Early-English education works no miracles

Cognitive and linguistic development of mainstream, early-English, and bilingual primary-school pupils in the Netherlands

Published by
LOT
Trans 10
3512 JK Utrecht
The Netherlands

phone: +31 30 253 6111

e-mail: lot@uu.nl
http://www.lotschool.nl

Cover illustration: Lysanne de Water

ISBN: 978-94-6093-316-5
NUR 616

Early-English education works no miracles

Cognitive and linguistic development of mainstream, early-English, and bilingual primary-school pupils in the Netherlands

Proefschrift

ter verkrijging van de graad van doctor

aan de Radboud Universiteit Nijmegen

op gezag van de rector magnificus prof. dr. J.H.J.M. van Krieken,

volgens besluit van het college van decanen

in het openbaar te verdedigen op vrijdag 22 februari 2019

om 12.30 uur precies

door

Claire Maria Michèle Goriot

geboren op 8 december 1990

te Boxmeer

Promotoren: Prof. dr. R.W.N.M. van Hout

Prof. dr. J.M. McQueen

Copromotoren: Dr. M.E. Broersma

Dr. S. Unsworth

Manuscriptcommissie:

Prof. dr. J.J.M. Schoonen

Prof. dr. W.B.T. Blom (Universiteit Utrecht)

Prof. dr. H.C.J. de Graaff (Universiteit Utrecht)

Prof. dr. C. Muñoz (Universitat de Barcelona, Spanje)

Dr. G.J. Poarch (Universität Münster, Duitsland)

Table of Contents

Chapter 1: Introduction

Chapter 1

Introduction

1.1 Introduction

A slow revolution is taking place in primary education in Europe. All across Europe, including the Netherlands, a trend can be observed towards lowering the starting age for foreign language lessons in primary education. This international trend has a strong bias for lessons in English. Today, approximately one in five primary schools in the Netherlands provide English lessons from the moment that pupils enter school, that is, when they are four years old (Nuffic, 2018a). While this number is rapidly increasing – from less than 100 schools in 2004 to a staggering 1150 in 2018 (Nuffic, 2018a) – little is known about the effects on pupils attending these schools. How does the early start with English education affect their cognitive and linguistic development? This crucial question is the topic of this dissertation.

Pupils who attend schools in which foreign language lessons are given from a young age are exposed to another language at school at a much younger age than their peers at mainstream schools, where English language instruction does not start until they are approximately ten years old. Mainstream-school pupils do however also have some exposure to English before English instruction starts, since in the Netherlands English is prominent in everyday society, as for example, English television programs are subtitled rather than dubbed (Kuppens, 2010).

Both mainstream and early-English pupils are thus likely to have knowledge of and at least some experience with English. Their experience does however not compare to that of simultaneous bilinguals, who learn two languages from a very young age (i.e., younger than four years) onwards, and who use both languages on a daily basis. Previous research has shown that bilingual experience may influence children's cognitive and linguistic development (see for example Barac, Bialystok, Castro, & Sanchez, 2014). It may be assumed that early-English pupils' English proficiency is higher than that of mainstream pupils, but lower than that of simultaneous bilinguals. On a continuum ranging between a more monolingual and a more bilingual side, early-English pupils can be placed somewhere in the middle, having more bilingual experience than mainstream pupils, but not as much as simultaneous bilinguals. The main aim of this thesis is to investigate whether early-English pupils, who are learning a second language (L2) in an instructed setting and for a limited amount of time, differ in their development of executive functions, phonological awareness, and the perception of English speech sounds from mainstream pupils, and to what extent their development in these domains resembles that of simultaneous Dutch-English bilinguals.

1.1.1 English education in Dutch primary schools: status quo and recent changes

The Dutch primary-school system consists of eight grades, of which the two first grades are Kindergarten. Almost all pupils (98%) enter the first grade when they turn four years old (although compulsory education only starts at the age of five), and leave grade eight when they are approximately twelve years old (Centraal Bureau voor de Statistiek, 2003). Currently, almost one in five primary schools in

the Netherlands start with foreign language lessons from the start of primary school. Although it is possible for schools to offer French, German, or Spanish, the large majority (> 90%) of those schools offer English as the foreign language (Nuffic, 2018b). This type of education, called 'early-English', is relatively new. At the introduction of English as an obligatory subject in primary education in 1986 the Minister of Education already allowed schools to start English lessons in an earlier grade than the penultimate one (Thijs, Trimbos, Tuin, Bodde, & de Graaff, 2011); yet it was not until the turn of the current century that the first schools started to actually do this. Before then, schools generally started English education in the penultimate grade (grade 7), as is still the case in the majority of schools. To date, four types of English education can be distinguished. In addition to mainstream education and early-English education, some schools start English lessons at an earlier point in the curriculum than grade 7 but not in the first grade. Very recently a fourth option has been added, which is bilingual primary education. In these schools, 30-50% of the lessons are in English from grade 1 onwards, as opposed to maximally 15% in early-English schools. This type of education is however still in its pilot phase, and only 19 schools were allowed to provide bilingual primary education (Driessen et al., 2016; Jenniskens et al., 2018; Nuffic, 2018). In this thesis, the focus is on the two most common types: mainstream English education and early-English education.

That many schools opt for English as the foreign language is not surprising. English was already the language taught as the primary foreign language in mainstream education. In addition, English is the lingua franca of the world. The Educational Council explicitly recommended English as the foreign language, since Dutch children are exposed to that language from a young age. The council suggested that German or French would be an option for schools in the border regions (Onderwijsraad, 2008). Dutch natives are known for their generally good command of English. According to the international English Proficiency Index by Education First, the Netherlands ranks first of 80 countries, indicating that Dutch citizens have a very high level of English proficiency (EF EPI, 2017). Despite the prominence of English in everyday life in the Netherlands, and despite the overall high English proficiency of its inhabitants, Dutch parents (Driessen et al., 2016) and policy makers (Onderwijsraad, 2008) are convinced that an early start of English lessons will help children to become fluent speakers of English.

Early-English and mainstream English lessons differ in several ways, other than in their starting point. First, they differ in the amount of time devoted to English. When English lessons were introduced in 1986, the idea was that by introducing English in the final grades of primary school, pupils would already have acquired basic knowledge of the English language when starting English lessons in secondary school. An important argument to introduce English in primary schools was European policy regarding foreign language education: With the goals of cultural diversity, tolerance, and European citizenship in mind, the European Union promoted language learning in primary education in all its countries (European Council, 2002; Thijs et al., 2011). It was decided that at the end of primary school, pupils should have had 80 hours of English education, which was deemed sufficient to reach these goals. If schools start in the penultimate grade with English lessons, this means that pupils should get one hour of English lessons

per week (Thijs et al., 2011). In practice, however, mainstream schools spend approximately 45 minutes per week on English lessons (Thijs et al., 2011). Early-English schools, on the contrary, are allowed to teach up to 15% of the teaching time (3.5 hours per week) in English. A recent investigation (Jenniskens et al., 2017) showed that in practice, most early-English schools do not spend more than two hours per week on English lessons, with the majority (68%) spending 30 minutes to one hour per week on English.

Another difference between early-English and mainstream English schools lies in the objectives associated with each programme. When the Ministry of Education started to actively encourage primary schools to lower the starting age for English lessons in 2004 (Ministerie van Onderwijs Cultuur en Wetenschap, 2005), rather than preparing them for secondary education, the goal was for them to become more proficient speakers of English in the long term. The rationale behind this policy was the assumption that children acquire a foreign language more easily at a younger than at a later age. The motivation for promoting early foreign language learning was merely economic in nature: according to the Ministry, being proficient in English is indispensable in a time of increasing globalization, and is important for the Dutch trade position. In 2008, the European Union stated that its goal was to strive to multilingualism: Every European citizen should be proficient in two languages, in addition to their own mother tongue (The Council of the European Union, 2008). One of these languages should be a lingua franca. To reach this goal, the European Union called for foreign language education for pupils under the age of ten. In that same year, the Dutch Educational Council, the independent advisory board for educational policies, advised that the starting age for English lessons should be lowered to either the start or the middle of primary school (Onderwijsraad, 2008). According to the council, young children would already be motivated to learn a foreign language, and by introducing this language at an earlier point in the curriculum, children would master English at a younger age.

As a consequence of these different expectations, the objectives for early-English education are higher than for mainstream English lessons. Pupils in mainstream schools should be able to reach minimally the lowest level (A1) of the Common European Framework of Reference for Languages (CEFRL; Council of Europe, 2018) (Bodde, Schippers, Klein Tank, & van der Linde-Meijerink, 2008). The objectives for pupils attending early-English schools depend on the type of secondary school they will afterwards attend. The Dutch system is selective. Depending on pupils' academic achievements, they will either go to lower general secondary education (Dutch: *VMBO*), higher general secondary education (*HAVO*), or pre-university education (*VWO*) after primary school. For those who will go to lower general secondary education, the objective, as for mainstream pupils, is set at A1. Pupils leaving primary school for higher general secondary education should be able to reach A2 level. Pupils who will attend pre-university education should reach A2 level for reading and writing, and even B1 level for communication in English (Nuffic, n.d.). Teachers are not required to test whether pupils actually reach these levels. Only 3% of the teachers who test pupils' proficiency level in English make use of a standardised test that measures multiple language development domains, and 18% of the teachers do not test pupils' level at all. Others make use of a test

that they designed themselves, or of a test provided by the makers of the teaching method they use (Thijs et al., 2011).

The third way in which Early-English and mainstream lessons differ is with respect to the required proficiency levels of the teachers. Nuffic, the organisation for internationalisation in education, states that teachers who are responsible for early-English lessons should have B2 level on the scale of the CEFRL, except for writing for which B1 level is deemed sufficient (Nuffic, n.d.). In addition, teachers should have followed a course of 'Classroom English', which not only aims at improving teachers' proficiency in English, but also at providing teachers with didactical skills that are needed to teach (in) English (Nuffic, n.d.). In reality, not all early-English teachers obtain the required B2 level. An investigation of 9756 teachers showed that 28% had a proficiency level lower than B2, and for more than half of the participating teachers the proficiency level was unknown (Jenniskens et al., 2017). For mainstream English education, on the other hand, no such standard has been defined. Before English was introduced in 1986, several measures were taken in order to ensure a smooth integration of English in primary education: teaching materials were developed, and English became an obligatory subject in teacher training institutes in 1983 already. Furthermore, in-service primary-school teachers got additional training in teaching English. Nowadays, English lessons are still part of the teacher training programme, but the way in which English lessons are structured differs between the various teacher-training institutes. In general, the subject receives little attention (Thijs et al., 2011). Nevertheless, every teacher who has obtained a diploma with a specialisation for teaching in the upper primary grades is qualified to teach mainstream English lessons.

Nuffic has developed a document with quality standards for early-English education (Nuffic, n.d.). In addition to the English proficiency level of teachers, the document contains information about how English lessons ought to be structured. It states for example that English should be used as the language of instruction and as the language of communication, that at least 60 minutes per week should be devoted to English lessons, and that pupils' progress in English should be monitored and documented. Early-English schools can ask one of three independent organisations to evaluate their English education, and upon meeting these requirements, a certificate is issued. Less than 10% of the early-English schools hold such a certificate (Nuffic, 2017).

1.1.2 Effects of early-English education on pupils' development of English

In the Dutch context, several studies have been conducted to investigate whether early-English education is beneficial to pupils' knowledge of English. One of the first (see also Aarts & Ronde, 2006) was conducted by Goorhuis-Brouwer and de Bot (2010), who investigated receptive and productive English vocabulary development in a longitudinal study, testing pupils' skills when they were in their first year of primary school, and a year later. The authors compared pupils' scores to the norm scores of the test. The conclusion was that in the first year of primary school, pupils substantially improved in their English vocabulary knowledge, as already after six months of schooling pupils' scores were, on average, comparable to the norm score for two-year-old English monolinguals. In this study unfortunately

no pre-test was conducted before the children entered primary school and no control group was used, limiting the interpretations that can be drawn from the results.

Lobo (2013) showed that children in first grade (4-5 years of age) and third grade (6-7 year-olds) showed significant improvement in their English receptive vocabulary and pronunciation scores after a total of only 10 hours of either art and crafts or physical education lessons in English. Another study investigating pupils' receptive English vocabulary (Unsworth, Persson, Prins, & de Bot, 2015) showed, like Lobo (2013), that early-English pupils outperformed their peers on an English vocabulary test, but they also showed that this was only the case for pupils who got more than 60 minutes of English lessons per week. In other words, children who followed early-English instruction for 60 minutes or less per week were not significantly better than their peers at mainstream schools. In addition, pupils who were educated by a teacher with a proficiency at the B level on the CEFRL only, showed slower vocabulary development than pupils who were educated by a teacher who had a higher proficiency level or a B-level teacher with a native English co-teacher. This effect of the teacher's proficiency was also found for receptive grammar; early-English pupils however always outperformed mainstream pupils, including those with 60 minutes or less of English lessons per week (Unsworth et al., 2015). Van der Leij, Bekebrede, and Kotterink (2010) investigated English language development in somewhat older pupils in a longitudinal study, with the first measurement when pupils were in grade 4 (approximately eight years old), and the second in grade 5. Pupils had started to learn to read in English from grade 3 onwards. At both time points, early-English pupils outperformed their peers from mainstream schools on an English vocabulary task, as well as a reading fluency task in English.

Up to now, only one study has investigated pupils' English language skills at the end of primary school. De Graaff (2015) compared English listening, reading, writing, and speaking skills of early-English and mainstream pupils in grade eight. Whereas the early-English pupils had received English lessons for five to eight years, the mainstream pupils had started English lessons only a little more than a year earlier (in grade seven). On average, early-English pupils scored better on all tasks in English, although there were large individual differences, with some mainstream pupils outperforming some early-English pupils.

In summary, it seems that pupils in early-English schools, at least those with more than 60 minutes of exposure to English and a teacher with a higher proficiency level, show more improvement than pupils from Dutch mainstream schools in English vocabulary knowledge (Goorhuis-Brouwer & de Bot, 2010; Unsworth et al., 2015; van der Leij et al., 2010), grammar (Unsworth et al., 2015), reading fluency (van der Leij et al., 2010), and listening, reading, writing, spelling, and speaking in English (de Graaff, 2015). At the same time, however, their *Dutch* vocabulary seems to develop at the same pace as that of their peers at mainstream schools (Goorhuis-Brouwer & de Bot, 2010; van der Leij et al., 2010). Van der Leij et al. (2010) even found that pupils from an early-English school outperformed their mainstream education peers on Dutch reading fluency and reading comprehension tests. Contrary to what some parents and teachers may think (Goorhuis-Brouwer & de Bot, 2010), learning English at a young age is (at least in this respect) not detrimental to the development of the mother tongue.

Early-English pupils are L2 learners, and as such they are expected to become more proficient in the L2 over time; moreover, they can be expected to become more proficient in the L2 than their mainstream peers. Evidence for this expectation has been provided for pupils in the lower grades of primary school (Unsworth et al., 2015; van der Leij et al., 2010). Although not much is known about what happens during and after several years of early-English education, when pupils are in the upper grades of primary school, the expectation is that pupils' proficiency in the L2 will only increase. These early-English pupils arguably start on the monolingual side of a continuum ranging from monolingual language proficiency to bilingual language proficiency. As they grow older and receive more English education, and thus become more experienced L2 learners, they shift towards the bilingual side of the continuum – although it is unlikely that they will reach full bilingual proficiency.

Learning an L2 may influence pupils' development in several ways. Not only will pupils become more proficient in the L2, learning an L2 may also influence other domains of their development. Research with proficient L2 learners and simultaneous bilingual children has suggested that the development of these children is different from that of monolinguals. Children who grow up with two languages will be more proficient speakers of those languages later on. They will, for example, show better discrimination of the speech sounds of these languages than those who learned the new language at a later age (Baker, Trofimovich, Flege, Mack, & Halter, 2008; McCarthy, Mahon, Rosen, & Evans, 2014). Bilingual children and proficient L2 learners should also show faster development of, for example, executive functions, metalinguistic skills including phonological awareness, theory of mind understanding, and memory skills (Adesope, Lavin, Thompson, & Ungerleider, 2010; Barac et al., 2014). Previous studies have yielded contradictory results however, and some researchers doubt the presence of a bilingual advantage in these domains (de Bruin, Treccani, & Della Sala, 2015; Duñabeitia et al., 2014; Paap et al., 2017; Paap, Johnson, & Sawi, 2014, 2015; Paap & Greenberg, 2013). Questions about whether such developmental differences between monolingual and bilingual children exist, for what groups of children, and under which circumstances they may occur, persist. As few studies have included pupils with only little exposure to the L2, and previous research on early-English education in the Netherlands has focused primarily on pupils' English language development, it is unknown whether such limited bilingual experience has an influence on pupils' linguistic development (other than receptive vocabulary and grammar skills), or on their cognitive development. These influences are investigated in this thesis.

1.2 Domains of enquiry and research questions

This thesis focuses on the development of Dutch-speaking pupils who are learning English as an L2. The development of these pupils is compared to that of Dutch-English simultaneous bilingual children, who have learned both languages from birth (or at least before attending primary school), and who use both languages on a daily basis. Despite this narrow focus on one language pair only, the results of this thesis are relevant to speakers of other language pairs. Although some of the outcomes reported in this thesis will be language-pair specific, the outcomes in

general may show whether limited bilingual experience influences pupils' development in different domains.

The main question of this thesis was the following: does the cognitive and linguistic development of early-English pupils differ from that of their mainstream peers, and to what extent is it comparable to that of simultaneous bilingual children? The focus of this thesis is on three domains of children's development: executive functions, phonological awareness, and English speech perception. They were chosen because they have been the subject to (considerable) research in recent years, children show development in these domains during their primary-school years (Diamond, 2013; Hazan & Barrett, 2000; Sodoro, Allinder, & Rankin-Erickson, 2002), and bilingual children may show differences in development in these domains compared to monolingual children (Barac et al., 2014; Bosch & Sebastián-Gallés, 2003a, 2003b).

1.2.1 Executive functions

The first topic of interest was pupils' development of executive functions. Executive functions are brain processes that are involved in the execution of goal-directed behaviour (Diamond, 2013). These brain processes are generally divided in three sub-processes, according to the model by Miyake et al. (2000): switching (the ability to shift between changing rules), inhibition (the capacity to overcome an irrelevant response), and working memory (the verbal or non-verbal ability to hold information in mind and to manipulate it). Because bilingual children have to constantly monitor the competition between their two activated languages, their executive functions should be more developed than those of monolingual children (Green, 1998; Green & Abutalebi, 2013). As a consequence, bilingual children might show more accurate and/or faster responses on tasks measuring executive functioning (Adesope et al., 2010; Barac et al., 2014).

The topic of the development of executive functions in bilingual children is among the most debated research domains in bilingualism research, with studies reporting contradictory findings. Some studies have provided evidence for the claim that bilingual children show better performance on executive functions tasks than monolingual children (Morales, Calvo, & Bialystok, 2014; Poarch & Bialystok, 2015; Poarch & van Hell, 2012), but others did not find such differences (Antón et al., 2014; Duñabeitia et al., 2014; Paap et al., 2017; Ratiu & Azuma, 2014).

The question arises which groups of (bilingual) children show developmental differences, and in which aspect of executive functions. Some researchers have suggested that especially L2 learners would show enhanced executive functioning development, since managing ongoing competition between two languages is not automatized yet for this group of emerging bilinguals (Gathercole et al., 2014). Several studies have shown that L2 learners may indeed show enhanced executive functioning development, but that such enhancement takes place especially when children's proficiency in their two languages is well matched (Blom, Küntay, Messer, Verhagen, & Leseman, 2014; Thomas-Sunesson, Hakuta, & Bialystok, 2018; Vega & Fernandez, 2011).

Whereas the L2 learners in previous studies were all immersed in the L2, and therefore relatively proficient in that language, the question thus remains whether pupils who only got little exposure to the L2 would also show enhanced

executive functioning performance. Chapter 2 asks whether Dutch early-English pupils show differences in the development of executive functions (switching, inhibition, working memory) in comparison to Dutch pupils attending a mainstream school. I investigated this question in a cross-sectional design. Children had just started primary school (4-5 year-olds), were halfway their primary-school career (8-9 year-olds), or were in the final grade of primary school (11-12 year-olds). I examined possible differences in executive functions between groups of mainstream and early-English pupils. In addition, I used a measure of lexical balance (English vocabulary relative to Dutch vocabulary) as a more continuous measure of bilingualism, and investigated whether balance scores are related to pupils' executive functions.

1.2.2 Phonological awareness

The second set of skills I investigated was pupils' phonological awareness skills. Phonological awareness is the ability to ignore the meaning of words, and to pay attention to the different sounds of words, as well as being able to manipulate those sounds (Sodoro et al., 2002). Just as for executive functions, research findings on phonological awareness in bilingual children have been mixed (Barac et al., 2014).

Previous research on phonological awareness in monolingual and bilingual children shows even more mixed results than research comparing monolingual and bilingual children on the development of executive functions. The development of phonological awareness is at its peak when children are in their preschool and early elementary school years (Anthony & Francis, 2005), at which age early-English pupils are also exposed to an L2. Since well-developed phonological awareness skills are an important predictor of successful reading and writing development (Anthony & Francis, 2005; Sodoro et al., 2002), it is important to investigate whether this type of education affects – that is, hinders or helps – the development of phonological awareness.

Early research on phonological awareness in bilingual children has suggested that bilinguals may have an advantage compared to monolingual children (Bruck & Genesee, 1995; Rubin & Turner, 1989). Later studies, however, have shown mixed results. Sometimes, bilinguals are found to outperform monolinguals on phonological awareness (Bialystok, Luk, & Kwan, 2005; Chen, Wu, & Shu, 2004; Kuo & Anderson, 2010; Loizou & Stuart, 2003; Marinova-Todd, Zhao, & Bernhardt, 2010), but other studies did not find any differences (Bialystok, Majumder, & Martin, 2003; Reder, Marec-Breton, Gombert, & Demont, 2013), or even found a monolingual advantage (Janssen, Segers, McQueen, & Verhoeven, 2015, 2017; Marinova-Todd et al., 2010).

