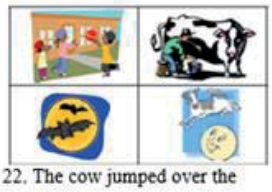
17. Can you see the yellow flower?

18. This elephant is playing with a ball.

19. The teacher is very old.

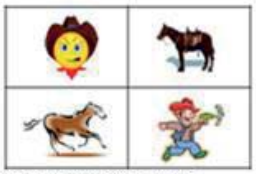
20. The mother is feeding the baby.

21. The boy has drawn his mother a picture.



22. The cow jumped over the moon.

23. Three children are playing.



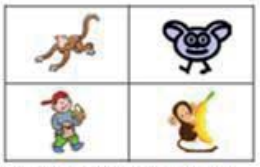
24. The horse is running.

25. She is holding the baby.

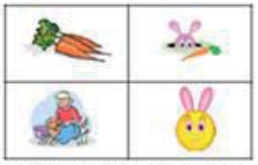
26. He is sleeping under a tree.

27. They are playing in the water.

28. The girl is holding the cat.



29. The monkey has a banana.



30. The rabbit is looking at the carrot.

Appendix 2

Sentence repetition test and scoring guide

Trials

Pen
Fish
Eight
Shoulder
Pencil
Seven
I am a boy/girl.
You are three years old.
Is it raining outside?

Test sentences and scoring guide

Sentence	*Morphemes*	*Total*
1. Birds fly.	bird+s fly+Ø	4
2. Horses run fast.	Horse+s run+Ø fast	5
3. The baby drinks milk.	the baby+Ø drink+s milk+Ø	7
4. She is jumping.	she+x is+x jump+ing	6
5. He is eating a banana.	he+x is+x eat+ing a/ banana+Ø	9
6. She has washed the dog.	she+x has+x wash+ed the dog+Ø	9
7. They have played.	they+x have+x play+ed	6
8. He is not young.	he+x is+x not young	6
9. Cats do not fly.	cat+s do+Ø not fly+Ø	7
10. Is the elephant playing?	is+x the elephant+Ø play+ing	7
11. Does she sing?	do+es she+x sing+Ø	6
12. Do you like to read?	do+Ø you+x like+Ø to read+Ø	9
13. Are the children at school?	are+x the child+ren at school+Ø	8
14. The cat is playing with a ball.	the cat+Ø is+x play+ing with a/ ball+Ø	11
15. The children's dog always sleeps on the floor.	the child+ren+'s dog+Ø always sleep+s on the floor+Ø	13
16. The boy has been playing with his friends for an hour.	the boy+Ø has+x be+en play+ing with his+x friend+s for an/ hour+Ø	18
17. The house with the blue door looks nice.	the house+Ø with the blue door+Ø look+s nice	11
Total:		142

Ø = point for zero morpheme
x = 1 point for auxiliary/pronoun
/ = 0.5 point for indefinite article, 0.5 for correct form

References

Adesope, 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*, 207-245 doi: 10.3102/0034654310368803

Aukrust, V. G. (2008). Turkish-speaking first graders in Norway acquiring second language vocabulary, listening comprehension and literacy skills. *Scandinavian Journal of Educational Research, 52*(3), 293-314. doi: 10.1080/00313830802025108

Bialystok, E. (2002). On the reliability of robustness. *Studies in Second Language Acquisition, 24*(3), 481-488. doi: 10.1017.S0272263102003054

Bialystok, E. (2009). Bilingualism: The good, the bad, and the indifferent. *Bilingualism: Language and Cognition, 12*(1), 3-11. doi: 10.1017/S1366728908003477

Bialystok, E. (2011). Coordination of executive functions in monolingual and bilingual children. *Journal of Experimental Child Psychology, 110*(3), 461-468. doi: 10.1016/j.cognition.2009.06.014

Bialystok, E., & Barac, R. (2012). Emerging bilingualism: Dissociating advantages for metalinguistic awareness and executive control. *Cognition, 122*(1), 67-73. doi: 10.1016/j.cognition.2011.08.003