Whether bilingual children differ in phonological awareness from their monolingual peers seems to depend on various factors. These factors include language-specific features, and the specific combination of languages spoken by the bilingual group (Barac et al., 2014), whether children already started literacy instruction or not (Bruck & Genesee, 1995; Reder et al., 2013), and children's age (Branum-Martin, Tao, Garnaat, Bunta, & Francis, 2012; Carroll, Snowling, Stevenson, & Hulme, 2003). Research on cross-linguistic transfer between phonological awareness skills in two languages suggest that other factors may also

play a role, such as the proficiency in the first language (L1) and the L2 (Atwill, Blanchard, Gorin, & Burstein, 2007), and memory development (Janssen et al., 2017). In contrast to the development of executive functions, amount of bilingual experience seems to play a much smaller role in the development of phonological awareness: even children with limited experience with an L2 can show advantages in phonological awareness relative to their monolingual peers (Chen, Xu, Nguyen, Hong, & Wang, 2010; Kang, 2012). Nevertheless, the pupils in the Chen et al. (2010) and Kang (2012) study got at least 80 minutes per week of English education, which is more than the 60 minutes that are common in Dutch early-English schools. In addition, the control groups in these studies also received English lessons, although to a much lesser extent. It is thus still unclear what level of L2 proficiency is needed for phonological awareness skills to surpass those of monolingual children.

The main question that is addressed in Chapter 3 is whether mainstream pupils, early-English pupils, and simultaneous Dutch-English bilingual children differ from each other in their development of phonological awareness. In addition, I investigated whether individual differences in short-term memory, working memory, Dutch vocabulary, English vocabulary, or lexical balance between Dutch and English vocabulary, influence the relation between bilingualism and phonological awareness. I included children in the first three grades of primary school (4-5 year-olds, 5-6 year-olds, and 6-7 year-olds, respectively). Thereby, I was able to investigate whether possible bilingual differences in phonological awareness are more present in preliterate children than in children who already started literacy instruction.

1.2.3 Perception of English phonetic contrasts

The third area of development I examined was children's perception of English phonetic contrasts. One of the reasons for introducing a foreign language early in the primary-school curriculum was that, according to the Educational Council, young pupils would not be hindered by their L1 when acquiring the L2, and would become more proficient speakers of the L2 later on than pupils who started at a later age (Onderwijsraad, 2008). It has been claimed that especially for the perception of non-native speech sounds, an early start is indispensable (Singleton & Ryan, 2004). At the same time, research has shown that pupils who do not learn an L2 from birth onwards, but who get immersed in a new language in primary school are still able to learn the speech sounds of that language (McCarthy et al., 2014).

Research has suggested that learning to perceive the speech sounds of a foreign language is easier at an early age than at a later age (Aoyama, Flege, Guion, Akahane-Yamada, & Yamada, 2004; Baker et al., 2008; Flege, MacKay, & Meador, 1999; Tsukada et al., 2005). Tsukada et al. (2005), for example, showed that children who were immersed in an L2 environment were able to distinguish sounds of that L2 after some time. Adults who had been immersed in the L2 for the same amount of time as the children were not able to do so (Tsukada et al., 2005). As previous research was almost exclusively performed with children who were immersed in the L2, either at school or in their country of immigration, the question remains whether the limited amount of English offered in different school types in

the Netherlands leads to measurable differences in pupils' perception of English speech sounds.

In Chapter 4, I investigated whether children in early-English schools are better at perceiving differences in English speech sounds than their peers attending mainstream schools. I again compared the performance of those two groups with that of a group of simultaneous bilingual children, who grew up with both Dutch and English. This was done because the latter group of children is likely to easily discriminate the English sounds, given that they are native speakers of that language. This way, it was possible to examine to what extent early-English pupils' performance reflects that of bilinguals.

To investigate this question, I presented children with four different English phonetic contrasts, which were chosen on the basis of the predictions of the Perceptual Assimilation Model (PAM) for L2 learners (Best & Tyler, 2007). According to the PAM-L2, learners will easily be able to discriminate two phonemes that are similar to two separate phonemes in their L1 (two category assimilation). They will have a harder time to do so when both L2 phonemes map onto a single native phoneme with one being a good match and the other one being a poor match with the native sound (category goodness assimilation), and an even harder time if both of these two L2 phonemes are equally good exemplars of the L1 phoneme (single category assimilation). I tested children on four phonetic contrasts, which I expected to vary in difficulty ranging from easy via intermediate to hard and, finally, very hard.

The PAM-L2 predicts that L2 learners may initially have a hard time distinguishing certain non-native speech sounds, but that they may learn to distinguish those sounds with increasing experience in the L2. I therefore included children from early-English schools who were likely to differ in their experience with the English language: 4-5-year-old pupils who were at the start of primary school (grade 1) and who were inexperienced English learners, 8-9-year-old pupils who were at the middle of primary school (grade 5) and thus somewhat experienced with regards to English, and 11-12-year-old pupils (grade 8) who had had eight years of early-English education. Note that for the mainstream children, English education did not start until they were around ten years old (grade 7); they were thus mostly inexperienced, or, in the case of the 11-12-year-olds, only slightly experienced in English. The bilingual children had at least one parent that was a native speaker of English, and bilingual children of all ages could thus be considered very experienced with respect to English.

The aim of this study was to investigate whether starting to learn English as an L2 at a young age is beneficial to pupils' perception of non-native speech contrasts. In addition, it sought to add to our knowledge about the development of the speech system, and to determine whether, and if so, how pupils learn to discriminate non-native speech sounds as they become more experienced in the L2.

1.2.4 Measuring vocabulary development in the second language

English language-learning programmes are relatively new, and many researchers, policy makers, teachers and parents are interested in the question whether these programmes are effective. Various studies have investigated whether such programmes enhance English vocabulary knowledge. Many of those studies

(e.g., Buyl & Housen, 2014; Cohen, 2016; Dahl & Vulchanova, 2014; Unsworth et al., 2015; van der Leij et al., 2010) have used the Peabody Picture Vocabulary Test (PPVT) (Dunn, 1959; Dunn & Dunn, 2007) .

The PPVT is a receptive vocabulary test that requires the participant to choose the picture out of four response alternatives that correctly matches an orally presented word. The PPVT is relatively quick and easy to administer, and it is therefore no surprise that it is often used with bilingual children as well as second language learners (Bialystok, Luk, Peets, & Yang, 2010; Buyl & Housen, 2014; Cohen, 2016; Dahl & Vulchanova, 2014; Unsworth et al., 2015; van der Leij et al., 2010). It was however originally developed as a test to measure vocabulary development in people whose L1 is English. Even though it is stated in the manual that the test may be used to investigate vocabulary knowledge in people who are L2 learners of English (Dunn & Dunn, 2007), the question addressed in Chapter 5 is whether it is appropriate to do so.

Previous research has suggested that vocabulary tests that are developed for mother-tongue speakers of that particular language may not always be reliably used to investigate vocabulary in speakers for whom that language is their L2 (Gathercole, Thomas, & Hughes, 2008). Since early-English pupils start to acquire a new language at the moment their L1 has already largely developed, they may make use of that L1 when they start to learn the new language (Dijkstra & Van Heuven, 2002; Kroll & Stewart, 1994). I investigated two aspects of the L1 that may influence L2 vocabulary testing. The first one was lexical frequency. As children have already a relatively large but not fully developed lexicon in Dutch at the moment they start to learn English, word knowledge in English may rely on whether children know the Dutch equivalent in the first place. This may be especially the case since their exposure to English is limited. I asked whether the lexical frequency of the L1 (Dutch) translation equivalents of the English words in the PPVT-4 plays a more important role in performance on the PPVT-4 than L2 (English) frequency. Second, I investigated cognates: English words that show phonological overlap with their Dutch translation equivalents. Previous research has not only shown that children are faster to recognize cognates than to recognize non-cognates (Brenders, van Hell, & Dijkstra, 2011), but also that they are more often correct on cognate items than on other items in a test (Bosma, Blom, Hoekstra, & Versloot, 2016). I therefore investigated whether the odds that pupils responded correctly to items in the PPVT-4 were related to the items' cognate status.

In Chapter 5, I therefore investigated to what extent the PPVT-4 (Dunn & Dunn, 2007) was a reliable measure for English vocabulary in primary-school and secondary-school pupils, and to what extent L1 (Dutch) frequency and cognate status had an influence on pupils' score in their L2. To do so, I examined the PPVT-4 results of 204 primary-school and 152 secondary-school pupils. I used a cross-sectional design to investigate whether the effects of L1 and L2 characteristics would change when pupils became more experienced learners of English. Primary-school pupils were either at the start (grade 1; 4-5-year-olds), in the middle (grade 5; 8-9-year-olds), or at the end of primary school (grade 8; 11-12-year-olds). Secondary-school pupils were in the first year (12-13-year-olds), second year (13-14-year-olds), or third year (14-15-year-olds). Approximately half of the pupils were enrolled in an educational programme with intensified English lessons (n = 105

primary-school pupils attending early-English schools; n = 117 secondary-school pupils enrolled in a bilingual track). All other pupils were attending mainstream programmes. As it has been shown that children seem to be better at recognizing identical than non-identical cognates (Bosma et al., 2016), I used the Levenshtein distance as a continuous measure of cognate distance between English-Dutch translation pairs.

1.3 Research design

All studies in this thesis had a comparable research design. Data were collected in three waves, in each of which I tested pupils from mainstream and early-English schools. The design was always cross-sectional, in the sense that in every wave I included children of three different age groups. In the first wave, I collected data on children's executive functions (Chapter 2), in the second wave, I collected data on children's phonological awareness skills (Chapter 3), and in the third wave on pupils' perception of English phonetic contrasts (Chapter 4). The test batteries in all waves included the PPVT-4 (Dunn & Dunn, 2007) and the PPVT-III-NL (Dunn, Dunn, & Schlichting, 2005). The PPVT was used not only to address whether the English version is suitable for L2 learners (Chapter 5, data collected in the first wave), but also to test the hypothesis that early-English pupils' have better English vocabularies than mainstream pupils, while their Dutch vocabularies are the same. Previous research has confirmed this hypothesis (Unsworth et al., 2015; van der Leij et al., 2010), but only for pupils in the first grades of primary school. There is no previous research available on older early-English pupils' vocabulary development. As I tested pupils from various grades, ranging from the first to the final grade of primary school, the results show whether early-English pupils of different ages differ from their mainstream peers in the development of English and Dutch vocabulary. They also show whether a possible advantage for early-English pupils in English vocabulary still remains after mainstream pupils have started their English lessons.

The data on the bilingual children were collected in two different waves, and some of the children participated in both waves. Again, the design was cross-sectional, such that the same child was never tested twice on the same test. The exceptions to this are the PPVT-4, the PPVT-III-NL, and a non-verbal working memory test, as these tests were part of the design of at least two separate experiments. In those cases, there was always at least a year between the two waves in which the child participated. The procedure was kept comparable to that of the mainstream and early-English pupils.

1.4 Conclusion

In summary, I investigated three aspects of the cognitive and linguistic development of mainstream, early-English, and bilingual pupils: executive functions (Chapter 2), phonological awareness (Chapter 3), and speech perception in English (Chapter 4). In addition, I asked whether L1 frequency and cognate status have an effect on pupils' score on a widely used test to assess English vocabulary: the PPVT-4 (Chapter 5).

The work described in this thesis contributes to our knowledge about the cognitive and linguistic development of bilingual children and the question of how

little (or how much) bilingual experience is needed for development to become different from that of monolingual children. This work also contributes to our understanding of the influence that early-English lessons have on pupils' development, other than their development in English per se. These findings are valuable given the ever-increasing number of primary schools in the Netherlands, and in other countries in Europe, that provide foreign-language lessons from kindergarten onwards.

1.5 References

Aarts, R., & Ronde, S. (2006). Tweetalig onderwijs met vervroegd Engels in het basisonderwijs [Bilingual education with early English in primary education]. *Levende Talen Tijdschrift*, *7*, 3–14.

Adesope, O. O., Lavin, T., Thompson, T., & Ungerleider, C. (2010). A systematic review and meta-analysis of the cognitive correlates of bilingualism. *Review of Educational Research*, *80*, 207–245. http://doi.org/10.3102/0034654310368803

Anthony, J. L., & Francis, D. J. (2005). Development of phonological awareness. *Current Directions in Psychological Science*, *14*(5), 255–259. http://doi.org/10.1111/j.0963-7214.2005.00376.x

Antón, E., Duñabeitia, J. A., Estévez, A., Hernández, J. A., Castillo, A., Fuentes, L. J., … Carreiras, M. (2014). Is there a bilingual advantage in the ANT task? Evidence from children. *Frontiers in Psychology*, *5*, 398. http://doi.org/10.3389/fpsyg.2014.00398

Aoyama, K., Flege, J. E., Guion, S. G., Akahane-Yamada, R., & Yamada, T. (2004). Perceived phonetic dissimilarity and L2 speech learning: the case of Japanese /r/ and English /l/ and /r/. *Journal of Phonetics*, *32*(2), 233–250. http://doi.org/10.1016/S0095-4470(03)00036-6

Atwill, K., Blanchard, J., Gorin, J. S., & Burstein, K. (2007). Receptive vocabulary and cross-language transfer of phonemic awareness in kindergarten children. *The Journal of Educational Research*, *100*(6), 336–346. http://doi.org/10.3200/JOER.6.336-346

Baker, W., Trofimovich, P., Flege, J. E., Mack, M., & Halter, R. (2008). Child–adult differences in second-language phonological learning: The role of cross-language similarity. *Language and Speech*, *51*(4), 317–342. http://doi.org/10.1177/0023830908099068

Barac, R., Bialystok, E., Castro, D. C., & Sanchez, M. (2014). The cognitive development of young dual language learners: A critical review. *Early Childhood Research Quarterly*, *29*, 699–714. http://doi.org/10.1016/j.ecresq.2014.02.003

Bialystok, E., Luk, G., & Kwan, E. (2005). Bilingualism, biliteracy, and learning to read: Interactions among languages and writing systems. *Scientific Studies of Reading*, *9*, 43–61. http://doi.org/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*, 525–531. http://doi.org/10.1017/S1366728909990423

Bialystok, E., Majumder, S., & Martin, M. M. (2003). Developing phonological awareness: Is there a bilingual advantage? *Applied Psycholinguistics*, *24*, 27–

44. http://doi.org/10.1017.S014271640300002X

Blom, E., Küntay, A. C., Messer, M., Verhagen, J., & Leseman, P. (2014). The benefits of being bilingual: Working memory in bilingual Turkish-Dutch children. *Journal of Experimental Child Psychology*, *128*, 105–119. http://doi.org/10.1016/j.jecp.2014.06.007

Bodde, M., Schippers, S., Klein Tank, M., & van der Linde-Meijerink, G. (2008). *TULE-Engels. Inhouden en activiteiten bij de kerndoelen van 2006 [TULE-English. Content and activities of the objectives of 2006]*. Enschede. Retrieved from http://tule.slo.nl

Bosch, L., & Sebastián-Gallés, N. (2003a). Language experience and the perception of a voicing contrast in fricatives: Infant and adult data. *Proceedings of the 15th International Conference of Phonetic Sciences*, 1987–1990.

Bosch, L., & Sebastián-Gallés, N. (2003b). Simultaneous bilingualism and the perception of a language-specific vowel contrast in the first year of life. *Language and Speech*, *2*, 217–243.

Bosma, E., Blom, E., Hoekstra, E., & Versloot, A. (2016). A longitudinal study on the gradual cognate facilitation effect in bilingual children's Frisian receptive vocabulary. *International Journal of Bilingual Education and Bilingualism. Advance Online Publication.* http://doi.org/10.1080/13670050.2016.1254152

Branum-Martin, L., Tao, S., Garnaat, S., Bunta, F., & Francis, D. J. (2012). Meta-analysis of bilingual phonological awareness: Language, age, and psycholinguistic grain size. *Journal of Educational Psychology*, *104*, 932–944. http://doi.org/10.1037/a0027755

Brenders, P., van Hell, J. G., & Dijkstra, T. (2011). Word recognition in child second language learners: Evidence from cognates and false friends. *Journal of Experimental Child Psychology*, *109*, 383–396. http://doi.org/10.1016/J.JECP.2011.03.012

Bruck, M., & Genesee, F. H. (1995). Phonological awareness in young second language learners. *Journal of Child Language*, *22*, 307–324. http://doi.org/10.1017/S0305000900009806

Buyl, A., & Housen, A. (2014). Factors, processes and outcomes of early immersion education in the Francophone community in Belgium. *International Journal of Bilingual Education and Bilingualism*, *17*, 178–196. http://doi.org/10.1080/13670050.2013.866627

Carroll, J. M., Snowling, M. J., Stevenson, J., & Hulme, C. (2003). The development of phonological awareness in preschool children. *Developmental Psychology*, *39*, 913–923. http://doi.org/10.1037/0012-1649.39.5.913

Centraal Bureau voor de Statistiek. (2003). Hoe gaat het met de kinderen? [How are the children doing?]. Retrieved August 29, 2018, from https://www.cbs.nl/nl-nl/nieuws/2003/29/hoe-gaat-het-met-de-kinderen-

Chen, X., Wu, X., & Shu, H. (2004). Phonological awareness of bilingual and monolingual Chinese children. *Journal of Educational Psychology*, *96*, 142–151. http://doi.org/10.1037/0022-0663.96.1.142

Chen, X., Xu, F., Nguyen, T.-K., Hong, G., & Wang, Y. (2010). Effects of cross-language transfer on first-language phonological awareness and literacy skills in Chinese children receiving English instruction. *Journal of Educational Psychology*, *102*, 712–728. http://doi.org/10.1037/a0018802

Cohen, C. (2016). Relating input factors and dual language proficiency in French-English bilingual children. *International Journal of Bilingual Education and Bilingualism*, *19*, 296–313. http://doi.org/10.1080/13670050.2014.982506

Council of Europe. (2018). Global scale - Table 1 (CEFR 3.3): Common Reference levels. Retrieved July 19, 2018, from https://www.coe.int/en/web/common-european-framework-reference-languages/table-1-cefr-3.3-common-reference-levels-global-scale

Dahl, A., & Vulchanova, M. (2014). Naturalistic acquisition in an early language classroom. *Frontiers in Psychology*, *5*, 329. http://doi.org/10.3389/fpsyg.2014.00329

de Bruin, A., Treccani, B., & Della Sala, S. (2015). Cognitive advantage in bilingualism: an example of publication bias? *Psychological Science*, *26*, 99–107. http://doi.org/10.1177/0956797614557866

de Graaff, R. (2015). Vroeg of laat Engels in het basisonderwijs: Wat levert het op? [Early or late English in primary education: Does it matter?]. *Levende Talen Tijdschrift*, *16*(2), 3–15.

Diamond, A. (2013). Executive Functions. *Annual Review of Psychology*, *64*(1), 135–168. http://doi.org/10.1146/annurev-psych-113011-143750

Dijkstra, T., & Van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, *5*, 175–197. http://doi.org/10.1017/S1366728902003012

Driessen, G., Krikhaar, E., de Graaff, R., Unsworth, S., Leest, B., Coppens, K., & Wierenga, J. (2016). *Startmeting pilot weetalig primair onderwijs. Startmeting schooljaar 2014/15 [Start measurement pilot bilingual primary-school education. Start measurement schoolyear 2014/15]*. Retrieved from https://www.rijksoverheid.nl/onderwerpen/basisonderwijs/documenten/rapporten/2016/02/01/startmeting-pilot-tweetalig-primair-onderwijs

Duñabeitia, J. A., Hernández, J. A., Antón, E., Macizo, P., Estévez, A., Fuentes, L. J., & Carreiras, M. (2014). The inhibitory advantage in bilingual children revisited: myth or reality? *Experimental Psychology*, *61*, 234–251.

Dunn, L. M. (1959). *Peabody Picture Vocabulary Test*. Minneapolis: American Guidance Service.

Dunn, L. M., & Dunn, D. M. (2007). *PPVT-4 Manual*. Bloomington: NCS Pearson.

Dunn, L. M., Dunn, L. M., & Schlichting, L. (2005). *Peabody Picture Vocabulary Test-III-NL* (2nd ed.). Amsterdam: Pearson Assessment and Information.