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. doi: 10.1177/1529100610387084

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. doi: 10.1207/s1532799xssr0901_4

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. doi: 10.1017/S1366728909990423

Bialystok, E., McBride-Chang, C., & Luk, G. (2005). Bilingualism, language proficiency, and learning to read in two writing systems. *Journal of Educational Psychology, 97*(4), 580-590. doi: 10.1037/0022-0663.97.4.580

Bialystok, E., Peets, K. F., & Moreno, S. (2014). Producing bilinguals through immersion education: Development of metalinguistic awareness. *Applied Psycholinguistics, 35*(01), 177 - 191. doi: 10.1017/S0142716412000288

Bloom, P. (2000). *How children learn the meanings of words*. Cambridge, MA: MIT Press.

Bloom, P. (2004). Myths of Word Learning. In D. G. Hall & S. R. Waxman (Eds.), *Weaving a lexicon* (pp. 205-224). Cambridge, MA: MIT Press.

Butler, Y. G. (2012). Bilingualism/multilingualism and second-language acquisition. In T. K. Bhatia & W. C. Ritchie (Eds.), *Blackwell Handbooks in Linguistics: Handbook of Bilingualism and Multilingualism* (2 ed., pp. 109-136). Somerset, NJ: Wiley.

Cardenas-Hagan, E., Carlson, C. D., & Pollard-Durodola, S. D. (2007). The cross-linguistic transfer of early literacy skills: The role of initial L1 and L2 skills and language of instruction. *Lang Speech Hear Serv Sch, 38*(3), 249-259. doi: 10.1044/0161-1461(2007/026)

Carroll, S. E. (2004). Segmentation: Learning how to 'hear words' in the L2 speech stream. *Transactions of the Philological Society, 102*(2), 227-254. doi: 10.1111/j.0079-1636.2004.00136.x

Cook, V. (1999). Going beyond the native speaker in language teaching. *TESOL Quarterly, 33*(2), 185-209. doi: 10.2307/3587717

Craik, F. F., & Bialystok, E. (2005). Intelligence and executive control: Evidence from aging and bilingualism. *Cortex, 41*(2), 222-224. doi: 10.1016/S0010-9452(08)70899-2

Cummins, J. (1979). Linguistic interdependence and the educational development of bilingual children. *Review of Educational Research, 49*(2), 222-251. doi: 10.3102/00346543049002222

Cummins, J. (2007). Promoting literacy in multilingual contexts *What Works? Research into Practice* (Vol. 5, pp. 1-4). Ontario: Literacy and Numeracy Secretariat.

Dahl, A. (n.d.). *Sentence repetition as a measure of early L2 development*. Trondheim, Norway.

Dahl, A. (to appear). Input and language competence in early-start foreign language classrooms. In T. Cadierno & S. W. Eskildsen (Eds.), *Proceedings: Thinking, Doing, Learning: Usage Based Perspectives on Second Language Learning*. Trondheim, Norway: Mouton.

De Houwer, A., Bornstein, M. H., & Putnick, D. L. (2013). A bilingual–monolingual comparison of young children's vocabulary size: Evidence from comprehension and production. *Applied Psycholinguistics, FirstView*, 1-23. doi: 10.1017/S0142716412000744

DeKeyser, R. M. (2000). The robustness of Critical Period effects in second language acquisition. *Studies in Second Language Acquisition, 22*(4), 499-533.

Dickinson, D. K., Flushman, T. R., & Freiberg, J. B. (2009). Vocabulary, reading and classroom supports for language. In B. J. Richards, H. M. Daller, D. D. Malvern, P. Meara, J. Milton & J. Treffers-Daller (Eds.), *Vocabulary studies in first and second language acquisition: the interface between theory and application* (pp. 23-39). Houndmills: Palgrave Macmillan.

Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test Manual* (4 ed.). Minneapolis, MN: Pearson Education, Inc.

Edwards, V. (2009). Vocabulary, education and diversity. In B. J. Richards (Ed.), *Vocabulary studies in first and second language acquisition: The interface between theory and application* (pp. XIX, 248 s. : ill.). Houndmills: Palgrave Macmillan.