EF EPI. (2017). EF English Proficiency Index. Retrieved August 7, 2018, from https://www.ef.nl/epi/

European Council. (2002). *Presidency conclusions. Barcelona European Council 15 and 16 March 2002. SN 100/1/02 REV 1*. Barcelona. Retrieved from http://ec.europa.eu/invest-in-research/pdf/download_en/barcelona_european_council.pdf

Flege, J. E., MacKay, I. R. A., & Meador, D. (1999). Native Italian speakers' perception and production of English vowels. *The Journal of the Acoustical Society of America*, *106*, 2973–2987. http://doi.org/10.1121/1.428116

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*, 678–720.

http://doi.org/10.1080/13670050802149283
Gathercole, V. C. M., Thomas, E. M., Kennedy, I., Prys, C., Young, N., Guasch, N. V., … Jones, L. (2014). Does language dominance affect cognitive performance in bilinguals? Lifespan evidence from preschoolers through older adults on card sorting, Simon, and metalinguistic tasks. *Frontiers in Psychology*, *5*, 11. http://doi.org/10.3389/fpsyg.2014.00011
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*, *14*, 289–302. http://doi.org/10.1177/1367006910367846
Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, *1*(2), 67–81. http://doi.org/10.1017/S1366728998000133
Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology*, *25*, 515–530. http://doi.org/10.1080/20445911.2013.796377
Hazan, V., & Barrett, S. (2000). The development of phonemic categorization in children aged 6–12. *Journal of Phonetics*, *28*, 377–396. http://doi.org/10.1006/JPHO.2000.0121
Janssen, C., Segers, E., McQueen, J. M., & Verhoeven, L. (2015). Lexical specificity training effects in second language learners. *Language Learning*, *65*, 358–389. http://doi.org/10.1111/lang.12102
Janssen, C., Segers, E., McQueen, J. M., & Verhoeven, L. (2017). Transfer from implicit to explicit phonological abilities in first and second language learners. *Bilingualism: Language and Cognition*, *20*, 795–812. http://doi.org/10.1017/S1366728916000523
Jenniskens, T., Leest, B., Wolbers, M., Bruggink, M., Dood, C., & Krikhaar, E. (2017). *Zicht op vroeg vreemdetalenonderwijs [Insight into early foreign language education]*. Nijmegen.
Jenniskens, T., Leest, B., Wolbers, M., Krikhaar, E., Teunissen, C., de Graaff, R., … Koppens, K. (2018). *Evaluatie pilot tweetalig primair onderwijs. Vervolgmeting schooljaar 2016/17 [Evaluation pilot bilingual primary education. Follow-up measurement schoolyear 2016/17]*.
Kang, J. Y. (2012). Do bilingual children possess better phonological awareness? Investigation of Korean monolingual and Korean-English bilingual children. *Reading and Writing*, *25*, 411–431. http://doi.org/10.1007/s11145-010-9277-4
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*, 149–174. http://doi.org/10.1006/jmla.1994.1008
Kuo, L.-J., & Anderson, R. C. (2010). Beyond cross-language transfer: Reconceptualizing the impact of early bilingualism on phonological awareness. *Scientific Studies of Reading*, *14*, 365–385. http://doi.org/10.1080/10888431003623470
Kuppens, A. H. (2010). Incidental foreign language acquisition from media exposure. *Learning, Media and Technology*, *35*, 65–85. http://doi.org/10.1080/17439880903561876

Lobo, V. R. (2013). *Teaching L2 English at a very early age : A study of Dutch schools*. Radboud University. Retrieved from http://www.lotschool.nl

Loizou, M., & Stuart, M. (2003). Phonological awareness in monolingual and bilingual English and Greek five-year-olds. *Journal of Research in Reading*, *26*, 3–18. http://doi.org/10.1111/1467-9817.261002

Marinova-Todd, S. H., Zhao, J., & Bernhardt, M. (2010). Phonological awareness skills in the two languages of Mandarin–English bilingual children. *Clinical Linguistics & Phonetics*, *24*, 387–400. http://doi.org/10.3109/02699200903532508

McCarthy, K. M., Mahon, M., Rosen, S., & Evans, B. G. (2014). Speech perception and production by sequential bilingual children: A longitudinal study of voice onset time acquisition. *Child Development*, *85*, 1965–1980. http://doi.org/10.1111/cdev.12275

Ministerie van Onderwijs Cultuur en Wetenschap. (2005). *Bijlage: Herziene kerndoelen basisonderwijs [Attachment: revised objectives primary education]*. The Hague.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "Frontal Lobe" tasks: a latent variable analysis. *Cognitive Psychology*, *41*(1), 49–100. http://doi.org/10.1006/cogp.1999.0734

Morales, J., Calvo, A., & Bialystok, E. (2014). Working memory development in monolingual and bilingual children. *Journal of Experimental Child Psychology*, *114*(2), 187–202. http://doi.org/10.1016/j.jecp.2012.09.002.Working

Nuffic. (n.d.). *Standaard vroeg vreemdetalenonderwijs Engels [Standard early foreign language education English]*.

Nuffic. (2017). Vvto-scholen met vvto-keurmerk [early foreign language schools with foreign language certificate]. Retrieved August 5, 2018, from https://www.nuffic.nl/bestanden/documenten/vvto-scholen-met-vvto-keurmerk.pdf

Nuffic. (2018). Locaties vvto-scholen [Locations early foreign language schools]. Retrieved July 18, 2018, from https://www.nuffic.nl/primair-onderwijs/vroeg-vreemdetalenonderwijs-vvto/locaties-vvto-scholen

Nuffic. (2018). Tweetalig primair onderwijs [bilingual primary education]. Retrieved August 7, 2018, from https://www.nuffic.nl/onderwerpen/tweetalig-primair-onderwijs/

Nuffic. (2018). Wat is vroeg vreemdetalenonderwijs? [What is early foreign language education?]. Retrieved July 18, 2018, from https://www.nuffic.nl/primair-onderwijs/vroeg-vreemdetalenonderwijs-vvto/wat-is-vroeg-vreemdetalenonderwijs

Onderwijsraad. (2008). *Vreemde talen in het onderwijs [foreign languages in education]*. The Hague: OBT bv.

Paap, K. R., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology*, *66*, 232–258. http://doi.org/10.1016/j.cogpsych.2012.12.002

Paap, K. R., Johnson, H. A., & Sawi, O. (2014). Are bilingual advantages dependent upon specific tasks or specific bilingual experiences? *Journal of Cognitive*

Psychology, *26*, 615–639. http://doi.org/10.1080/20445911.2014.944914

Paap, K. R., Johnson, H. A., & Sawi, O. (2015). Bilingual advantages in executive functioning either do not exist or are restricted to very specific and undetermined circumstances. *Cortex*, *69*, 265–278. http://doi.org/10.1016/j.cortex.2015.04.014

Paap, K. R., Myuz, H. A., Anders, R. T., Bockelman, M. F., Mikulinsky, R., & Sawi, O. M. (2017). No compelling evidence for a bilingual advantage in switching or that frequent language switching reduces switch cost. *Journal of Cognitive Psychology*, *29*, 89–112. http://doi.org/10.1080/20445911.2016.1248436

Poarch, G. J., & Bialystok, E. (2015). Bilingualism as a model for multitasking. *Developmental Review*, *35*, 113–124. http://doi.org/10.1016/j.dr.2014.12.003

Poarch, G. J., & van Hell, J. G. (2012). Executive functions and inhibitory control in multilingual children: Evidence from second-language learners, bilinguals, and trilinguals. *Journal of Experimental Child Psychology*, *113*, 535–551. http://doi.org/10.1016/j.jecp.2012.06.013

Ratiu, I., & Azuma, T. (2014). Working memory capacity: Is there a bilingual advantage? *Journal of Cognitive Psychology*, *27*, 1–11. http://doi.org/10.1080/20445911.2014.976226

Reder, F., Marec-Breton, N., Gombert, J.-E., & Demont, E. (2013). Second-language learners' advantage in metalinguistic awareness: A question of languages' characteristics. *British Journal of Educational Psychology*, *83*, 686–702. http://doi.org/10.1111/bjep.12003

Rubin, H., & Turner, A. (1989). Linguistic awareness skills in grade one children in a French immersion setting. *Reading and Writing*, *1*, 73–86. http://doi.org/10.1007/BF00178839

Singleton, D., & Ryan, L. (2004). *Language acquisition: The age factor* (2nd Ed.). Clevedon: Multilingual Matters.

Sodoro, J., Allinder, R. M., & Rankin-Erickson, J. L. (2002). Assessment of phonological awareness: Review of methods and tools. *Educational Psychology Review*, *14*, 223–260. http://doi.org/10.1023/A:1016050412323

The Council of the European Union. (2008). *Council resolution of 21 November 2008 on a European strategy for multilingualism (2008/C 320/01)*. Barcelona.

Thijs, A., Trimbos, B., Tuin, D., Bodde, M., & de Graaff, R. (2011). *Engels in het basisonderwijs [English in primary education]*. Enschede.

Thomas-Sunesson, D., Hakuta, K., & Bialystok, E. (2018). Degree of bilingualism modifies executive control in Hispanic children in the USA. *International Journal of Bilingual Education and Bilingualism*, *21*, 197–206. http://doi.org/10.1080/13670050.2016.1148114

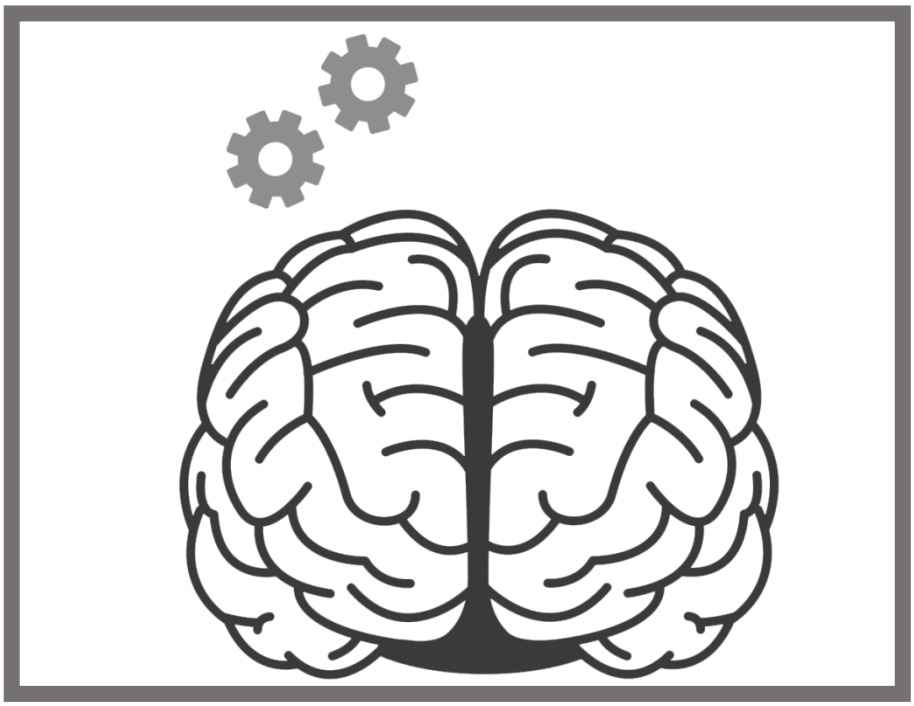
Tsukada, K., Birdsong, D., Bialystok, E., Mack, M., Sung, H., & Flege, J. (2005). A developmental study of English vowel production and perception by native Korean adults and children. *Journal of Phonetics*, *33*, 263–290. http://doi.org/10.1016/j.wocn.2004.10.002

Unsworth, S., Persson, L., Prins, T., & de Bot, K. (2015). An investigation of factors affecting early foreign language learning in the Netherlands. *Applied Linguistics*, *36*, 527–548. http://doi.org/10.1093/applin/amt052

van der Leij, A., Bekebrede, J., & Kotterink, M. (2010). Acquiring reading and

vocabulary in Dutch and English: The effect of concurrent instruction. *Reading and Writing*, *23*, 415–434. http://doi.org/10.1007/s11145-009-9207-5

Vega, C., & Fernandez, M. (2011). Errors on the WCST correlate with language proficiency scores in Spanish-English bilingual children. *Archives of Clinical Neuropsychology*, *26*, 158–164. http://doi.org/10.1093/arclin/acq097



Chapter 2: Language balance and switching ability in children acquiring English as a second language

This chapter is based on:
Goriot, C., Broersma, M., McQueen, J. M., Unsworth, S., & van Hout, R. (2018). Language balance and switching ability in children acquiring English as a second language. *Journal of Experimental Child Psychology, 173*, 168-186. doi: 10.1016/j.jecp.2018.03.019.

Chapter 2

Language balance and switching ability in children acquiring English as a second language

Abstract

This study investigated whether relative lexical proficiency in Dutch and English in child second-language learners is related to executive functioning. Participants were Dutch primary school pupils of three different age groups (4-5, 8-9, and 11-12 year-olds) who either were enrolled in an early-English schooling programme or were age-matched controls not on that early-English programme. Participants performed tasks that measured switching, inhibition and working memory. Early-English programme pupils had greater knowledge of English vocabulary, and more balanced Dutch-English lexicons. In both groups, lexical balance, a ratio measure obtained by dividing vocabulary scores in English by those in Dutch, was related to switching but not inhibition or working memory performance. These results show that for children who are learning an L2 in an instructional setting, and for whom managing two languages is not an automatized process yet, language balance may be more important than L2 proficiency in influencing the relation between childhood bilingualism and switching abilities.

2.1 Introduction

Learning two languages instead of one might affect not only language acquisition and processing but also cognitive development, especially in the domain of executive functions. It has been shown, for instance, that bilingual children outperform monolinguals on tasks requiring cognitive flexibility and inhibition skills (Barac, Moreno, & Bialystok, 2016; Poarch & van Hell, 2012). Here we examined whether the balance in first language (L1) and second language (L2) proficiency in child L2 learners is related to executive functioning.

'Executive functioning' is an umbrella term for a set of processes that together foster cognitive skills that are needed for goal-directed behaviour itself, as well as for reflection on one's own behaviour (Diamond, 2013). In general, three key processes are identified (Miyake et al., 2000): (a) inhibition, the ability to control one's attention in order to replace preliminary responses by more deliberate ones; (b) switching (or shifting), the process that fosters flexibility and adaptation to changed circumstances, including the ability to behave according to different rules or demands, and (c) (verbal and nonverbal) working memory, the capacity to hold information in mind, and to manipulate it. Executive functions develop during childhood and do not fully mature until early adulthood (Diamond, 2013). The development of executive functions may be positively influenced by different factors, including socioeconomic background, intelligence (Diamond, 2013), and, according to some, bilingualism (Adesope, Lavin, Thompson, & Ungerleider, 2010; Bialystok, 2009).

One of the most widely accepted theories about the relation between bilingualism and cognitive development is the inhibitory control model (Green, 1998). This model proposes that control of the lexico-semantic system is more demanding for bilinguals than for monolinguals, because bilinguals need to control

two languages that are simultaneously active. Monitoring the competing semantic activation between words requires goal maintenance, conflict monitoring, and, specifically for bilinguals, suppression of interference from the other language. According to this model, dealing with this linguistic competition strengthens bilinguals' linguistic control as well as behavioural-related control processes, in particular executive functions (Green, 1998; Green & Abutalebi, 2013).

Recently, however, the latter has been challenged (de Bruin, Treccani, & Della Sala, 2015; Paap & Greenberg, 2013; Paap, Johnson, & Sawi, 2015). Whereas some studies have failed to find any bilingual advantages in inhibition (Duñabeitia et al., 2014), switching (Paap et al., 2017), working memory (Ratiu & Azuma, 2014) or attentional control mechanisms (Antón et al., 2014; Duñabeitia et al., 2014), others have reported that children growing up bilingually from birth do show more developed inhibitory skills (Barac et al., 2016), conflict resolution (Poarch & van Hell, 2012), working memory (Morales, Calvo, & Bialystok, 2014), and attentional control (Poarch & Bialystok, 2015). These inconclusive findings indicate that it is still not clear which specific bilingual groups show advantages in executive functioning, when such differences manifest themselves and in which specific components. This study aimed to address all three of these issues.

2.1.1 Executive functioning in L2 learners

Gathercole et al. (2014) presented simultaneous bilingual children and adults with card-sorting and Simon tasks. Participants were either monolingual or bilingual (English dominant, Welsh dominant, or balanced in language use). For kindergartners and primary-school children, no general bilingual advantage was found. The authors suggested that for simultaneous bilinguals, language switching may be an automatic and effortless process. They theorized that this may be different for L2 learners (i.e., sequential bilinguals), because linguistic selection requires a greater level of control in this group, which in turn strengthens their executive functions.

Indeed, in contrast to Gathercole et al.'s (2014) findings for early bilinguals, findings with bilinguals who are in the process of learning an L2 indicate that there is a relation between language balance and executive functioning performance (Blom, Küntay, Messer, Verhagen, & Leseman, 2014; Thomas-Sunesson, Hakuta, & Bialystok, 2018). Thomas-Sunesson and colleagues (2018) suggest that, if we assume that managing ongoing linguistic competition between two languages results in executive functioning benefits, those who are more equally proficient in both languages, and hence have most experience in managing two languages, would show the greatest benefits. It may thus be that for L2 learners who are relatively proficient in both languages, but for whom language monitoring is not yet automatized, controlling two languages places demands on executive functioning. If this is the case, enhanced executive functioning performance should hold for these bilinguals in particular. Consequently, given that managing linguistic competition is dependent on proficiency in the two languages, language balance, rather than L2 proficiency alone, should be an important predictor of executive functioning.

2.1.2 Language balance

Several studies have investigated executive functioning performance in L2 learners or sequential bilingual children. Studies that have compared L2 learners, monolinguals, and sometimes also early bilinguals, have assumed that differences between the groups could be attributed to differences in exposure and proficiency: those who are most exposed to the L2 are likely to have greater proficiency in that language and more practice in monitoring two languages at the same time, which in turn enhances their performance on executive functioning measures.

These studies have yielded variable results, however. Carlson and Meltzoff (2008), for example, found that Spanish-English bilingual children outperformed both English monolingual children and children enrolled in immersion education on tasks involving conflicting attention. The authors suggested that by attending an immersion program for six months only, immersion pupils may not have had enough exposure to the additional language to show the same advantages as the bilingual group.

In another study (Poarch & van Hell, 2012), German-English bilinguals outperformed German monolingual children on conflict resolution in a Simon task, but German children enrolled in English immersion education did not significantly differ from either the monolingual or the bilingual group. The pupils had attended immersion education for less than two years and thus had more limited experience in controlling two languages than early bilinguals, which could, according to the authors, explain why the immersion group also had less developed control skills.

Purić, Vuksanović, & Chondrogianni (2017) compared monolingual children to a group of pupils enrolled in foreign-language education for five hours a day (high exposure group) and a group exposed to the L2 for one and a half hours a day (low exposure group). The high exposure group outperformed the other groups on working memory tasks, whereas the low exposure and monolingual groups' scores did not significantly differ. Purić et al. (2017) conclude that the high exposure group had more practice in working memory performance as a result of having to continuously monitor the vocabulary and grammar of both languages.

All of the aforementioned studies focused on L2 development, assuming that as the L2 developed, children would become more balanced in proficiency across their two languages. However, none of these studies included a measure of L1 proficiency or of language balance, despite large individual variation in L1 vocabulary scores (Carlson & Meltzoff, 2008; Poarch & van Hell, 2012). In this paper we argue that executive functioning depends on children's proficiencies in the L2 *and* the L1 and hence on language balance. Language balance accounts for the fact that two children who are equally proficient in the L2 do not have to be equally balanced bilinguals: someone with low proficiency in the L2 and the L1 is balanced, whereas someone with low proficiency in the L2 but high proficiency in the L1 is unbalanced. Assessing language balance in relation to executive functions thus requires a measure that comprises both the L1 and the L2.

Several researchers have indeed directly investigated the relation between relative L1 and L2 proficiency and executive functions. For example, Vega and Fernandez (2011) classified sequential Spanish-English bilingual children as 'balanced' or 'unbalanced' based on their scores on Spanish and English vocabulary tests. The balanced bilinguals performed significantly better than the unbalanced

bilinguals on a card-sorting task measuring switching abilities. Blom et al. (2014) presented Turkish-Dutch children with verbal and nonverbal working memory tasks. Language balance was operationalised as the difference in lexical proficiency between Dutch and Turkish, by dividing children's highest vocabulary score by their lowest vocabulary score. A significant positive association was revealed between language balance and performance on verbal working memory tasks. In a study on executive functioning performance of Spanish-English bilingual children enrolled in English monolingual or Spanish-English immersion education, Thomas-Sunesson et al. (2016) operationalised language balance as the difference between the (standardized) English and Spanish vocabulary scores. A positive relation was found between balance and both executive control and working memory, but not inhibition. It should be noted that previous research failed to find a relation between language balance and executive functioning in (young) adults (Paap et al., 2017; Paap, Johnson, & Sawi, 2014). Such a relation thus might exist only in children.

In previous studies, children acquired their L2 in (relatively) naturalistic language environments, often being immersed in the L2 for at least 50% of the time. It remains unclear whether there is a similar relation between language balance and executive functioning for children who acquire the L2 via instruction in which there is more limited input of the L2. That question is the focus of this study. It examines the relation between executive functioning and language balance in L2 learners, operationalising balance using lexical proficiency, but now examining pupils in early-English primary schools in The Netherlands where the input in English was much more limited compared to earlier studies. In previous studies, immersion pupils received much more input in English than the pupils in our study, but some pupils in our study have been exposed to the L2 for a longer time, as we also included pupils in the highest grade of primary school.

2.1.3 English education in The Netherlands

In The Netherlands, English has been an obligatory subject from the penultimate grade of primary school (grade 7, 10-11 year-olds) since 1986. In this type of education, pupils have approximately 1 hour of English instruction per week, during the final two years of primary school only. Currently, approximately 18% of the primary schools provide children with English lessons from grade 1 (i.e., kindergarten) onwards (EP-Nuffic, 2015). Schools are allowed to use English for up to 4 hours of teaching time (EP-Nuffic, 2015), but most do so for less than one hour per week (Thijs, Trimbos, Tuin, Bodde, & de Graaff, 2011).

Previous studies showed that despite the limited exposure to English, children make significant progress in their English vocabulary (Goorhuis-Brouwer & de Bot, 2010; Lobo, 2013; Unsworth, Persson, Prins, & de Bot, 2015; van der Leij, Bekebrede, & Kotterink, 2010), pronunciation (Lobo, 2013), and grammar (de Graaff, 2015; Unsworth et al., 2015). All these studies, however, tested children in only one age group. In most studies, pupils had had maximally two years of English lessons; in one study (de Graaff, 2015) participants had had either two or five to eight years of English lessons. In our study, we included pupils who had just started, were halfway, or were at the end of primary school. By doing so, we also aimed to shed more light on the effects of early-English education on the vocabulary knowledge of pupils after different amounts of English education.

The previous studies on early-English education show that there is considerable variation in the extent to which pupils have acquired the L2. Considering language balance, only two studies took pupils' L1 development into account. In both studies, early-English pupils' Dutch vocabulary development was comparable to that of same-aged peers (Goorhuis-Brouwer & de Bot, 2010; van der Leij et al., 2010). Despite that, there was individual variation in Dutch vocabulary development. Primary-school pupils' L1 vocabularies are still developing. As was hypothesized by Cummins (1979), the development of an L1 and L2 are related, and so, once again, it is important to take into account not only the development of the L2, but also that of the L1, as we did in this study.

Previous studies showed advantages for early-English pupils in English language skills. Pupils from control schools, however, are also found to have some knowledge of English (de Graaff, 2015; Goorhuis-Brouwer & de Bot, 2010; Lobo, 2013; Unsworth et al., 2015), and some individual control school pupils have greater knowledge of English than some enrolled in early-English education (de Graaff, 2015). Out-of-school exposure may have played a role in these findings. Although English is not an official language of The Netherlands, it is very present in everyday life: for example, advertisements are often in English, and movies are subtitled rather than dubbed (Kuppens, 2010). Indeed, primary-school pupils' knowledge of English is influenced by watching subtitled movies and playing English computer games (Kuppens, 2010). Given these findings, we assume that there will be considerable variability in pupils' Dutch-English language balance, both for early-English and control pupils.

2.1.4 The current study

This study investigated whether differences in lexical balance are related to executive functioning measures in children who have limited exposure to the L2. Such findings should contribute to the ongoing debate on the relation between bilingualism and cognitive development by helping to specify the conditions of bilingual experience that are most likely to stimulate these functions. We chose to focus on Dutch children who are learning English as an L2, either because of contact with English in everyday life or by means of a special educational programme. Although the first group was not completely monolingual, they were functionally monolingual because their knowledge of English was very limited and they did not actively use the language. We included children of different age groups to be able to evaluate the development of vocabulary in the L1 and L2 as well as executive functions as children get older: 4-5 year-olds (grade 1, i.e. kindergarten), 8-9 year-olds (grade 5), and 11-12 year-olds (grade 8; final grade of primary school).

Our main research question was whether individual differences in lexical balance between the L1 (Dutch) and the L2 (English) are related to individual differences in executive functions. We expected that more balanced L2 learners would perform better on executive functioning tasks. Whereas previous research mostly took into account one or two specific executive functioning processes (Blom et al., 2014; Carlson & Meltzoff, 2008; Poarch & van Hell, 2012; Purić et al., 2017; Vega & Fernandez, 2011), we included all key factors in Miyake's model (switching, inhibition, verbal and non-verbal working memory). Given that previous research showed that balance is an important factor in L2 learners' executive

functioning performance (Blom et al., 2014; Thomas-Sunesson et al., 2018), we expected that also for this group of L2 learners, there would be a relation between lexical balance and executive functions.

Lexical balance was expected to be a better predictor of executive functioning than L1 or L2 development alone because, as argued above, it is assumed that differences between L2 learners and monolinguals stem from the fact that L2 learners have to control two languages that are simultaneously active (Green, 1998; Green & Abutalebi, 2013). This becomes more demanding when both languages are mastered and the interplay between the two languages gets stronger.

Our second research question was whether children enrolled in an early-English school would differ in development from pupils who were not enrolled in such a programme. Given previous studies' results on vocabulary development (Goorhuis-Brouwer & De Bot, 2010; Unsworth et al., 2014), we expected early-English pupils to have greater knowledge of English vocabulary than pupils from control schools, and therefore to have more balanced lexicons.

Previous studies showed that a bilingual immersion programme can positively influence executive functioning development (Poarch & van Hell, 2012; Purić et al., 2017). The participants in our study were exposed to the L2 to a lesser extent than previous participants, in terms of hours per week. Some of them were, however, exposed for a longer time, since we also included 11-12 year-old children who had been exposed to the L2 for eight years. As sufficient exposure to the L2 seems a prerequisite for executive functioning advantages to show (Carlson & Meltzoff, 2008), we expected that 11-12 year-old early-English pupils would outperform their peers from control schools.

In summary, this study examined three hypotheses: 1) there is a positive relation between lexical balance and executive functioning performance; 2) early-English pupils have on average a greater vocabulary in English and are more balanced across their two lexicons, and 3) older early-English pupils (i.e., the 11-12 year-old group) have better developed executive functions than control pupils of the same age.

2.2 Method

2.2.1 Participants

Four early-English schools in The Netherlands that had at least eight years of experience with teaching English were recruited. Four control schools were matched to the early-English schools on neighbourhood (average income), area (urbanized or rural), religious denomination and educational philosophy.

We asked the head teacher of each school to select ten children in each age group to participate. Criteria for selection were that children should not have any developmental disorders, severe motor, sight or hearing impairments, and should not be exposed to another language at home. Data were collected from 241 primary-school pupils. Participants were part of one of the following age groups: 4-5 years old (1st grade), 8-9 years old (5th grade), or 11-12 years old (final grade). On average, the 11-12-year-old children in early-English education had had 640 hours of English lessons, as opposed to 80 hours for control pupils who started their English lessons in the penultimate grade (grade 7). Of the 241 participants, 37 were removed from the analyses, because they did not complete enough tasks (N = 8),

were enrolled in extracurricular language-learning activities (N = 1), were clearly not concentrating during testing (n = 1), were exposed to another language at home (n = 23), or had a diagnosis of dyslexia (n = 4). Parents gave informed consent for their child's participation. Table 2.1 provides the number (n) of participants per group, their mean (M) age and the standard deviations (SD). The exposure to English measure was based on the estimates that parents provided in a questionnaire. There was a response rate of 37.2% on the questionnaire, which is a lower completion rate than in previous research (e.g., Unsworth et al., 2014).

Table 2.1
Participants: Descriptive Statistics per Age Group

	Control schools			Early-English schools		
	4-5 year-olds	8-9 year-olds	11-12 year-olds	4-5 year-olds	8-9 year-olds	11-12 year-olds
n	38	34	26	40	38	28
Girls (*n*)	21	17	16	23	19	19
Boys (*n*)	17	17	10	17	19	9
Age (*M, SD*)	4.90 (0.27)	9.08 (0.40)	12.15 (0.53)	4.81 (0.35)	8.95 (0.41)	11.97 (0.38)
Exposure to English at home in hours per week *(M, SD)*	5.09 (6.08)	9.27 (6.35)	16.10 (14.11)	9.14 (7.97)	7.44 (6.41)	22.5 (16.43)

Note: n = number of participants; *M* = mean; *SD* = standard deviation.

2.2.2 Executive function measures

Switching. In the Dimensional Change Card Sort task (DCCS; Zelazo, 2006) children were first presented with two cards, a red boat and a blue rabbit. These cards remained visible throughout the task. Experimental cards depicted a blue boat (n = 12) or a red rabbit (n = 12). In the *pre-switch phase*, participants sorted six cards on colour. In the *post-switch phase* six cards were sorted on shape. If children responded correctly to at least five trials, they passed on to the *border phase*: children sorted cards with a black border (6) on colour, and cards without a border (6) on shape. Instructions were repeated half-way. The order of the cards was the same for all children. The *pre-switch* and *border* phases were preceded by two explanation trials. No more than two cards with a border or with the same picture were shown after each other, and cards of blue boats and red rabbits were equally distributed in each phase. The switching score was computed as the total number of correctly performed trials over all three conditions.

Inhibition. In this version of the Simon task (Hedge & Marsh, 1975; Simon & Small, 1967), children saw a blue triangle or red square on either the left or the right side of a computer screen. Instructions were counterbalanced: half of the

children were asked to press the left button on the button box when they saw a square, and the right one when they saw a triangle, and the other half were asked the opposite. Buttons were marked with a sticker in the corresponding colour. Response location was thus either congruent or incongruent with stimulus location. The trial was terminated by a response, which immediately triggered the start of a new trial. Otherwise, the trial was terminated after 2500 ms for kindergartners, or 1000 ms for older children. Stimuli were preceded by a fixation cross, presented in the middle of the screen for 500 ms. Children were told to react accurately and as fast as possible. The task started with eight practice trials, which were repeated until the child understood the task, and were followed by 240 experimental trials.

Working memory. Two subtests of the Automated Working Memory Assessment (AWMA; Alloway, Gathercole, Kirkwood, & Elliott, 2008), a PC-based test for working memory skills, were used. Nonverbal working memory was assessed by the *Odd One Out* subtest. A trial started with a set of three shapes in a three-by-three matrix being presented together in a row. Children had to indicate the odd one out, which had a different shape. The shapes then disappeared, leaving only the boxes. Children had to recall the location of the shape identified as the odd one out. The test started with four practice trials. The test phase started with a block of trials in which the location of one shape had to be remembered, building up to a block in which the location of seven shapes had to be remembered in order.

In the *Backwards Digit Recall test*, which taps into verbal working memory, children orally repeated a string of spoken digits in reverse order. After four practice trials, the test phase started with a string of two digits up to a maximum of nine. In both subtests, a block consisted of four trials. When a trial was performed incorrectly, additional trials of the same length were administered, with a maximum of two trials. The tests ended after three incorrect trials. The test-retest reliability coefficients range between .86 for the Backwards Digit Recall and .88 for the Odd One Out (Alloway et al., 2008).

2.2.3 Language measures

English and Dutch vocabulary. The Peabody Picture Vocabulary Task – Fourth edition (PPVT-4; Dunn & Dunn, 2007) and the Peabody Picture Vocabulary Task-III Dutch (Dunn, Dunn, & Schlichting, 2005) were administered to assess English and Dutch vocabulary, respectively. The PPVT is a standardized receptive vocabulary task in which children hear a word and have to choose the corresponding picture from a set of four. The PPVT-4 consists of 228 words. The PPVT-III Dutch consists of 204 words. The test-retest reliability coefficients for children between the age of 4;0 and 13;0 range between .91 and .94 for the English version (Dunn & Dunn, 2007), and from .91 to .96 for the Dutch version (Dunn et al., 2005). The score was determined by the number of correctly identified words. Raw scores were used instead of age-related norm scores, since norm scores are based on native speakers and are therefore inappropriate for L2 learners.

Balance. Lexical balance was the natural logarithm of the proportion correct on the English vocabulary test divided by the proportion correct on the Dutch vocabulary test. An outcome of 0 indicated a turning point from being more proficient in Dutch (negative value) to being more proficient in English (positive

value). A comparable calculation of language balance can be found in Blom et al. (2014).

2.2.4 Intelligence measure

Intelligence was measured with the brief version of the Wechsler Nonverbal Scale of Ability (Wechsler & Naglieri, 2008). All children performed the *Matrices* subtest: They had to complete an incomplete figural matrix by choosing the correct piece out of four or five alternatives. The total number of trials was 41. As for the subtest *Recognition* described below, testing ended when four out of five consecutive trials were incorrect.

Kindergartners performed the *Recognition* subtest in which they were presented with a geometric design for three seconds. Subsequently, they had to identify the design they had seen out of four or five designs given on a new page. The total number of trials was 21.

Older children did the *Spatial span* subtest. The examiner tapped a pattern on ten irregularly placed blocks. The task consisted of a forward and a backward phase, in which children repeated the sequence in the same order or in reverse order, respectively. Testing ended after two failed attempts at sequences of the same length (max. 32 trials).

For all subscales, raw scores were computed as the total number of correctly performed trials. The raw score for intelligence was determined by the total score of both subtests. The internal consistency of the various subtests ranges from α = .63 to α =.78 (Wechsler & Naglieri, 2008).

2.2.5 Procedure

Children were tested individually in a quiet room at school. Testing took place in two sessions of approximately 30 minutes. The first session started with the DCCS, followed by the AWMA (Odd One Out and then Backwards Digit Recall), and the PPVT-4. The second session began with the WNV (Matrix Reasoning and then Recognition or Spatial Span), followed by the PPVT-III-NL, and the Simon task. All tests were carried out following the procedure outlined in the task manuals, except for the DCCS and the Simon task for which general descriptions were used (Hedge & Marsh, 1975; Zelazo, 2006). The two sessions were separated by at least one and no more than 28 days (M = 5.87, SD = 4.63), except for one participant who performed the two sessions on the same day, but separated from one another by more than four hours. Responses were registered on a laptop (AWMA and Simon task) and/or noted down by the experimenter (all tasks except for the Simon task).

2.2.6 Analyses

Given the data's hierarchical (multilevel) structure, with child level gathered in the context of a specific school, the Linear Mixed Models function in SPSS 22.0 was used for the analysis. In all models, School was entered as random factor.

First, we investigated whether there were relations between each of the language measures and Intelligence with Age Group (4-5, 8-9, and 11-12 years), Type of Education (early-English and control), and the interaction between Age Group and Type of Education. Thereafter, we investigated which variables could

explain executive functioning development. We started with a base model with Age Group, Type of Education, and their interaction. Subsequently, we compared this model to one in which we added Intelligence and the language measure Balance. The aim was to investigate whether individual differences in these latter two variables influenced executive functioning outcomes. If Balance was included in the model, we checked whether there were interaction effects of this measure with Type of Education and/or Age Group. The overall aim was to find the most parsimonious model: a model that included the variables that could explain the variance in the executive functioning measure with the least degrees of freedom. We performed all of the aforementioned steps for each executive function measure separately (Switching, Inhibition, Verbal and Nonverbal Working Memory). To determine if a newer model was a better fit than a previous one, the deviance score (-2*Log Likelihood ratio; *-2LL*) was used as a goodness of fit measure (Heck, Thomas, & Tabata, 2014). All models were tested against .05 significance levels. The direction and size of the effect parameters were checked only after the goodness of fit had been checked. We used the MuMIn package in R (version 3.4.1) to determine the R^2 for each model. We report both the marginal and conditional R^2 (Nakagawa & Schielzeth, 2013). The marginal R^2 (R^2_m) shows the variance that is explained by fixed factors, whereas the conditional R^2 (R^2_c) is concerned with the variance explained by both fixed and random factors. We follow the general rule of thumb that R^2 values of .1, .3, and .5 are respectively small, moderate, and large effects.

2.3 Results

2.3.1 General results

The results for all measures are presented in Table 2.2.

Table 2.2
Raw Scores on Executive Functioning Measures, the Intelligence Measure, and Vocabulary Measures

			4-5 year-olds		8-9 year-olds		11-12 year-olds	
			Control	Early-English	Control	Early-English	Control	Early-English
Switching (accuracy, max. 24)		*N*	29	30	33	37	26	28
		M	18.45	18.10	19.70	19.32	20.77	21.25
		(SD)	(1.76)	(1.47)	(3.03)	(3.09)	(2.86)	(2.29)
Inhibition	Congruent (accuracy, max. 120)	*N*	36	35	34	37	25	27
		M	105.39	105.97	108.06	107.51	113.56	113.41
		(SD)	(11.15)	(8.71)	(5.78)	(6.74)	(4.08)	(4.30)
	Incongruent (accuracy, max. 120)	*M*	99.56	90.54	99.97	99.43	105.76	105.00
		(SD)	(13.77)	(13.77)	(9.20)	(9.26)	(9.26)	(8.20)
	Simon Effect (RT)	*M*	64.00	72.03	39.81	45.51	36.51	39.77
		(SD)	(41.04)	(60.93)	(18.71)	(24.50)	(20.03)	(17.35)
Verbal working memory (accuracy, max. 162)		*N*	37	39	32	37	26	28
		M	4.08	3.46	12.38	11.73	16.38	15.07
		(SD)	(3.44)	(2.61)	(3.70)	(3.44)	(4.22)	(3.64)
Nonverbal working memory (accuracy, max. 36)		*N*	37	39	32	37	26	28
		M	11.62	10.28	20.00	19. 89	25.00	23.93
		(SD)	(3.56)	(2.80)	(4.02)	(3.96)	(4.72)	(4.40)
Intelligence (raw score, max. 62 for 4-5 year-olds, max. 73 for older pupils)		*N*	38	40	34	38	26	28
		M	21.39	21.63	32.26	30.66	36.96	36.57
		(SD)	(5.24)	(5.83)	(4.26)	(4.04)	(5.36)	(4.61)
Dutch vocabulary (raw score, max. 204)		*N*	38	40	34	38	26	28
		M	72.05	72.58	110.68	111.13	133.65	134.86
		(SD)	(10.28)	(12.21)	(7.53)	(7.56)	(10.14)	(9.80)
English vocabulary (raw score, max. 228)		*N*	38	40	34	38	26	28
		M	13.39	18.18	55.94	56.13	93.77	111.50
		(SD)	(8.33)	(10.47)	(26.67)	(16.77)	(32.00)	(27.14)
Balance		*N*	38	40	34	38	26	28
		M	-1.97	-1.65	-0.89	-0.84	-0.52	-0.33
		(SD)	(0.61)	(0.56)	(0.44)	(0.32)	(0.29)	(0.22)

Note: A higher score means poorer performance for Inhibition, and better performance for all other measures. *n* = number of participants; *M* = mean; *SD* = standard deviation.

Table 2.2 reveals that pupils in the highest groups obtained the best scores on all measures. Descriptively, there is little difference in scores between early-English and control pupils, except that early-English pupils seem to have higher scores on the English vocabulary test and more balanced lexicons than pupils from control schools, although differences are very small for 8-9 year-olds. Note that a higher score on the English than on the Dutch vocabulary test, does not necessarily indicate a larger vocabulary in English, due to differences in the number of items in the two tests. When comparing norm scores for both vocabulary tests, all pupils except one had a higher norm score for Dutch than for English.

A mixed models analysis on Age with Schools as random variable and Age Group, Type of Education, and the interaction between Age Group and Type of Education revealed that only Age Group was significantly related to Age ($p < .001$). There were thus no differences in age between the groups of pupils from the two types of schools. An ANOVA with Age Group and Type of Education as independent variables and Out-of-school exposure to English as dependent variable showed a significant relation between Age Group and Out-of-school exposure ($p < .001$), indicating that the pupils in the higher grades were more exposed to English than the pupils in the lower grades. There were again no differences between the pupils from the two types of schools.

2.3.2 Correlations between measures

First, we examined the partial correlations between the measures, controlling for school-specific effects. Table 2.3 shows that there is a very strong, positive correlation between Balance, English Vocabulary and Dutch Vocabulary. For the theoretical reasons outlined in the Introduction, Balance rather than English Vocabulary will be used as the primary measure in the following analyses.

Figures 2.1 and 2.2 give a graphical overview of the relations between the raw scores on the English and Dutch vocabulary tests. As shown in Figure 2.1, the difference between Dutch and English vocabulary becomes smaller for children in the higher age groups. Figure 2.2 shows the relation between developments of both vocabularies showing that the development of Dutch vocabulary is less prone to individual variation than that of English. The benefit of Balance over the other two vocabulary measures is that it captures the relation between English and Dutch vocabulary. That is, it takes into account the differences between L1 and L2 proficiency shown in these figures.

Table 2.3
Partial Correlations between Age Group, Type of Education, Intelligence, Language Measures, and Executive Functions

		1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
1.	Age Group	1									
2.	Type of Education	-.100	1								
3.	Intelligence	.770**	-.080	1							
4.	Balance	.764**	.150	.620**	1						
5.	English Vocabulary	.817**	.070	.686**	.878**	1					
6.	Dutch Vocabulary	.929**	-.011	.807**	.750**	.832**	1				
7.	Switching	.359**	-.077	.346**	.336**	.401**	.390**	1			
8.	Inhibition	-.377**	.086	-.267**	-.218**	-.303**	-.341**	-.060	1		
9.	Nonverbal Working Memory	.798**	-.080	.716**	.624**	.680**	.796**	.387**	-.342**	1	
10.	Verbal Working Memory	.790**	-.074	.765**	.626**	.647**	.804**	.355**	-.263**	.787**	1

*$p < .05$; **$p < 0.01$ (2-tailed).

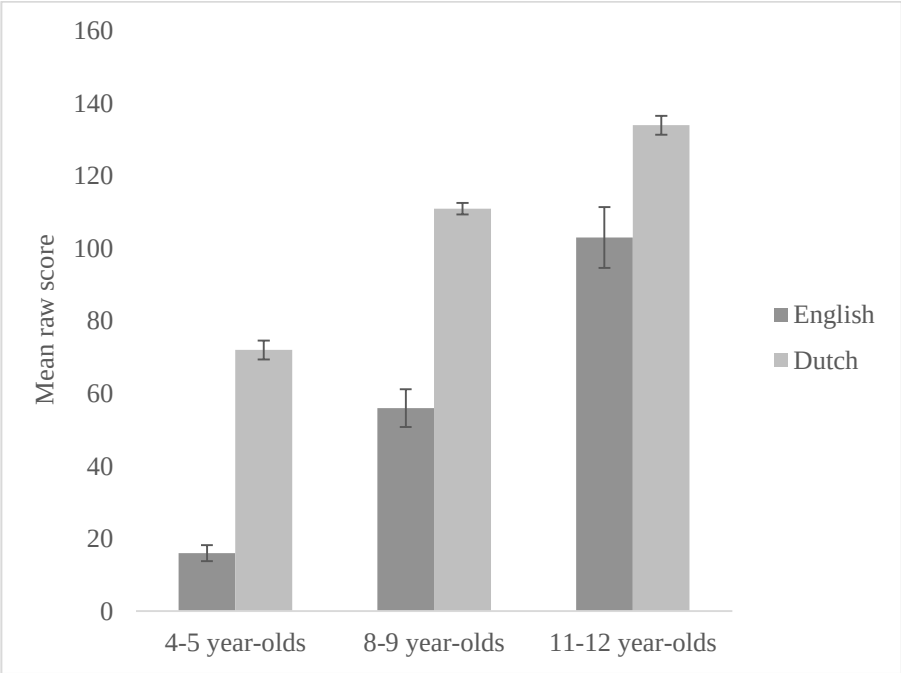


Figure 2.1. Average scores for English and Dutch vocabulary for the different age groups.