Elllo. (2013a). Listening Games: Fruit. Retrieved 08.29.13, 2013, from http://www.elllo.org/english/Games/G008-Fruit.html

Elllo. (2013b). Listening Games: Pets. Retrieved 08.29.13, 2013, from http://www.elllo.org/english/Games/G039-Pets.html

Farrant, B. M., & Zubrick, S. R. (2013). Parent-child book reading across early childhood and child vocabulary in the early school years: Findings from the Longitudinal Study of Australian Children. *First Language, 33*(3), 280-293. doi: 10.1177/0142723713487617

Gathercole, V. C. M., Thomas, E. M., & Hughes, E. (2008). Designing a normed receptive vocabulary test for bilingual populations: A model from Welsh. *International Journal of Bilingual Education and Bilingualism, 11*(6), 678-720. doi: 10.1080/13670050802149283

Genesee, F. (2004). What do we know about bilingual education for majority-language students? In T. K. Bhatia & W. C. Ritchie (Eds.), *The Handbook of Bilingualism* (pp. 547-576). Malden, MA: Blackwell Pub.

Gibson, E., & Pearlmutter, N. J. (1998). Constraints on sentence comprehension. *Trends in Cognitive Sciences, 2*(7), 262-268. doi: 10.1016/S1364-6613(98)01187-5

Gorman, B. K. (2012). Relationships between vocabulary size, working memory, and phonological awareness in Spanish-speaking English language learners. *American Journal of Speech-Language Pathology, 21*(2), 109-123. doi: 10.1044/1058-0360(2011/10-0063)

Grosjean, F. (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language*(36), 3-15. doi: 10.1016/0093-934X(89)90048-5

Grosjean, F. (1997). The bilingual individual. *International Journal of Research and Practice in Interpreting, 2*(1), 163 –187. doi: 10.1075/intp.2.1-2.07gro

Grosjean, F. (2010). *Bilingual: Life and reality*. Cambridge, MA: Harvard University Press.

Hyltenstam, K., & Abrahamsson, N. (2003). Maturational constraints in SLA. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 539-588). Oxford: Blackwell.

Jessop, L., Suzuki, W., & Tomita, Y. (2007). Elicited imitation in second language acquisition research. *The Canadian Modern Language Review/La Revue canadienne des langues vivantes, 64*(1), 215-238. doi: 10.3138/cmlr.64.1.215

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology, 21*(1), 60-99.

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Brief Intelligence Test, second edition manual*. Minneapolis, MN: Education, Inc.

Kroll, J. F., & Dussias, P. E. (2004). The comprehension of words and sentences in two languages. In T. K. Bhatia & W. C. Ritchie (Eds.), *The handbook of bilingualism* (pp. 169-200). Malden, MA: Blackwell.

Lauchlan, F., Parisi, M., & Fadda, R. (2013). Bilingualism in Sardinia and Scotland: Exploring the cognitive benefits of speaking a 'minority' language. *International Journal of Bilingualism, 17*(1), 43-56. doi: 10.1177/1367006911429622

Lee, J. (2011). Size matters: Early vocabulary as a predictor of language and literacy competence. *Applied Psycholinguistics, 32*(1), 69-92. doi: 10.1017/S0142716410000299

Lenneberg, E. H. (1967). *Biological foundations of language*. New York, NY: Wiley.

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 Child Psychiatry, 51*(5), 612-620. doi: 10.1111/j.1469-7610.2009.02185.x

Locke, J. L. (1997). A theory of neurolinguistic development. *Brain and Language, 58*(2), 265-326. doi: 10.1006/brln.1997.1791

Long, M. H. (1990). Maturational constraints on language development. *Studies in Second Language Acquisition, 12*(3), 251-285. doi: 10.1017/S0272263100009165

Luk, G., De Sa, E., & Bialystok, E. (2011). Is there a relation between onset age of bilingualism and enhancement of cognitive control? *Bilingualism, 14*(4), 588-595. doi: 10.1017/S1366728911000010

Meisel, J. (2004). The bilingual child. In T. K. Bhatia & W. C. Ritchie (Eds.), *The handbook of bilingualism* (pp. 91-113). Malden, MA: Blackwell.