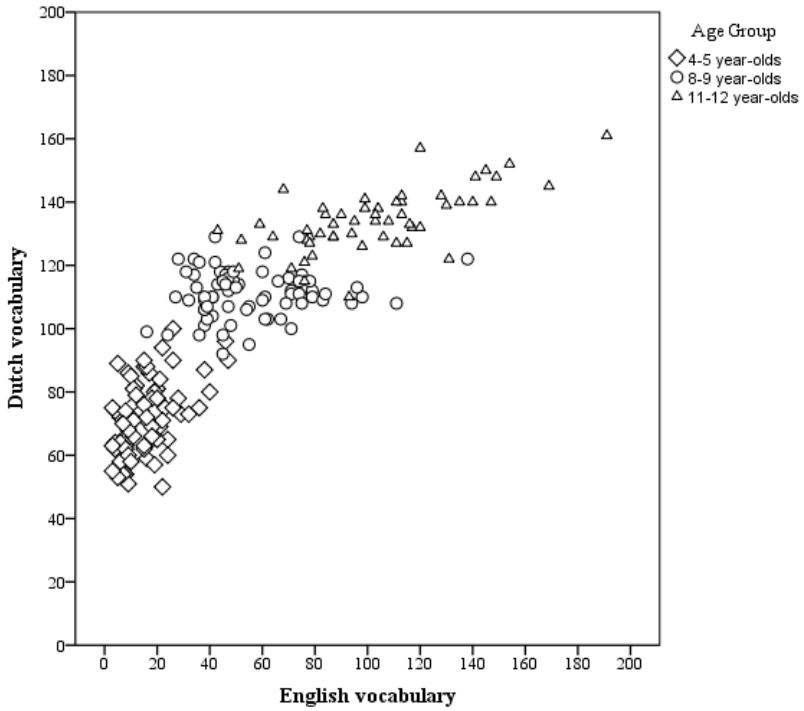


Figure 2.2. Relationship between Dutch and English vocabulary for the different age groups.

All three languages measures correlated with the executive functioning measures (cf. Table 2.3), with inhibition showing the weakest correlations of the four in general. Note that the correlations for inhibition are negative because lower scores indicate better inhibitory control.

2.3.3 Intelligence and Language Measures

First, we used a Linear Mixed Model with Schools as random factor, and Age Group, Type of Education and the interaction between the two variables as fixed factors to investigate whether there were any significant differences between the early-English and control pupils in intelligence, or any of the language measures. The results are shown in Table 2.4.

Table 2.4.
F-values of Fixed Effects for Intelligence and Language Measures

		Intelligence	Dutch Vocabulary	English Vocabulary	Balance
Intercept		2130.04***	17630.22***	1477.99***	1064.33***
Age Group		173.37***	683.22***	276.88***	171.30***
Type of Education		.26	.24	6.24(*)	8.65**
Age Group*Type of Education		.67	.03	2.82(*)	1.61
Variance components	Schools	2.40	1.21	0.73	0.00
	Residual	22.56	94.39	434.80	0.20
R^2_m		.616	.871	.737	.636
R^2_c		.653	.872	.737	.636

*(*) $p < .1$; ** $p < .01$; *** $p < .001$*

For all measures, Age Group was a significant predictor, with children in the higher age groups obtaining significantly higher scores on the intelligence tasks and language measures. For Balance, Type of Education was a significant predictor as well: children from early-English schools had on average more balanced lexicons than children from control schools. Finally, there was a marginally significant effect of Type of Education ($p = .056$), and a marginally significant interaction between Age Group and Type of Education ($p = .062$) on English Vocabulary. The model parameter estimates show that there was a significant effect of Type of Education (Beta = -17.69; *SD* = 5.72; $p = .004$ for control schools compared to early-English schools) and a significant interaction effect for 8-9 year-olds and control schools (Beta = 17.48; *SD* = 7.52; $p = .021$). This shows that the relation between Type of Education and English Vocabulary is not the same for all age groups (see Figure 2.3). Differences between control and early-English pupils in English vocabulary were found only in the highest grade of primary school.

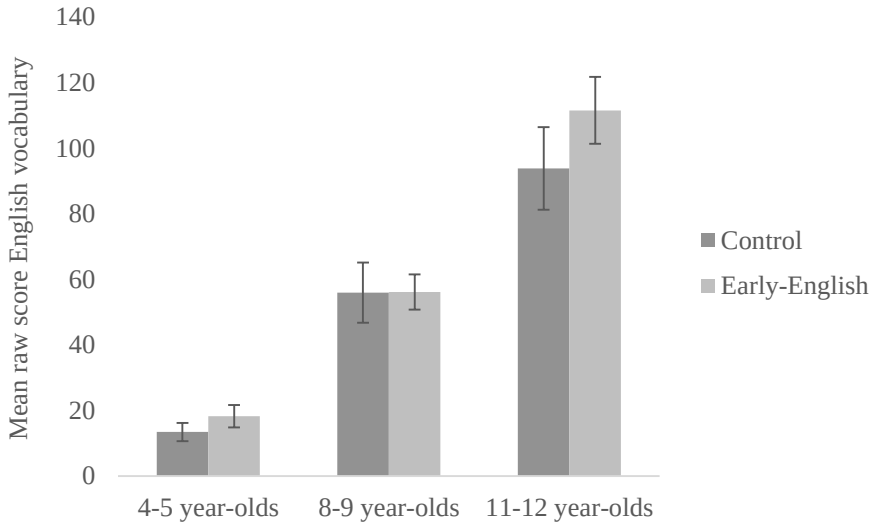


Figure 2.3. Mean scores on the English vocabulary test split out by age group and type of education.

2.3.4 Executive functions

2.3.4.1 Switching

We assume that all children who were not able to sort the cards correctly after the sorting feature in the DCCS had changed from colour to shape did not understand the task (nineteen 4-5 year-olds, two 8-9 year-olds; ten early-English, eleven control pupils). Their scores on the DCCS were therefore excluded from the analyses[1].

We started with the basic model in which Age Group, Type of Education and the interaction between the two variables were included as fixed factors. This model revealed that only Age Group was significantly related to switching outcomes (*-2LL* = 850.03; *df* = 13; R^2_m = .159; R^2_c = .186).

Subsequently, we ran the model adding either Intelligence or Balance. The model with Intelligence was not significantly different from the basic model (*-2LL* = 851.96, *df* = 14; R^2_m = .171; R^2_c = .186). Adding Balance as a fixed effect to the basic model resulted in a significant improvement of fit (*-2LL* = 844.92, *df* = 14; R^2_m = .181; R^2_c = .215). Therefore, we proceeded with Balance rather than Intelligence in the following analyses. We checked if there were any significant interactions between Balance and Age Group or Type of Education. Although the model improvement was just significant (*-2LL* = 823.58, *df* = 25; R^2_m = .212; R^2_c = .236), the *p*-values of the interactions were all $p > 0.05$. We therefore continued with the model without the interactions.

In the model with Balance as well as in the basic model, Type of Education showed no significant relation with switching scores. We therefore removed Type of

[1] Including these children in the analyses did not change the pattern of results.

Education and the interaction from the model (*-2LL* = 849.88, *df* =6; R^2_m = .177; R^2_c = .208). This latter model was not significantly different from the basic model or from the model with Type of Education and Balance both included, but the model with Age Group and Balance as predictors did have considerably fewer degrees of freedom, and was thus considered the most parsimonious model. In this model, Balance showed a significant and positive relation with Switching. Age Group reached borderline significance, but removing this factor from the model resulted in a significant deterioration in goodness of fit (*-2LL* = 857.09, *df* = 3; R^2_m = .151; R^2_c = .184), which is why it was retained.

Continuing with the model that included Age Group and Balance, we investigated whether Dutch Vocabulary, English Vocabulary, or scores on both vocabulary tests could replace Balance. Replacing Balance by scores on both vocabulary tests (*-2LL* = 856.59, *df* = 7; R^2_m = .207; R^2_c = .223), English only (*-2LL* = 853.45, *df* = 6; R^2_m = .196; R^2_c = .212) or Dutch Vocabulary only (*-2LL* = 853.28, *df* = 6; R^2_m = .188; R^2_c = .207) resulted in a significant deterioration in model fit. For both models, removing Age Group did not result in a better model fit (*-2LL* = 855.34, *df* = 3; R^2_m = .196; R^2_c = .210, and *-2LL* = 857.48, *df* = 3; R^2_m = .188; R^2_c = .207, for English and Dutch Vocabulary, respectively).

In summary, switching scores seem to be best predicted in a model that includes both Age Group and Balance. In this model, Balance showed a positive relation with switching scores: children with more balanced Dutch and English lexicons showed better switching abilities. Age Group also showed a positive relation with switching scores: children in the higher age groups scored better on the switching task than children in the lower age groups. Table 2.5 shows the model parameter estimations for the variables (with the standard errors in parentheses). For Age Group, 11-12 year-olds are the reference category.

Table 2.5
Model Parameter Estimations for Executive Functioning Measures (SE in parentheses)

		Switching	Inhibition	Verbal WM	Nonverbal WM
Intercept		21.34 (.41)***	-	-1.32 (3.18)	17.56 (2.08)***
4-5 year-olds		-1.53 (.75)*	-	-7.35 (-0.90)***	-10.71 (1.07)***
8-9 year-olds		-1.15 (.49)*	-	-2.05 (.62)**	-3.61 (.74)***
Balance		.92 (.43)*	-	-	-
Intelligence		-	-	.30 (.046)***	.19 (.05)***
Variance components	Schools	0.23	-	0.252	0.366
	Residual	6.04	-	9.75	13.832

*p < .05; **p < .01; ***p < .001

2.3.4.2 Inhibition

Two children were excluded from the analyses: One child did not complete the Simon task due to technical errors with the equipment and one child, according to the experimenter's assessment, was not concentrating during the task. All trials to which children gave no response, or an incorrect or anticipatory response (RT < 200 ms) were excluded from the analyses (4.3% removed). Thereafter, all responses with RTs above or below 2.5 *SD*s of the participant's mean were removed (1.6%), comparable to procedures used in previous studies (Davidson, Amso, Anderson, & Diamond, 2006; Duñabeitia et al., 2014). For eight participants, less than 60 trials (1 block in the Simon task) remained. Their scores were left out of the analyses. The Simon effect for the remaining 194 children was calculated as the difference between RTs on incongruent trials minus the RTs on congruent trials. Table 2.3 shows the correlations between Inhibition scores and Age Group, Type of Education, languages measures and Intelligence. All measures except for Type of Education correlated negatively with Inhibition: in general, the children in the higher age groups, those with more balanced lexicons, a larger vocabulary in Dutch, a larger vocabulary in English, and those obtaining higher scores on the intelligence tasks, showed a smaller Simon effect. Except for the correlation with Balance, these results converge with previous findings with monolinguals and highly proficient bilinguals (Davidson et al., 2006; Diamond, 2013; Tse & Altarriba, 2014).

We started again with the basic model, first adding Age Group, Type of Education, and the interaction as fixed factors to the model (*-2LL* = 1895.38, *df* = 13; R^2_m = .128; R^2_c = .142). Only Age Group turned out to be a significant predictor ($p < .001$) of inhibition scores. Thereafter, we added Intelligence to the model. The model fit was not significantly different from the previous model (*-2LL* = 1892.00, *df* = 14; R^2_m = .140; R^2_c = .152) and Intelligence was not significant. We removed Intelligence from the model and added Balance. This resulted in an improvement in model fit compared to the basic model (*-2LL* = 1886.21, *df* = 14; R^2_m = .145; R^2_c = .150). There was a significant and positive effect of Balance ($p = .049$): children with more balanced lexicons showed larger Simon effects (indicating less inhibition). This result is difficult to interpret.[2] Figure 2.4 shows that it is probably due to the positive relation between Balance and Inhibition in 4-5 year-olds. The cause might be that the 4-5 year-olds had more time to respond. From Table 2.2 it is clear that the average Simon Effect is substantially larger for 4-5 year-olds than for 8-9 and 11-12 year-olds. We therefore decided to leave the 4-5 year-olds out of the analyses. The basic model (*-2LL* = 1068.62, *df* = 4; R^2_m = .020; R^2_c = .105) was significantly different from the model with Balance (*-2LL* = 1063.27, *df* = 5; R^2_m = .021; R^2_c = .104), but Balance was no longer significant. Balance was first replaced by English Vocabulary (*-2LL* = 1070.59, *df* = 11; R^2_m = .032; R^2_c = .107) and then by Dutch Vocabulary (*-2LL* = 1067.88, *df* = 11; R^2_m = .037; R^2_c = .099), but both language measures showed a non-significant relation with Inhibition ($p > .05$). The analysis was stopped at this point as there was no optimal model.

[2] A negative effect was expected, as children with more balanced lexicons were predicted to show a smaller Simon effect.

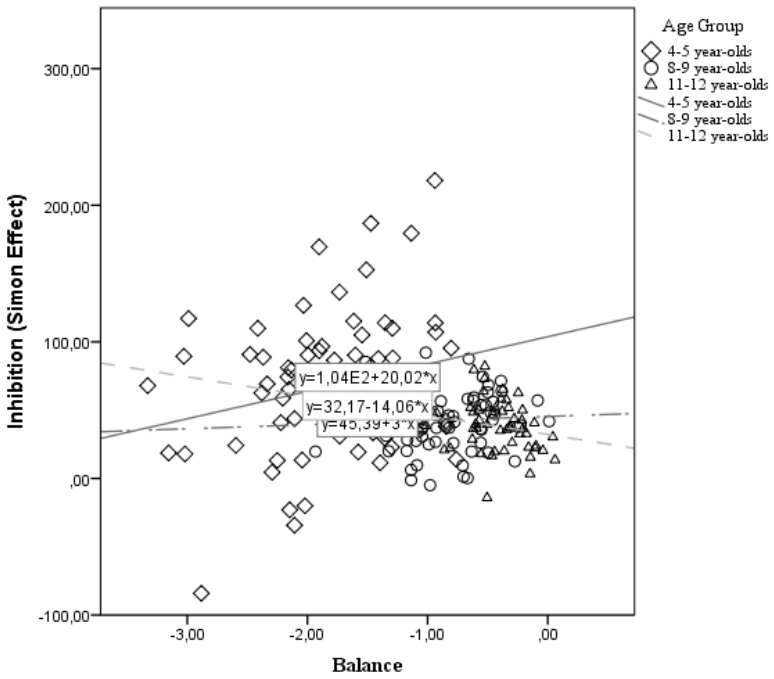


Figure 2.4. The relation between balance and inhibition split out per age group.

2.3.4.3 Working memory

Five children did not complete the working memory measures due to technical errors. Their scores were excluded from the analyses. Table 2.3 shows the correlations between working memory measures and Age Group, Type of Education, language measures, and Intelligence. Except for Type of Education, working memory measures are highly and positively correlated with all measures.

2.3.4.4 Verbal Working memory

We started with the same basic model again (*-2LL* = 1039.68; *df* =13; R^2_m = .676; R^2_c = .681). Thereafter, we checked if the individual measures Intelligence or Balance could improve the basic model. Adding Intelligence significantly improved model fit (*-2LL* = 1015.55; *df* = 14; R^2_m = .733; R^2_c = .738). Including Balance in the basic model did not result in a better *-2LL* fit (*-2LL* = 1047.41; *df* = 14; R^2_m = .677; R^2_c = .683). We therefore continued with the model with Intelligence as the only individual-differences variable. Adding interactions between Intelligence, Age Group and Type of Education did not improve the model (*-2LL* = 1015.6; *df* = 25; ; R^2_m = .743; R^2_c = .745). In the basic model, Type of Education was not significant. Removing this factor from the model containing Intelligence resulted in a model that did not differ significantly from the previous one, but had fewer degrees of freedom (*-2LL* = 1021.83, *df* = 6; R^2_m = .732; R^2_c = .739), and therefore this was considered to be a more parsimonious model. Including English Vocabulary resulted in

significantly decreased model fit (*-2LL* = 1028.84, *df* = 7; R^2_m = .732; R^2_c = .738). Although including Dutch Vocabulary in the model revealed a significant relation between Dutch Vocabulary and Verbal Working Memory (*p* = .026), the model fit did not improve (*-2LL* = 1022.44, *df* = 7; R^2_m = .738; R^2_c = .745). Therefore, the model with Age Group and Intelligence was considered the best model (see Table 2.5). In summary, Verbal Working Memory was best predicted by Age Group and Intelligence: older children and children with higher scores on the intelligence test performed, in general, better on the verbal working memory task.

2.3.4.5 Nonverbal Working memory

Again, we started with the basic model (*-2LL* = 1090.28; *df* = 13; R^2_m = .681; R^2_c = .692). Adding Intelligence resulted in a significantly better model fit (*-2LL* = 1082.00, *df* = 14; R^2_m = .700; R^2_c = .706). Adding Balance to the basic model did not improve model fit (*-2LL* of 1088.28; *df* = 14; R^2_m = .681; R^2_c = .695). As the *-2LL* was not significantly better than in the basic model, and Balance was not significant, we rejected this model. We continued with the model including Age Group, Type of Education and Intelligence. Adding the interactions between Intelligence, Age Group, and Type of Education did not improve the model (*-2LL* = 1088.8; *df* = 25; ; R^2_m = .698; R^2_c = .700). Type of Education showed no significant relation with Nonverbal Working Memory. Removing Type of Education did not improve model fit compared to the basic model plus Intelligence, but this model had fewer degrees of freedom (*-2LL* = 1090.10, *df* = 6; R^2_m = .670; R^2_c = .707), and was therefore considered the most parsimonious model. Both Age Group and Intelligence showed a significant relation with Nonverbal Working Memory. Neither including Dutch (*-2LL* = 1093.03, *df* = 7; R^2_m = .702; R^2_c = .710) nor English Vocabulary (*-2LL* = 1096.55, *df* = 7; R^2_m = .699; R^2_c = .708) resulted in a better model fit. In conclusion, Nonverbal Working Memory performance is best predicted by Age Group and Intelligence, with older children and children with higher intelligence scores performing better on the nonverbal working memory assessment (see Table 2.5).

2.4 Discussion and conclusion

This study investigated the relation between lexical balance and executive functioning in Dutch primary-school pupils learning English as an L2. Pupils, who were 4-5 years old (grade 1), 8-9 years old (grade 5) or 11-12 years old (final grade of primary school), were enrolled in an early-English educational programme or not, and both groups were exposed to English in everyday life, for example via media. Children performed executive functioning tasks, and Dutch and English vocabulary tasks. We investigated whether individual differences in Dutch-English lexical balance were related to differences in executive functioning, and whether early-English children performed differently on the tasks than children from control schools.

We hypothesized that there would be a positive relation between lexical balance and executive functioning performance, because lexical competition would be more demanding for L2 learners with more balanced language proficiencies (Blom et al., 2014; Vega & Fernandez, 2011). This hypothesis was confirmed, but only for switching. Our results are in line with previous findings showing that

children being more balanced in language proficiency show advantages in executive functioning (Blom et al., 2014; Thomas-Sunesson et al., 2018), and in particular, with the results of Vega and Fernandez (2011), who found that child bilinguals who are balanced in language proficiency perform better than less balanced bilinguals in switching, but not in inhibition. Our findings demonstrate for the first time that even for children who are exposed to the L2 in an instructed rather than naturalistic setting, and who in addition have minimal input in the L2, lexical balance is related to executive functioning performance, and to switching in particular. Our results on the other hand contradict the findings of previous studies with children and (young) adults (Gathercole et al., 2014; Paap et al., 2017, 2014) that found no relation between balanced bilingualism and switching RTs in card sorting tasks. The participants in those studies, unlike those in our study, were either early bilinguals or bilinguals who started L2 learning many years previously. A relation between L1/L2 proficiency and switching may only exist for individuals, like our participants, who are in the process of learning an L2. Such a relation may fade when managing two languages becomes automatized.

All tasks were chosen because they are considered good measures of executive functions (Diamond, 2013), and because they were suitable for use with 4-12-year-old children. For switching, the average performance was high: 76.1% in the youngest, and 87.6% in the oldest group. Our analyses revealed an effect of Age Group for all measures, thereby indicating that the tasks were suitable for measuring developmental differences in executive functioning. Despite careful selection of the tasks, we did not find a relation between either inhibition or working memory and any of the language measures. One possible reason for this is our measure of inhibition: RTs in the Simon task. Vega and Fernandez (2011) have suggested that only older participants show a bilingual advantage on a timed inhibition task, because in young children brain processes needed for optimal performance on such a task are not yet mature. Two studies provide evidence for this. Mohades et al. (2014) found that 8-11-year-old bilingual and L2 learners showed over-recruitment of brain areas such as the bilateral cingulate cortex and larger Simon effects than monolinguals, whereas Yow and Li (2015) found a relation between lexical balance and inhibition in adult participants. In line with those studies, we found a trend for the expected relation between language balance and inhibition for pupils only in the highest age group (r = -.205, p = .146; i.e. more balanced pupils show a smaller Simon effect, indicating better inhibition), but not for younger children (r = .052, p = .665 for 8-9 year-olds; r = .234, p = .049 for 4-5 year-olds). Furthermore, following the standard procedure, participants were instructed to respond accurately and as fast as possible in the Simon task. Young children tend to rush responses on timed tasks, thereby making errors (Diamond, 2013). Altogether, it may be that a relation between lexical balance and inhibition does exist, but is only observable in younger participants when using a non-timed task, like the Day-Night test (Gerstadt, Hong, & Diamond, 1994).