Meisel, J. (2009). Second language acquisition in early childhood *Zeitschrift für Sprachwissenschaft* (Vol. 28, pp. 5-34).

Melby-Lervåg, M., & Lervåg, A. (2013). Reading comprehension and its underlying components in second-language learners: A meta-analysis of studies comparing first- and second-language learners. *Psychological Bulletin*, 409-433. doi: 10.1037/a0033890

Nicolay, A.-C., & Poncelet, M. (2013). Cognitive advantage in children enrolled in a second-language immersion elementary school program for three years. *Bilingualism, 16*(3), 597-607. doi: 10.1017/S1366728912000375

Oller, D. K., Pearson, B. Z., & Cobo-Lewis, A. B. (2007). Profile effects in early bilingual language and literacy. *Applied Psycholinguistics, 28*(2), 191-230. doi: 10.1017/S0142716407070117

Ortega, L. (2014). Ways forward for a bi/multilingual turn in SLA. In S. May (Ed.), *The multilingual turn: Implications for SLA, TESOL and bilingual education* (pp. 32-53). New York: Routledge.

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.

Penfield, W., & Roberts, L. (1959). *Speech and brain-mechanisms*. Princeton, NJ: Princeton University Press.

Pérez-Cañado, M. L. (2012). CLIL research in Europe: Past, present, and future. *International Journal of Bilingual Education and Bilingualism, 15*(3), 315-341. doi: 10.1080/13670050.2011.630064

Ricketts, J., Nation, K., & Bishop, D. V. M. (2007). Vocabulary is important for some, but not all reading skills. *Scientific Studies of Reading, 11*(3), 235-257. doi: 10.1080/10888430701344306

Ruben, R. J. (1997). A time frame of critical/sensitive periods of language development. *Acta oto-laryngologica, 117*(2), 202-205. doi: 10.1007/BF02996542

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. doi: 10.1017/S0142716409990191

Sears, C. (1998). *Second language students in mainstream classrooms: A handbook for teachers in international schools*. Clevedon: Multilingual Matters.

Seliger, H. W. (1978). Implications of a multiple critical periods hypothesis for second language learning. In W. C. Ritchie (Ed.), *Second language acquisition research* (pp. 11-19). New York, NY: Academic Press.

Serratrice, L. (2012). The bilingual child. In T. K. Bhatia & W. C. Ritchie (Eds.), *Handbook of bilingualism and multilingualism* (2 ed., pp. 87-108). Somerset, NJ: Wiley.

Van Moere, A. (2012). A psycholinguistic approach to oral language assessment. *Language Testing, 29*(3), 325-344. doi: 10.1177/0265532211424478

Verhoeven, L., van Leeuwe, J., & Vermeer, A. (2011). Vocabulary growth and reading development across the elementary school years. *Scientific Studies of Reading, 15*(1), 8-25. doi: 10.1080/10888438.2011.536125

Vulchanova, M., Vulchanov, V., Sarzhanova, D., & Eshuis, H. (2012). The role of input in early bilingual lexical development. *Lingue e linguaggio*(2), 181-198. doi: 10.1418/38785

Wei, L. (2000). Dimensions of bilingualism. In L. Wei (Ed.), *The bilingualism reader* (pp. 2-21). London: Routledge.

Wesche, M. B. (2002). Early French immersion: How has the original Canadian model stood the test of time? In P. Burmeister, T. Piske & A. Rohde (Eds.), *An integrated view of language development: Papers in honour of Henning Wode* (pp. 357-379). Trier: Wissenschaftlicher Verlag Trier.

Wode, H. (1999). Language learning in European immersion classes. In J. Masih (Ed.), *Learning through a foreign language: Models, methods and outcomes* (pp. 16-29). Grantham: Grantham Book Services.