Our results also tie in with previous research (Blom et al., 2014) showing no relation between lexical balance and nonverbal working memory. However, contrary to Blom et al. (2014), we did not find any relation between lexical balance and verbal working memory. These diverging results may be attributable to differences in language balance in both samples. Blom et al.'s sample was quite

balanced: only two children showed a difference between Dutch and Turkish vocabulary of 10 or more points on scales ranging from 0 to 30 (Turkish) and 0 to 45 (Dutch) (Blom et al., 2014), whereas in our sample, the average difference in the scores on the Dutch and English vocabulary test was 56.47 points (*SD* = 11.12) on scales of 0 to 228 (English) and 0 to 204 (Dutch). It is worth noting that previous research has not consistently observed a bilingual advantage in working memory (Barac, Bialystok, Castro, & Sanchez, 2014). If concurrently managing two languages indeed requires greater working memory demands (Adesope et al., 2010), such a relation might exist only in more advanced L2 learners.

We expected that lexical balance would be a better predictor of executive functioning than L1 or L2 development alone, since lexical competition should be more effortful once L2 learners are more proficient in two languages (Blom et al., 2014; Vega & Fernandez, 2011). Our results show that, indeed, lexical balance rather than L1 or L2 vocabulary knowledge best explained differences in switching. Our results thereby suggest that rather than L1 or L2 differences, differences in lexical balance, even if they are small, can explain variation in switching outcomes.

Our results confirmed our second hypothesis, namely that pupils enrolled in early-English programmes would have larger English vocabularies and more balanced Dutch-English vocabularies than pupils from control schools. This suggests that pupils in early-English schools expand their knowledge of English at greater speed than their peers at control schools. When pupils start learning English, their level of English is much lower than their (already more greatly developed) Dutch. They can make larger improvements in English, and in theory, eventually English may approximate the level of Dutch.

Previous research investigating pupils' knowledge of English at either the start or end of primary school (de Graaff, 2015; Goorhuis-Brouwer & de Bot, 2010; Lobo, 2013; Unsworth et al., 2015), showed that pupils enrolled in an early-English programme outperform children from control schools. We extended previous studies by also including pupils who were halfway through their primary-school career. However, we only found a significant difference in English vocabulary between the two types of schools for the oldest pupils. There are many factors that could influence English vocabulary size, including the amount of classroom exposure to English, the teachers' English proficiency (Unsworth et al., 2015), and out-of-school exposure to English (de Graaff, 2015), and these may also have an effect on the development of executive functions. All early-English schools reported teaching English for at least 60 minutes per week, but personal communication with teachers revealed that they sometimes switched to Dutch during those lessons, so the actual time devoted to English may be less. Even though the available data from the questionnaire did not show differences in out-of-school exposure between early-English and control pupils, differences may actually exist since, unfortunately, data for the majority of the pupils (62.8%) are unavailable.

Our third hypothesis was that the group of 11-12-year-old early-English pupils would outperform control pupils of the same age on executive functioning tasks. Previous research (Purić et al., 2017) found advantages in working memory for pupils enrolled in a foreign-language learning programme for five hours a day, but the group that received only 1.5 hours of instruction per day scored similarly to the monolingual group. Our participants' exposure to the L2 was more limited (in

terms of hours per week), although overall length of exposure was longer (8 years for the oldest groups). No significant differences in executive functioning performance were found between the early-English and the control pupils, in any of the age groups. Our results show that devoting 15% of the teaching time or less to foreign-language learning does not foster advantages in executive functioning, or at least if it does, these were not strong enough to be detected using the tasks employed in the present study. By including children from different age groups we investigated linguistic and cognitive development over time. We made sure that children in the highest age groups had been enrolled in an early-English programme since the start of primary school, but several factors may have influenced their exposure to English over that period: the programme may have changed over time, and children were educated by different teachers who likely differed in proficiency. Future research should follow children over time, thereby taking into account factors that influence the development of L1 and L2 vocabulary, and probably also of executive functions in L2 learners. We found a relation between lexical balance and switching, arguing that those who are more balanced in their lexicons show better switching abilities. It could equally well be the case that children who have better switching abilities are better in mastering an L2. Again, a longitudinal study could allow for stronger claims about the causality of this relation. In addition, future research with a similar population but with higher exposure to the L2 could investigate whether the relation between language balance and executive functioning also holds for this group. Despite these limitations, our study with its cross-sectional design is a first and important exploration of L2 development in children enrolled in early-English programmes.

Research on bilingual advantages in executive functioning has been highly controversial in recent years. Some authors claim that there is no convincing evidence of such advantages or that they may appear only under specific conditions (de Bruin et al., 2015; Paap & Greenberg, 2013; Paap et al., 2015). It has been suggested that studies should include at least two measures for each process of executive functioning, to show that bilingual advantages are not task specific (Paap et al., 2015). Like in most studies with primary-school pupils and teachers, time constraints made it impossible to include more tasks. Future research could include multiple tasks, especially for switching, to reveal whether the relation between lexical balance and switching can be generalized to other tasks.

In conclusion, the results of our study show that being exposed to an L2, even for a limited amount of time, is beneficial for pupils' vocabulary development in the L2, and consequently for their balance in lexical proficiency. In turn, language balance is related to switching abilities, but not to inhibition or working memory. We did not observe an overall, global positive effect of bilingualism on executive functions. Our findings contribute to the ongoing debate about executive functioning development in (emerging) bilingual children and in addition, we also show that language balance may be a more important predictor of executive functioning than L2 proficiency alone. These findings support the view that specific relations between bilingualism and executive functions may be determined by various variables such as age, type of bilingual setting, and stage of acquisition. Our study also supports the use of a relative proficiency measure that balances L1 and L2 vocabulary knowledge, because it reflects the potential competition between the two

languages involved. This measure seems particularly relevant when managing two languages is not automatized yet. It shows that, for children at that stage of development who are learning their L2 in an instructional setting, language balance is associated with the cognitive flexibility to switch between tasks.

2.5 References

Adesope, O. O., Lavin, T., Thompson, T., & Ungerleider, C. (2010). A systematic review and meta-analysis of the cognitive correlates of bilingualism. *Review of Educational Research*, *80*, 207–245. http://doi.org/10.3102/0034654310368803

Alloway, T. P., Gathercole, S. E., Kirkwood, H., & Elliott, J. (2008). Evaluating the validity of the Automated Working Memory Assessment. *Educational Psychology*, *28*, 725–734. http://doi.org/10.1080/01443410802243828

Antón, E., Duñabeitia, J. A., Estévez, A., Hernández, J. A., Castillo, A., Fuentes, L. J., … Carreiras, M. (2014). Is there a bilingual advantage in the ANT task? Evidence from children. *Frontiers in Psychology*, *5*, 398. http://doi.org/10.3389/fpsyg.2014.00398

Barac, R., Bialystok, E., Castro, D. C., & Sanchez, M. (2014). The cognitive development of young dual language learners: A critical review. *Early Childhood Research Quarterly*, *29*, 699–714. http://doi.org/10.1016/j.ecresq.2014.02.003

Barac, R., Moreno, S., & Bialystok, E. (2016). Behavioral and electrophysiological differences in executive control between monolingual and bilingual children. *Child Development*, *87*, 1277–1290. http://doi.org/10.1111/cdev.12538

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

Blom, E., Küntay, A. C., Messer, M., Verhagen, J., & Leseman, P. (2014). The benefits of being bilingual: Working memory in bilingual Turkish-Dutch children. *Journal of Experimental Child Psychology*, *128*, 105–119. http://doi.org/10.1016/j.jecp.2014.06.007

Carlson, S. M., & Meltzoff, A. N. (2008). Bilingual experience and executive functioning in young children. *Developmental Science*, *11*, 282–298. http://doi.org/10.1111/j.1467-7687.2008.00675.x

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

Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia*, *44*, 2037–2078. http://doi.org/10.1016/j.neuropsychologia.2006.02.006

de Bruin, A., Treccani, B., & Della Sala, S. (2015). Cognitive advantage in bilingualism: an example of publication bias? *Psychological Science*, *26*, 99–107. http://doi.org/10.1177/0956797614557866

de Graaff, R. (2015). Vroeg of laat Engels in het basisonderwijs: Wat levert het op? [Early or late English in primary education: Does it matter?]. *Levende Talen*

Tijdschrift, *16*(2), 3–15.

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*, 135–168. http://doi.org/10.1146/annurev-psych-113011-143750

Duñabeitia, J. A., Hernández, J. A., Antón, E., Macizo, P., Estévez, A., Fuentes, L. J., & Carreiras, M. (2014). The inhibitory advantage in bilingual children revisited: myth or reality? *Experimental Psychology*, *61*, 234–251.

Dunn, L. M., & Dunn, D. M. (2007). *PPVT-4 Manual*. Bloomington: NCS Pearson.

Dunn, L. M., Dunn, L. M., & Schlichting, L. (2005). *Peabody Picture Vocabulary Test-III-NL* (2nd ed.). Amsterdam: Pearson Assessment and Information.

EP-Nuffic. (n.d.). Locaties vvto-scholen. Retrieved September 8, 2016, from https://www.epnuffic.nl/primair-onderwijs/talenonderwijs/vroeg-vreemdetalenonderwijs-vvto/locaties-vvto-scholen

EP-Nuffic. (2015). *Stap in de wereld. Starten met vroeg vreemdetalenonderwijs. Stappenplan. [Step into the world. Starting with early foreign language education. Roadmap]*.

Gathercole, V. C. M., Thomas, E. M., Kennedy, I., Prys, C., Young, N., Guasch, N. V., … Jones, L. (2014). Does language dominance affect cognitive performance in bilinguals? Lifespan evidence from preschoolers through older adults on card sorting, Simon, and metalinguistic tasks. *Frontiers in Psychology*, *5*, 11. http://doi.org/10.3389/fpsyg.2014.00011

Gerstadt, C. L., Hong, Y. J., & Diamond, A. (1994). The relationship between cognition and action: performance of children 3 1/2-7 years old on a Stroop-like day-night test. *Cognition*, *53*(2), 129–153. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7805351

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*, *14*, 289–302. http://doi.org/10.1177/1367006910367846

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, *1*(2), 67–81. http://doi.org/10.1017/S1366728998000133

Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology*, *25*, 515–530. http://doi.org/10.1080/20445911.2013.796377

Heck, R. H., Thomas, S. L., & Tabata, L. N. (2014). *Multilevel and Longitudinal Modeling with IBM SPSS* (2nd ed.). New York: Taylor & Francis.

Hedge, A., & Marsh, N. W. A. (1975). The effect of irrelevant spatial correspondences on two-choice response time. *Acta Psychologica*, *39*, 427–439.

Kuppens, A. H. (2010). Incidental foreign language acquisition from media exposure. *Learning, Media and Technology*, *35*(1), 65–85. http://doi.org/10.1080/17439880903561876

Lobo, V. R. (2013). *Teaching L2 English at a very early age : A study of Dutch schools*. Radboud University. Retrieved from http://www.lotschool.nl

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "Frontal Lobe" tasks: a latent variable analysis.

Cognitive Psychology, *41*(1), 49–100. http://doi.org/10.1006/cogp.1999.0734

Mohades, S. G., Struys, E., Van Schuerbeek, P., Baeken, C., Van De Craen, P., & Luypaert, R. (2014). Age of second language acquisition affects nonverbal conflict processing in children: An fMRI study. *Brain and Behavior*, *4*, 626–642. http://doi.org/10.1002/brb3.246

Morales, J., Calvo, A., & Bialystok, E. (2014). Working memory development in monolingual and bilingual children. *Journal of Experimental Child Psychology*, *114*(2), 187–202. http://doi.org/10.1016/j.jecp.2012.09.002.Working

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*(2), 133–142. http://doi.org/10.1111/j.2041-210x.2012.00261.x

Paap, K. R., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology*, *66*(2), 232–258. http://doi.org/10.1016/j.cogpsych.2012.12.002

Paap, K. R., Johnson, H. A., & Sawi, O. (2014). Are bilingual advantages dependent upon specific tasks or specific bilingual experiences? *Journal of Cognitive Psychology*, *26*, 615–639. http://doi.org/10.1080/20445911.2014.944914

Paap, K. R., Johnson, H. A., & Sawi, O. (2015). Bilingual advantages in executive functioning either do not exist or are restricted to very specific and undetermined circumstances. *Cortex*, *69*, 265–278. http://doi.org/10.1016/j.cortex.2015.04.014

Paap, K. R., Myuz, H. A., Anders, R. T., Bockelman, M. F., Mikulinsky, R., & Sawi, O. M. (2017). No compelling evidence for a bilingual advantage in switching or that frequent language switching reduces switch cost. *Journal of Cognitive Psychology*, *29*, 89–112. http://doi.org/10.1080/20445911.2016.1248436

Poarch, G. J., & Bialystok, E. (2015). Bilingualism as a model for multitasking. *Developmental Review*, *35*, 113–124. http://doi.org/10.1016/j.dr.2014.12.003

Poarch, G. J., & van Hell, J. G. (2012). Executive functions and inhibitory control in multilingual children: Evidence from second-language learners, bilinguals, and trilinguals. *Journal of Experimental Child Psychology*, *113*, 535–551. http://doi.org/10.1016/j.jecp.2012.06.013

Purić, D., Vuksanović, J., & Chondrogianni, V. (2017). Cognitive advantages of immersion education after one year: Effects of amount of exposure. *Journal of Experimental Child Psychology*, *159*, 296–309. http://doi.org/10.1016/j.jecp.2017.02.011

Ratiu, I., & Azuma, T. (2014). Working memory capacity: Is there a bilingual advantage? *Journal of Cognitive Psychology*, *27*, 1–11. http://doi.org/10.1080/20445911.2014.976226

Simon, J. R., & Small, A. M. (1967). Processing auditory information : interference. *Journal of Applied Psychology*, *53*(5), 433–435.

Thijs, A., Trimbos, B., Tuin, D., Bodde, M., & de Graaff, R. (2011). *Engels in het basisonderwijs [English in primary education]*. Enschede.

Thomas-Sunesson, D., Hakuta, K., & Bialystok, E. (2018). Degree of bilingualism modifies executive control in Hispanic children in the USA. *International Journal of Bilingual Education and Bilingualism*, *21*, 197–206.

http://doi.org/10.1080/13670050.2016.1148114

Tse, C. S., & Altarriba, J. (2014). The relationship between language proficiency and attentional control in Cantonese-English bilingual children: Evidence from Simon, Simon switching, and working memory tasks. *Frontiers in Psychology*, *5*, 954. http://doi.org/10.3389/fpsyg.2014.00954

Unsworth, S., Persson, L., Prins, T., & de Bot, K. (2015). An investigation of factors affecting early foreign language learning in the Netherlands. *Applied Linguistics*, *36*, 527–548. http://doi.org/10.1093/applin/amt052

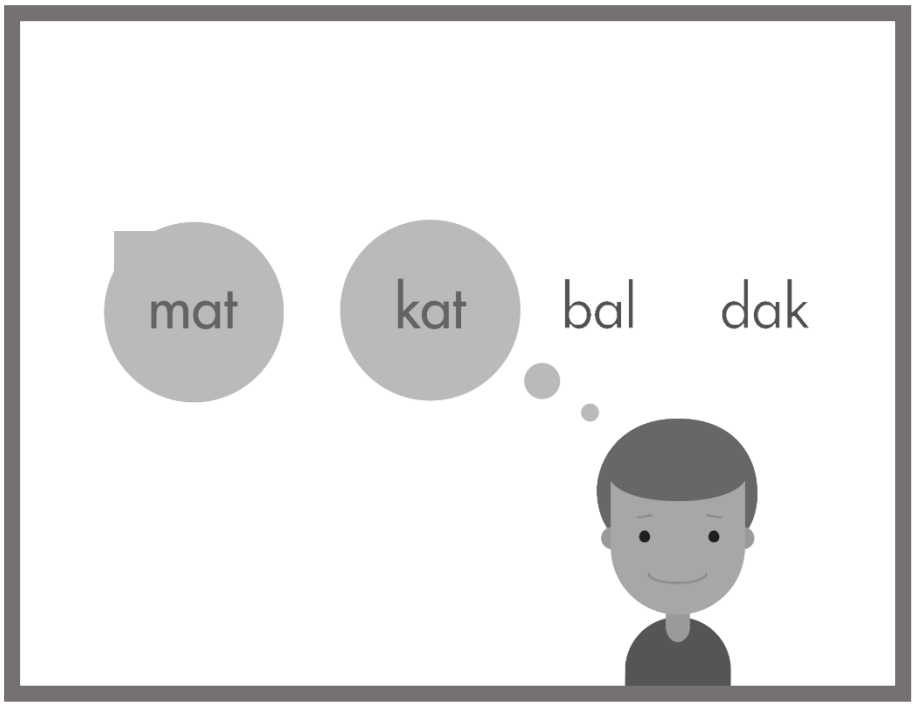
van der Leij, A., Bekebrede, J., & Kotterink, M. (2010). Acquiring reading and vocabulary in Dutch and English: The effect of concurrent instruction. *Reading and Writing*, *23*, 415–434. http://doi.org/10.1007/s11145-009-9207-5

Vega, C., & Fernandez, M. (2011). Errors on the WCST correlate with language proficiency scores in Spanish-English bilingual children. *Archives of Clinical Neuropsychology*, *26*, 158–164. http://doi.org/10.1093/arclin/acq097

Wechsler, D., & Naglieri, J. A. (2008). *WNV NL. Wechsler Nonverbal Scale of ability. Nederlandstalige bewerking.* Amsterdam: Pearson.

Yow, W. Q., & Li, X. (2015). Balanced bilingualism and early age of second language acquisition as the underlying mechanisms of a bilingual executive control advantage: Why variations in bilingual experiences matter. *Frontiers in Psychology*, *6*, 164. http://doi.org/10.3389/fpsyg.2015.00164

Zelazo, P. D. (2006). The Dimensional Change Card Sort (DCCS): a method of assessing executive function in children. *Nature Protocols*, *1*(1), 297–301. http://doi.org/10.1038/nprot.2006.46



Chapter 3: Differences in phonological awareness development: Are there positive or negative effects of bilingual experience?

This chapter is based on:
Goriot, C., Unsworth, S., van Hout, R., Broersma, M., McQueen, J. M. (2018). Differences in phonological awareness development: Are there positive or negative effects of bilingual experience? Manuscript submitted for publication

Chapter 3

Differences in phonological awareness development: Are there positive or negative effects of bilingual experience?

Abstract

This study asked whether bilingual experience helps or hinders the development of phonological awareness. It also asked how much bilingual experience is needed for differences in phonological awareness to show and whether individual differences in language and memory development alter the relation between bilingual experience and phonological awareness development. Participants were Dutch-speaking pupils attending either mainstream Dutch primary schools or early-English schools in which English lessons are given from the beginning of primary school, and simultaneous Dutch-English bilinguals. Children were four to seven years old and were in the first three years of primary school. We investigated the acquisition of phonological awareness (rhyming, phoneme blending, onset phoneme identification, and phoneme deletion) and its relation to Dutch vocabulary, English vocabulary, working memory and short-term memory development, and balance between Dutch and English vocabulary. Significant but small effects of bilingualism were found on onset phoneme identification and phoneme deletion, but post-hoc comparisons revealed no robust pairwise differences between the groups. Furthermore, the effects of bilingualism sometimes disappeared when individual differences in language or memory development were taken into account. These results show that learning two languages simultaneously is not beneficial to – and importantly, also not detrimental to – the development of phonological awareness.

3.1 Introduction

Well-developed phonological awareness skills are an important precursor for learning to read and write (Sodoro, Allinder, & Rankin-Erickson, 2002). This is therefore a topic of interest in language learning research as well as in educational practice, with researchers and teachers paying much attention to the development of these skills. Phonological awareness development starts in preschool, but develops especially quickly once literacy instruction begins (Anthony & Francis, 2005). Now that many schools in Europe have lowered the starting point of early foreign language instruction to exactly the point in time that phonological awareness skills are developing (i.e., kindergarten), an important question is whether early foreign language education has an influence, either positive or negative, on the development of phonological awareness. That question is addressed here for children in the Netherlands with early foreign language instruction in English.

On the one hand, it could be assumed that early foreign language education may help the development of phonological awareness. Many parents and teachers on the other hand have concerns that this type of education may negatively influence pupils' first language (L1) development, or their development in both languages (Goorhuis-Brouwer & de Bot, 2010). Previous research has shown mixed results. Several studies with proficient second language (L2) learners and bilinguals have suggested that bilingualism may positively influence phonological awareness skills (e.g. Bruck & Genesee, 1995; Marinova-Todd, Zhao, & Bernhardt, 2010). Other

studies did however not find differences between monolingual and bilingual groups (e.g. Bialystok, Majumder, & Martin, 2003), or even found an advantage in favour of the monolingual children (e.g. Janssen, Segers, McQueen, & Verhoeven, 2015). In the present study, we investigated whether children who are following an early foreign language learning programme at school differ from their monolingually educated peers in the development of phonological awareness, and if the development of foreign-language learners resembles that of simultaneous bilingual children who are highly proficient in two languages. We thus asked if different types of experience with two languages affects - either helping or hindering - the development of phonological awareness.

3.1.1 Development of phonological awareness in monolinguals and bilinguals

Phonological awareness is the ability to detect and manipulate the different sounds in a language, and to focus on the phonological structure of spoken language instead of on the meaning of the words (Sodoro et al., 2002). Phonological awareness develops during the preschool and early elementary school years, and generally follows a fixed pattern in which children become sensitive to increasingly smaller word units: children first become sensitive to syllables, then learn to detect and manipulate onsets and rimes (the second part of syllable, starting with the vowel), and finally learn to detect and manipulate individual phonemes (Anthony & Francis, 2005).

Although speakers of different languages generally show the same developmental pattern for phonological awareness, the speed with which this development takes place appears to differ according to the linguistic complexity of the language spoken. Language features such as the saliency of syllables, the saliency and complexity of onsets, and the proportion of rime neighbours, are related to children's ability to detect syllables, their onset and phoneme awareness, and their ability to separate onset from rime (onset-rime awareness), respectively (Anthony & Francis, 2005). For example, native speakers of a language with simple syllable structures are likely to develop syllable awareness more quickly than speakers of a language with more complex syllable structure (Anthony & Francis, 2005). Experiences with a specific language thus influence the development of phonological awareness.

If children have experience with two languages, the development of phonological awareness may follow a different path than that of children who only have experience with one language. In an early study on this topic, Rubin and Turner (1989) examined phonological awareness in children learning an L2 via an educational programme, and compared English-speaking pupils in French immersion classrooms to monolingual English pupils. The immersion pupils performed better than monolingual children on an English syllable deletion task ('say cowboy without cow') and a phoneme deletion task ('say mine without /m/'). The authors hypothesized that this heightened phonological awareness was due to the immersion pupils learning an L2, because they would have had to analyse the L2 constantly and explicitly (Rubin & Turner, 1989).

In later studies phonological awareness advantages for immersion and simultaneous bilingual pupils have not always been consistent. Whereas some studies indeed report an advantage in favour of the immersion or bilingual group

(Chen, Xu, Nguyen, Hong, & Wang, 2010; Kang, 2012; Kuo & Anderson, 2010; Marinova-Todd et al., 2010), others show that immersion and bilingual pupils show advantages on some tasks, but disadvantages on others (Chen, Wu, & Shu, 2004; Kuo, Uchikoshi, Kim, & Yang, 2016; Loizou & Stuart, 2003), and some researchers even found only disadvantages for the bilingual group (Janssen et al., 2015; Janssen, Segers, McQueen, & Verhoeven, 2017; Lesniak, Myers, & Dodd, 2014). Appendix A provides an overview of prior studies on this topic. There is no clear pattern of results. Of the 19 studies, six show an advantage for immersion and bilingual pupils, four show a disadvantage, eight show mixed results and one found only null effects.

Various reasons have been suggested for the (absence of) differences between monolingual and bilingual children's phonological awareness development, and these reasons sometimes contradict each other. Bruck and Genesee (1995), for example, suggested that phonological awareness advantages for L2 learners should be found for specific phonological units that are more salient in the L2 than in the L1. They found that in kindergarten, French immersion pupils performed better than monolingual English pupils on onset-rime awareness and syllable counting tasks. The authors reasoned that this was due to the fact that syllables are more salient in French than in English. Another study (Bialystok et al., 2003; study III), however, used the opposite reasoning. There, it was found that Spanish-English bilingual children showed an advantage over English monolingual children on a phoneme segmentation task, but Chinese-English bilinguals scored significantly lower than the two other groups on the task. The authors provided two possible reasons for the advantage shown by the Spanish-English bilinguals, the first being that English and Spanish are more similar in their sound structure than English and Chinese, which may provide easy access to the phonological structure of the languages. This hypothesis is thus exactly the opposite of that of Bruck and Genesee (1995).

The second reason Bialystok et al. provided was that the simple phonological structure of Spanish promotes phonological awareness, which may in turn enhance children's phonological awareness in English (Bialystok et al., 2003). Loizou and Stuart (2003) had the same reasoning as Bialystok and colleagues (2003), namely that having knowledge of an L2 with a more simple phonological structure may help the development of phonological awareness in the L1. They found that English-Greek bilinguals growing up in the United Kingdom outperformed English monolinguals on phonological awareness tasks, but Greek-English bilinguals growing up in Greece did not show an advantage over Greek monolinguals (Loizou & Stuart, 2003). The explanation is that English-Greek bilinguals in the UK had an L2 (Greek) that was phonologically simpler than their L1 (English), which would make them phonologically aware more easily and rapidly. For the Greek-English children in Greece it was the other way around, and therefore they would not show such a benefit (Loizou & Stuart, 2003).

Other studies, however, have yielded findings that support the opposite reasoning. For example, Chen et al. (2010) found that nine-year-old Chinese children who received 80 minutes of English lessons per week showed better performance on Chinese Pinyin phonological awareness tasks than their monolingual peers, leading to the conclusion that English lessons enhance Chinese phonological awareness. However, in contrast to what Bialystok et al. (2003) and Loizou and Stuart (2003) reasoned, Chen et al. argued that the Chinese-English L2

learners had an advantage because English is phonologically *more* complex than Chinese (Chen et al., 2010).

Whilst the aforementioned researchers all attributed differences in phonological awareness of monolingual and bilingual children to characteristics of specific languages or language combinations, Marinova-Todd et al. (2010) argued that such differences were due to a more general effect of bilingualism. They found that Mandarin-English bilinguals outperformed Mandarin monolinguals on a Mandarin tone discrimination task, and outperformed English monolinguals on an English phonemic awareness test. They argued that this heightened tone sensitivity and phonemic awareness cannot be explained by having knowledge of English and Mandarin, respectively, and therefore must be an effect of bilingualism more generally (Marinova-Todd et al., 2010).

In summary, studies have found positive as well as negative relations between bilingualism and phonological awareness, and different and sometimes contradictory reasons have been provided for these findings. A possible explanation for the mixed results is that various factors that are not specifically related to bilingualism are likely to have an influence on the development of phonological awareness as well as on the relation between bilingualism and phonological awareness.

One such factor is the influence of reading instruction (Bruck & Genesee, 1995). Since phonological awareness is related to learning to read and write, any possible phonological awareness advantage for L2 learners and bilinguals over monolinguals may disappear once children learn to read and write: when monolinguals and bilinguals are learning to read in the same language, literacy instruction will help monolinguals to catch up with the bilinguals. At least two studies provide evidence for this. Bruck and Genesee (1995) found that immersion pupils had an advantage over monolingual pupils before, but not after literacy instruction had started. In a study that compared French-German six-year-old partial immersion education pupils to French monolingual pupils, no differences in phonological awareness abilities were found (Reder, Marec-Breton, Gombert, & Demont, 2013), which was attributed to the fact that both groups were already learning to read and write in French, and pupils were already paying attention to the phonological structure of the language.

A meta-analysis on studies investigating the development of phonological awareness in bilingual children showed that learner-specific characteristics influence the development of phonological awareness in English (Branum-Martin, Tao, Garnaat, Bunta, & Francis, 2012). One of these characteristics is children's age. A longitudinal study showed that monolingual English children performed significantly better on a rhyme matching task (matching words that sound the same) when they were 50 months old than when they were 46 months old (Carroll, Snowling, Stevenson, & Hulme, 2003). Although it is known that older children are likely to perform better on phonological awareness tasks than younger children, researchers do not always take age into account when assessing bilingual children's phonological awareness (Branum-Martin et al., 2012).

Branum-Martin et al. (2012) suggested that other learner characteristics, such as cognitive development, may play a role in the development of phonological awareness as well, but due to lack of information in the studies included in their

meta-analysis this could not be investigated. In line with this suggestion, Bialystok et al. found a significant correlation between phonological awareness and working memory for monolinguals and bilinguals (Bialystok et al., 2003). Similarly, a study on Dutch-Turkish children showed that both for Dutch monolingual and Dutch-Turkish bilingual pupils, the level of Dutch vocabulary was positively correlated with their performance on rhyme awareness and phoneme blending tasks in Dutch (Janssen et al., 2017). Likewise, research on Spanish-English bilingual children has shown that children's level of vocabulary in the L1 (Spanish) is positively correlated with their performance on English phonological awareness tasks (Atwill, Blanchard, Gorin, & Burstein, 2007). As most of the previous studies did not include such learner-specific variables, it remains unknown to what extent such developmental differences relating to them influenced differences between monolinguals and bilinguals in phonological awareness.

In addition to learner characteristics, the amount of exposure to the L2 may also influence the relation between bilingual experience and phonological awareness. Research suggested that even a limited amount of exposure to the L2 may already influence pupils' phonological awareness skills in the L1, but that more exposure to the L2 is related to bigger influences on the L1. Chen et al. (2010) showed that pupils who received a limited amount of L2 instruction (80 minutes per week) outperformed their peers from a monolingual programme on onset and rime awareness tasks in the L1. In a follow-up study, children who were enrolled in a regular English programme (80 minutes per week) and children in an intensive English programme (over 10 hours per week) were tested five times over the course of two years on their phonological awareness in Chinese and English. The pupils following the intensive programme outperformed the pupils in the regular programme on Chinese phonological awareness, but only at the end of grade two (Chen et al., 2010). This suggests that a certain proficiency level must be reached before more proficient L2 learners start to outperform their less proficient peers.

Subsequent research confirmed the findings of Chen et al. (2010). Korean pupils attending English immersion kindergarten obtained higher scores on both English and Korean phonological awareness tasks than Korean peers who only got 15 minutes of English lessons per day. In both groups, English phonological awareness predicted Korean phonological awareness and vice versa (Kang, 2012). This indicates that even a small amount of L2 instruction may already influence pupils' phonological awareness in their native language. However, in both Chen et al. (2010; study II) and Kang (2012), no monolingual group was included, and hence it remains unknown how much bilingual experience is needed for effects on phonological awareness to show.

In summary, bilingual children and L2 learners sometimes outperform monolingual children on phonological awareness tasks, and sometimes they do not. Various, and sometimes contradicting, reasons have been put forward to account for this, including the characteristics of the languages in question, the specific language combinations, and the environment (i.e., factors such as the start of literacy instruction). In addition, children's own level of linguistic and/or cognitive development, and the amount of exposure to the L2 may influence the relation between bilingual experience and phonological awareness. The research on these influences is however scarce. It remains unanswered how little or much bilingual

experience is needed for differences in phonological awareness to show, and how individual differences between L2 learners may modulate the relation between bilingual experience and phonological awareness. The present study addresses those questions.

The Netherlands provide the perfect situation to investigate such questions. In general, Dutch primary schools have a monolingual (Dutch) curriculum, and English lessons are usually not given until the penultimate grade when children are around ten years old. Just like in many other European countries, however, primary schools have started to lower the starting age of English lessons to the start of primary school (Enever et al., 2011). Nowadays, as many as one in five primary schools provide English lessons from the moment that children enter primary school when they are four years old. Unlike the participants in previous studies, who were instructed for at least 80 minutes per week in English or even attended immersion classes, Dutch pupils in early-English school usually do not get more than 60 minutes of English per week (Jenniskens et al., 2017; Thijs, Trimbos, Tuin, Bodde, & de Graaff, 2011). Dutch pupils are thus likely not as experienced in the L2 as the children in the previous studies. It remains unknown whether this limited experience influences the development of phonological awareness and, if so, whether that influence is comparable to that of children with more language experience, namely children who are raised as simultaneous bilinguals.

3.1.2 The current study

The aim of this study was threefold. As the previous literature provides a contradictory picture, the first aim was to investigate whether bilingualism helps or hinders the development of phonological awareness. Our second aim was to examine, if there are differences in phonological awareness, how much bilingual experience is needed for them to emerge. The third aim was to investigate whether individual developmental differences alter the relation between bilingualism and phonological awareness. To answer these questions, we compared phonological awareness skills of three groups of children: Dutch pupils enrolled in mainstream schools who did not have any foreign-language lessons, Dutch pupils enrolled in early-English schools who had English lessons from the moment they entered primary school at the age of four, and children who were being raised bilingually in Dutch and English. To investigate how phonological awareness develops as instruction progresses, we included children from the first three grades of primary school: grade 1 (kindergarten, 4-5 year-olds), grade 2 (kindergarten, 5-6 year-olds), and grade 3 (start of reading instruction, 6-7 year-olds).

Despite the fact that many of the previous studies on the relation between bilingual experience and phonological awareness showed mixed results, most of the previous studies showed a bilingual advantage on a number of tasks. Assuming that the positive effects of bilingual experience can be replicated, our first hypothesis was that early-English education helps rather than hinders the development of phonological awareness. Previous research showed that pupils who had had more exposure to the L2 outperformed pupils with less exposure (Chen et al., 2010; Kang, 2012). Our second hypothesis therefore was that more bilingual experience would be associated with greater phonological awareness skills. In other words, we expected the early-English pupils to outperform the mainstream pupils, and the bilingual

children to outperform both other groups. In line with the findings of previous studies (Bruck & Genesee, 1995; Reder et al., 2013), we expected these advantages to show in the first two grades of primary school, and to disappear as soon as reading instruction started in grade 3.

Previous research (Branum-Martin et al., 2012; Carroll et al., 2003) has suggested that within-grade age differences play a role in phonological awareness development of bilingual children. We therefore investigated not only differences in phonological awareness between children in different grades, but also within-grade age differences.

Our third aim was to investigate whether the relation between bilingual experience and phonological awareness would change if learner-specific characteristics such as cognitive development and language development are taken into account. We therefore incorporated various measures which have previously been shown to be involved in the development of phonological awareness: working memory (Bialystok et al., 2003), short-term memory (Sodoro et al., 2002), vocabulary knowledge in the L1 (Janssen et al., 2017), and vocabulary knowledge in the L2 (Atwill et al., 2007). Since it has previously been argued that especially childhood bilingualism cannot be defined as a simple categorical variable, because proficiency can vary in the two languages of a bilingual child (Luk & Bialystok, 2013), we also included a measure of language balance. A child who has equal proficiency in both languages is a balanced bilingual, one who is more proficient in one language than the other is unbalanced. We operationalised language balance as the ratio between Dutch and English vocabulary development. Given that in previous work (Goriot, Broersma, McQueen, Unsworth, & Van Hout, 2018) we found that lexical balance may be a more important predictor for developmental differences between monolingual and bilingual children than grouping children in a monolingual or bilingual group, we expected lexical balance to be positively related to phonological awareness. In summary, our third hypothesis was that possible relations between bilingualism and phonological awareness would be affected by individual differences in language and memory.

The results of this study contribute to knowledge about whether learning two languages at the same time influences the development of phonological awareness, either positively or negatively, and, if there is such an effect, how much bilingual experience is needed before it starts to influence development. In addition, this study addresses how other differences between learners, such as differences in age or memory development, may influence the relation between bilingualism and phonological awareness.

3.2 Method

3.2.1 Participants

Participants were 294 children (151 boys and 143 girls), who were either L2 learners of English at early-English schools, pupils at mainstream schools (the control group), or English-Dutch bilinguals. The L2 learners (n = 123) attended an early-English school (n = 4) at which English lessons started from the moment children enter primary school (i.e., kindergarten; age four). These four schools had a certificate from an independent organization that they taught at least 60 minutes of English lessons per week, and that those lessons were given by a teacher who had at

least B2 level (intermediate) of English in terms of the Common European Framework of Reference. The control group consisted of 121 children who attended a mainstream Dutch primary school in which English education did not start before the penultimate grade (n = 5 schools). These children were functionally monolingual speakers of Dutch, with some exposure to English via e.g. media and music. All nine schools participated voluntarily. In each school, up to approximately 10 children from each grade (1, 2, or 3) participated. We asked the head teachers to select children who were not exposed to another language at home, and who did not have any developmental disorders or hearing or sight problems.

The remaining children (n = 50) were Dutch-English bilinguals who were raised bilingually at home, and had at least one parent who was a native speaker of English. Two additional children were tested, but their data were removed since it turned out that they failed to meet this criterion. According to parental report, none of the participating bilingual children had any known developmental disorders, or sight or hearing impairments.

Children were in one of three grades: grade 1 (kindergarten year 1; 4-5 year-olds; n = 93), grade 2 (kindergarten year 2; 5-6 year-olds, n = 95), or grade 3 (first year of formal schooling; 6-7 year-olds; n = 106). Parents of all children gave informed consent for participation.

3.2.2 Instruments

Phonological awareness. The development of phonological awareness was assessed with several tasks of the *Screeninginstrument Beginnende Geletterdheid* [Diagnostic Instrument for Emergent Literacy] (Vloedgraven, Keuning, & Verhoeven, 2009). The specific tasks were dependent on the grade of the participant. For children in grade 1, we assessed rhyming and phoneme blending, for those in grade 2, rhyming, phoneme blending and onset phoneme identification, and for those in grade 3, phoneme blending, onset phoneme identification, and phoneme deletion. All tasks consisted of two practice trials and 15 test trials. In each task, children were presented with three full colour pictures that appeared on the screen one-by-one while the pre-recorded name of the picture was played over the computer's speakers. The task was orally presented after the final response alternative appeared on the screen (see Figure 3.1 for the design). Scores were computed as the total number of correct responses on the sub-task. For each of the sub-tasks, an example is presented below:

Rhyming. Children were asked to identify the picture that rhymed with the target, for example: "*Kat, bal, dak; wat rijmt op mat?" ["Cat, ball, roof; what rhymes with mat?"]*.

Phoneme blending. Children were asked to identify a word based on the individual phonemes of that word. An example is: "*Sport, spons, storm; 's' 'p' 'o' 'r' 't'." [Sport, sponge, storm; 's' 'p' 'o' 'r' 't'].*

Onset phoneme identification. Children were asked to indicate which word started with the same phoneme as the target, for example: *"Nek, maan, hol; de n van neus." ["Neck, moon, cave, the n of nose"].*

Phoneme deletion. Children were asked to identify a word after removing one sound from another word. An example item is: *"Net, nek, bed; nest, laat de 's' weg."* ["*Net, neck, bed; nest leave out the 's'"]*

Figure 3.1. Design of the Diagnostic Instrument for Emergent Literacy. The pictures appeared on the screen one-by-one. The task was orally presented after the last picture appeared on the screen.

Vocabulary. English and Dutch vocabulary were assessed with the PPVT-4, and the PPVT-III-NL, respectively (Dunn & Dunn, 2007; Dunn, Dunn, & Schlichting, 2005). The English version consists of 228 items, grouped in 19 sets of 12 items each, in which the child is presented with 4 pictures. The name of the picture is orally presented (in this study a recording by a native speaker was played), and the child has to indicate which picture corresponds to the spoken description. The starting set depends on the age of the child. Administration rules as stated in the manual were followed, which means that the basal set is the lowest set in which a child made maximally one error, and the final set is the highest set in which the child made eight or more errors. The raw score is calculated by subtracting the number of errors from the number of the highest item that the child made. The administration of the PPVT-III (Dutch) is similar to the PPVT-4, except for the fact that the test consists of 204 items and the basal set and ceiling set are determined by the lowest and highest set in which the child makes maximally four and nine or more errors, respectively.

Working memory. The subtest 'Odd One Out' from the Automated Assessment of Working Memory (Alloway, Gathercole, Kirkwood, & Elliott, 2008) was assessed. This is a computerised test in which the child is presented with three pictures, surrounded by rectangles. The child has to indicate which of the pictures is the odd one out, and remember its location. After the pictures disappear, only the rectangles remain and the participant has to indicate the location of the odd one out. The test starts with a trial of one sequence of pictures, after which the correct location has to be remembered. After four correct trials, a trial of two sequences of three pictures is presented, and both locations have to be remembered and indicated by the participant. The test stops after the participant responds erroneously to three trials of the same length. The maximum number of sequences is seven. The raw score is calculated as the number of correctly performed trials.

Short-term memory. Short-term memory was examined by a word span task, taken from the subtest *Geheugen* [Memory] in the standardized *Screeningstest voor Taal- en Leesproblemen* [Diagnostic Test for Language and Literacy Problems] (Verhoeven, 2005). The child was presented with a sequence of two pre-recorded monosyllabic words, which she had to repeat in the same order. After two sequences of the same length, the length increased by one word. The test was stopped after the

child had responded erroneously to four consecutive trials. The score was calculated as the number of correctly repeated trials.

3.2.3 Procedure

All children were tested individually in a quiet room at their school, or, in the case of the bilingual children, at home. For pupils tested at schools, testing was done in two sessions of twenty minutes. The first session included the word span task, followed by the Odd One Out and the PPVT-4, and the second session included the Screening instrument followed by the PPVT-III Dutch. For bilingual children, testing consisted of two or three sessions, depending on their age. For them, sessions contained the tasks reported here, as well as other tasks as part of a larger test battery (the results of which are reported elsewhere, Goriot et al., 2018); crucially, the relative order of the tasks reported on here was the same as for the other groups. For all participants, there was always at least one day between the sessions.

3.2.4 Analyses

We performed four separate ANCOVAs to investigate the effects of Bilingualism, Grade, and Age on performance on each of the four phonological awareness tasks. In the first step, three-way and two-way interactions were included in each model, and removed if they were non-significant. After we established the base models, we investigated the effect of individual differences. We took the base model that was established in the first step, and added the covariates (Dutch vocabulary, English vocabulary, lexical balance, short-term memory, and working memory) one-by-one to each model. Finally, we performed similar analyses for the mainstream and early-English pupils without the bilingual children. These multilevel linear mixed-effects model analyses allowed us to include the effect of School. The ANCOVAs were performed in SPSS (version 23.0). The multilevel linear mixed-effects analyses were performed in R (package lme4, in platform R, version 3.4.1).

3.3 Results

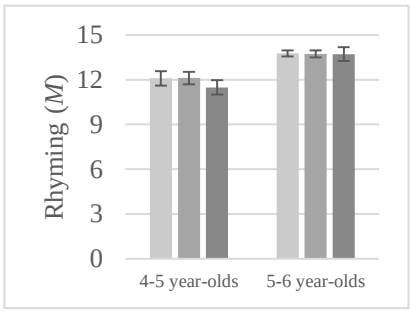
Table 3.1 shows the descriptive statistics for all groups. Figures 3.2a to 3.2d show the average outcomes for each of the phonological awareness measures for all nine groups (three grades by three bilingual categories). To investigate whether there is an effect of degree of bilingualism on phonological awareness, ANCOVAs with Grade (G1, G2, G3), Bilingual Category (mainstream, early-English, bilingual) and Age in months were performed. Age in months was included as covariate since given the fast development of phonological awareness in the age range involved, the pupils' age may provide additional information (in addition to grade) about differences in children's performance. Moreover, an ANOVA with Grade and Bilingual Category showed that the three groups of children differed in age ($F(2,83) = 5.80$, $p = .002$, Tukey HSD: mainstream > early-English > bilingual).

One ANCOVA for each phonological awareness measure was conducted. The results are shown in Table 3.2. The models reported in Table 3.2 are the base models, which were used for all further analyses. Table 3.2 shows that there was a main effect of Age, and a main effect of Grade for Rhyming, Phoneme Blending, and Onset Phoneme identification. The effect of Age also differed with Grade,

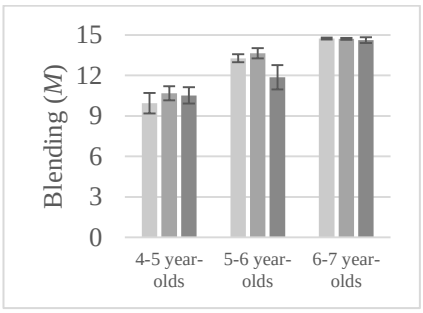
resulting in significant interaction effects. Figures 3.3a to 3.3c show that the children in the higher grades were all at ceiling for the phonological awareness tasks, and therefore age was a relevant predictor of phonological awareness only in the younger groups. For Onset Phoneme Identification and Phoneme Deletion, there were main effects of Age, and Bilingual Category, and interaction effects between these two variables. Figures 3.4a and 3.4b show that there is a stronger relation between Age and performance on the two tasks for bilingual children than for mainstream and early-English pupils, especially for Phoneme Deletion.

Table 3.1
Descriptive Statistics for Age and the Phonological Awareness Measures for All Groups

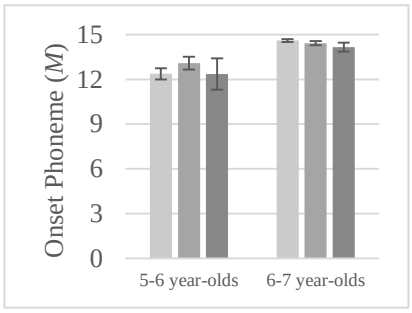
			Grade 1 (4-5 year-olds)			Grade 2 (5-6 year-olds)			Grade 3 (6-7 year-olds)		
			Control	Early-English	Bilingual	Control	Early-English	Bilingual	Control	Early-English	Bilingual
Number of participants		*N*	32	38	23	41	40	14	48	45	13
Age (in years)		*M*	5.0	4.9	4.8	6.0	5.8	5.8	6.9	6.8	6.8
		(SD)	(0.4)	(0.3)	(0.4)	(0.3)	(0.4)	(0.5)	(0.3)	(0.5)	(0.5)
Phonological awareness (max. score of 15 on each task)	Rhyming	*M*	12.09	12.11	12.09	13.76	13.73	13.71	-		-
		(SD)	(2.77)	(2.59)	(2.77)	(1.36)	(1.50)	(1.73)			
	Phoneme blending	*M*	9.94	10.68	9.94	13.27	13.63	11.86	14.73	14.71	14.62
		(SD)	(4.26)	(3.21)	(4.26)	(1.87)	(2.42)	(3.42)	(0.49)	(0.55)	(0.77)
	Onset Phoneme	*M*	-	-	-	12.36	13.08	12.36	14.60	14.42	14.15
		(SD)				(2.34)	(2.70)	(3.91)	(0.68)	(0.94)	(1.07)
	Phoneme Deletion	*M*	-	-	-	-	-	-	9.14	10.33	9.40
		(SD)							(3.58)	(3.17)	(4.30)



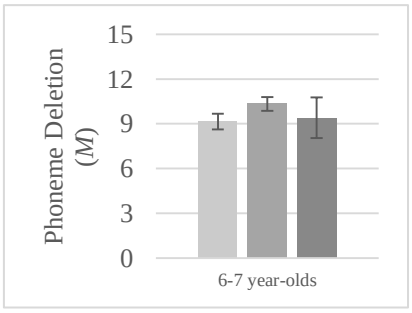
a) Rhyming



b) Phoneme blending



c) Onset Phoneme



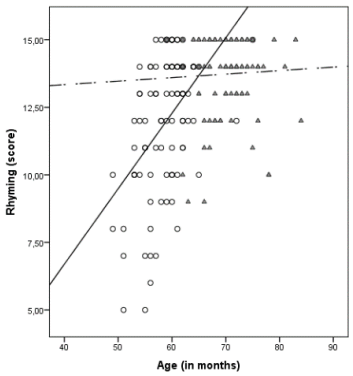
d) Phoneme Deletion

Figures 3.2a to 3.2d. Scores on the Phonological Awareness Tasks, by Grade and Bilingual Status (■ mainstream, ■ early-English, ■ bilingual) with *SE*s.

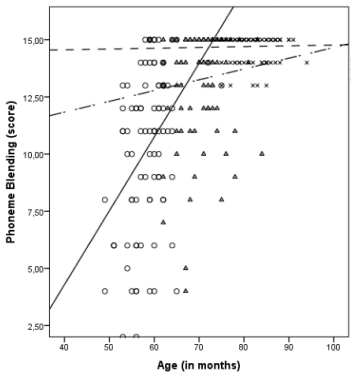
Table 3.2
ANOVAs with Grade, Age and Bilingual Category as Independent Variables and the Four Phonological Awareness Tasks as Dependent Variables

	Rhyming				Phoneme blending				Onset phoneme identification				Phoneme deletion			
	df	*F*	η_p^2	Post hoc	*df*	*F*	η_p^2	Post hoc	*df*	*F*	η_p^2	Post hoc	*df*	*F*	η_p^2	Post hoc
Bilingual Category	2	0.21			2	1.81			2	3.48*	.036	No diffs.	2	4.32*	.082	No diffs.
Grade	1	17.19***	.088		2	12.29***	.081		1	5.96*	.031			-		
Grade x Category	2	0.18			4	1.57			2	2.39				-		
Age	1	20.13***	.102		1	17.58***	.059		1	7.23**	.037		1	4.64*	.046	
Age x Bilingual Category	2	-			2	-			2	3.32*	.024		2	4.33*	.082	
Age x Grade	1	16.77***	.086		2	11.02***	.073		1	5.30*	.027			-		
Error	176				278				189				97			
R^2	.273				.428				.220				.073			

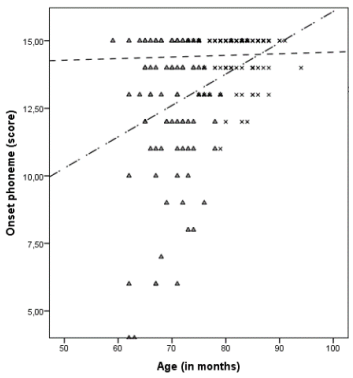
*p < .05; **p < .01; ***p < .001

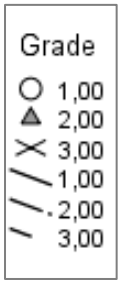


a) Rhyming



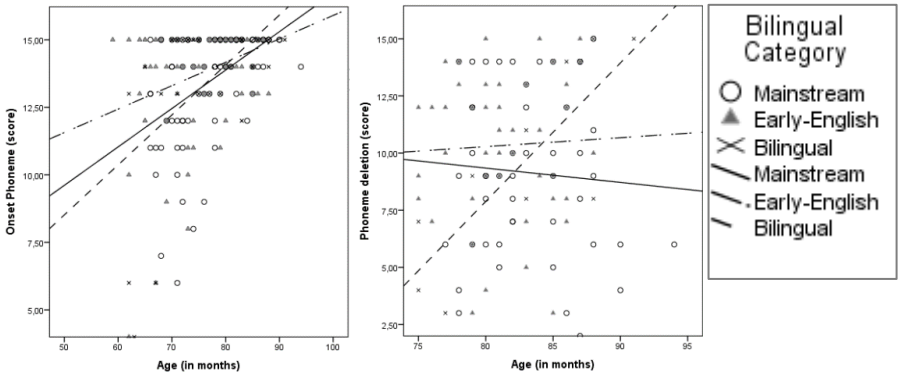
b) Phoneme Blending





c) Onset Phoneme

Figure 3.3. Relation between Age and Rhyming scores (a), Age and Blending scores (b), and Age and Onset Phoneme scores (c) for the different grades.



a) Onset Phoneme b) Phoneme Deletion

Figure 3.4. Relation between Age and Onset Phoneme scores (a) and Age and Phoneme deletion scores (b) for the different groups.

3.3.1 Including covariates in the model

After having established the base models, we checked whether individual differences in any of the other variables that we measured (Dutch Vocabulary, English vocabulary, short-term memory, or working memory) significantly contributed to phonological awareness, and if these variables changed the relation between bilingual category and phonological awareness. For reasons outlined in the introduction, a measure of lexical balance was also included. In line with our previous study (Goriot et al., 2018), lexical balance was calculated as follows: $\ln(\frac{\text{English vocabulary}}{\text{Dutch vocabulary}})$. A score of 0 means that someone is perfectly balanced, a negative score indicates a greater proficiency in Dutch than in English, and a positive score greater proficiency in English than in Dutch. The descriptive statistics for the covariate measures are shown in Table 3.3.

To check whether the groups differed in any of the covariate measures, ANCOVAs were performed with Grade and Bilingual Category as fixed factors, and Age as covariate. The two-way interactions between Age and Grade and Age and Bilingual Category were included if they were significant. If they were not, they were removed from the model. Table 3.4 shows the results. For all phonological awareness measures, there were effects of Age: older pupils generally obtained higher scores. First, for short-term memory, there was a main effect of Bilingual Category, with bilinguals obtaining better scores than mainstream pupils, and the early-English pupils not differing from either the mainstream or the bilingual children in their performance. Second, for Dutch vocabulary there were no significant differences between the bilingual groups. For English vocabulary and balance, finally, there was a main effect of Bilingual Category: Bilingual pupils had higher scores on the English vocabulary task and a higher lexical balance score than early-English pupils, who in turn had higher scores than the mainstream pupils.

Next, for each of the four phonological awareness measures, we added each covariate (Dutch vocabulary, English vocabulary, lexical balance, short-term

memory, or working memory) to the base model. If there was a significant effect of the covariate, we checked if there were any significant two-way interactions between the covariate and the fixed effects. If that was not the case, these interactions were removed from the analysis. The results are shown in Table 3.5. For Rhyming, adding the covariates to the model did not change the pattern of results that we found in the base model. There were always significant effects of Grade, Age, and the interaction between Grade and Age. There was never a significant effect of Bilingual Category. All covariates except Balance significantly contributed to the Rhyming scores. For Phoneme Blending we found that same pattern of results, except that among the covariates only Dutch vocabulary and Short-term memory showed a significant effect on Phoneme Blending scores.

For Onset Phoneme and Phoneme Deletion, the results were more complex. For Onset Phoneme, adding English vocabulary, Balance, or Short-term memory to the model did not change the pattern of results found in the base model: there were significant effects of Bilingual Category, Grade, Age, and interactions between Bilingual Category and Age, and Grade and Age. Short-term memory also showed a significant effect. The effects of Bilingual Category became non-significant when adding either Dutch vocabulary or Working memory to the model. Neither Dutch vocabulary nor Working memory was significant.

For Phoneme Deletion, adding Dutch vocabulary showed the same results as in the base model: there were significant effects of Bilingual Category, Age, and the interaction between the two. Dutch vocabulary was also significant. The main effect of Age become non-significant when adding English vocabulary, Balance or Short-term memory to the model, while none of these covariates showed a significant effect. Adding Working memory to the base model resulted in a significant effect of Working memory, while all the other effects were non-significant.

In summary, phonological awareness skills seem to be related to various developmental differences in language and memory skills. Especially the development of Dutch vocabulary and short-term memory seem to play an important role, as there were significant effects of these variables for three of the four phonological awareness skills. When any effects of bilingualism were found, they were small and unstable. Despite the main effect of Bilingual Category, pairwise comparisons between the three groups were never significant, meaning that the scores of the three groups did not significantly differ from each other. There were thus no positive, but also no negative effects of bilingualism on phonological awareness skills. Moreover, some effects of Bilingual Category disappeared when covariates were taken into account.

Table 3.3
Descriptive Statistics for the Covariates

		Grade 1 (4-5-year-olds)			Grade 2 (5-6-year-olds)			Grade 3 (6-7-year-olds)		
		Mainstream	Early-English	Bilingual	Mainstream	Early-English	Bilingual	Mainstream	Early-English	Bilingual
Dutch vocabulary (raw score, max. 204)	*M* *(SD)*	74.66 (13.75)	75.08 (10.24)	65.22 (16.26)	86.05 (9.08)	86.43 (9.64)	86.64 (13.41)	95.38 (8.22)	96.91 (9.18)	93.38 (11.2)
English vocabulary (raw score, max. 228)	*M* *(SD)*	18.43 (9.48)	22.13 (9.03)	81.57 (19.25)	23.15 (13.50)	30.73 (11.62)	96.64 (24.05)	33.83 (16.40)	41.49 (15.12)	108.15 (17.37)
Lexical balance	*M* *(SD)*	-1.52 (0.50)	-1.35 (0.63)	0.22 (0.36)	-1.51 (0.70)	-1.11 (0.44)	0.09 (0.23)	-1.15 (0.48)	-0.90 (0.32)	0.14 (0.16)
Nonverbal working memory (accuracy, max. 36)	*M* *(SD)*	9.28 (2.65)	10.45 (3.78)	9.00 (3.01)	13.07 (4.41)	12.75 (2.87)	14.80 (3.36)	17.38 (3.00)	16.91 (4.42)	17.10 (4.9)
Short-term memory (accuracy, max. 12)	*M* *(SD)*	2.81 (1.38)	3.34 (1.51)	3.10 (1.61)	3.66 (1.28)	3.60 (1.17)	4.79 (2.45)	3.88 (1.35)	3.89 (1.25)	5.23 (1.42)

Table 3.4
ANCOVAs with Language Measures and Memory Measures as Dependent Variables

	Dutch vocabulary				English vocabulary				Balance			
	df	*F*	η_p^2	Post hoc	*df*	*F*	η_p^2	Post hoc	*Df*	*F*	η_p^2	Post hoc
Bilingual Category	2	2.57			2	479.47***	.773	M<EE<B	2	7.23**	.049	M<EE<B
Grade	2	4.09*	.028		2	0.12			2	2.95		
Grade x Category	4	1.84			4	0.99			4	0.73		
Age	1	27.22***	.088		1	24.27***	.079		1	5.80*	.020	
Age x Bilingual Category		-				-			2	3.51*	.025	
Age x Grade		-				-			2	3.26*	.023	
Error	282				282				278			
R^2	.492				.776				.582			

*p < .05, **p < .01, ***p<.001
N.B.: See next page

Table 3.4
ANCOVAs with Language Measures and Memory Measures as Dependent Variables

	Working Memory				Short-term memory			
	df	*F*	η_p^2	Post hoc	*df*	*F*	η_p^2	Post hoc
Bilingual Category	2	0.87			2	9.72***	.065	M<B
Grade	2	6.53**	.046		2	5.18**	.036	
Grade x Category	4	1.23			4	2.43*	.034	
Age	1	5.14*	.019		1	10.85**	.038	
Age x Bilingual Category	2	-			2	-		
Age x Grade	2	-			2	5.27**	.037	
Error	269				278			
R^2	.415				.166			

*p < .05, **p < .01, ***p < .001
N.B.: See previous page

Table 3.5
ANCOVAs on the Phonological Awareness Data for each Covariate

		Rhyming		**Phoneme blending**		**Onset phoneme identification**		**Phoneme deletion**	
		Significant effects	R^2	Significant effects	R^2	Significant effects	R^2	Significant effects	R^2
	Base model	Grade Age Grade x Age	.273	Grade Age Grade x Age	.435	Bilingual [no sign. comparisons] Grade Age Bilingual x Age Grade x Age	.220	Bilingual [no sign. comparisons] Age Bilingual x Age	.073
	Dutch vocabulary	Grade Age Grade x Age Dutch	.325	Grade Age Grade x Age Dutch	.453	Grade Age Grade x Age	.228	Bilingual [no sign. comparisons] Age Bilingual x Age Dutch	.128
Covariate	English vocabulary	Grade Age Grade x Age English	.291	Grade Age Grade x Age	.430	Bilingual [no sign. comparison] Grade Age Bilingual x Age Grade x Age	.219	Bilingual [no sign. comparisons] Bilingual x Age	.096
	Balance	Grade Age Grade x Age	.281	Grade Age Grade x Age	.426	Bilingual [no sign. comparison] Grade Age Bilingual x Age Grade x Age	.216	Bilingual [no sign. comparisons] Bilingual x Age	.071

N.B.: See next page

Table 3.5
ANCOVAs on the Phonological Awareness Data for each Covariate

	Rhyming		**Phoneme blending**		**Onset phoneme identification**		**Phoneme deletion**	
	Significant effects	R^2	Significant effects	R^2	Significant effects	R^2	Significant effects	R^2
Base model	Grade Age Grade x Age	.273	Grade Age Grade x Age	.435	Bilingual [no sign. comparisons] Grade Age Bilingual x Age Grade x Age	.220	Bilingual [no sign. comparisons] Age Bilingual x Age	.073
Working memory (WM)	Grade Age Grade x Age WM	.292	Grade Age Grade x Age	.434	Grade Age Grade x Age	.212	WM	.141
Covariate								
Short-term memory (STM)	Grade Age Grade x Age STM	.295	Grade Age Grade x Age STM	.437	Bilingual [no sign. comparison] Grade Age Bilingual x Age Grade x Age STM	.266	Bilingual [no sign. comparisons] Bilingual x Age	.081

N.B.: See previous page

3.3.2 Multilevel analyses for mainstream and early-English pupils

The next step was to analyse the data of the early-English and mainstream pupils only, without the bilingual children. Since these pupils came from a selection of schools (five mainstream schools, four Early-English schools), it may be that school-specific characteristics have an influence on the results we found. In order to control for possible school effects, we performed linear mixed-effects model analyses for each of the phonological awareness measures in which we entered School as a random factor. Type of education (mainstream or Early-English), Grade, and the interaction between the two were added as fixed factors, and Age as a covariate, the same variables that were investigated in the ANCOVAs. We checked whether there were any significant two-way interactions between Age and the fixed factors. When that was not the case, the interactions were removed from the model. Table 3.5 shows the results. The random effect of School defines the amount of variance assigned to schools (reflecting the differences between the schools) as part of the total random variance. We used the Log Likelihood (*-2LL*) deviance score to compare the models (Heck, Thomas, & Tabata, 2014). The aim was to find the model that was the most parsimonious: a model that could explain the variance in the phonological awareness measures, with the least degrees of freedom. A significance level of .05 was used.

For Rhyming and Phoneme Blending, there were again significant effects of Grade, Age, and the interaction between the two. These results did not differ from the results in the ANCOVA. For Onset Phoneme, there was only a significant effect of Grade. Contrary to the ANCOVA, there were no main effects of or interaction effects with Bilingual Category. For Phoneme Deletion, none of the effects in the base model were significant. Unlike the results of the ANCOVA, Bilingual Category and Age were not significant. In conclusion, for mainstream and early-English pupils, age and grade are important predictors of phonological awareness skills.

Next, we again added each covariate to the base model, to examine whether that changed the pattern of results. For Rhyming, that was not the case. There was a significant main effect of Dutch vocabulary, and an interaction effect between Dutch vocabulary and Age. The model with Dutch vocabulary did not have a significantly better fit than the base model. There were also significant main effects of Working Memory and Short-Term memory, but only the model with Short-term Memory had a significantly better fit than the base model. The effect of English vocabulary was only marginally significant (p = .051), and removing it from the model resulted in a model with a better model fit in terms of fewer degrees of freedom and a lower *-2LL*. Including Balance in the model did not result in significant effects, nor in an improvement in model fit.

The results for Phoneme Blending were very similar to the results for Rhyming: there were significant main effects of Dutch Vocabulary (but no interaction this time), Working Memory, and Short-Term Memory. Only adding Short-Term Memory to the base model improved the model fit. All other covariates were non-significant and did not improve the model. These results are similar to the results found in the ANCOVA, except for the effects of Working Memory, which were not significant in the ANCOVA.

For Onset Phoneme Identification, Grade was the only significant predictor, regardless of which covariate was added to the model, with the exception of Short-

Term Memory, which was significantly related to Onset Phoneme Identification. Unlike in the ANCOVA, Short-Term Memory also interacted with Age. Adding the covariates did not improve the model, nor did it alter the relation between bilingualism and phonological awareness skills. Contrary to the ANCOVA, there were never any main or interaction effects of/with Bilingual Category.

For Phoneme deletion, adding Dutch vocabulary or Working Memory to the model resulted in significant effects. Contrary to the results of the ANCOVA, adding Short-Term Memory to the model also resulted in a significant effect, and the interactions between Short-Term Memory and Age, and Dutch vocabulary and Age were significant as well. The effect of Bilingual Category was never significant. Only the model with Working Memory showed a significant improvement in fit compared to the base model.

In conclusion, the mixed-effects model analyses largely reflected the outcomes of the ANCOVAs, except for the effects of bilingualism on onset phoneme identification and phoneme deletion, which were significant in the ANCOVAs but not in the linear mixed-effects models. All linear mixed-effects models with Phoneme deletion as the dependent variable gave a relatively large variance component for School, suggesting that the effect of Bilingual Category found in the ANCOVAs might in fact reflect accidental differences between the selected schools.

Table 3.6
Results of the Multilevel Analyses for Each of the Four Phonological Awareness Measures

		Rhyming				Phoneme blending				Onset phoneme identification				Phoneme deletion			
		df	Significant effects	*-2LL*	Effect of School	*df*	Significant effects	*-2LL*	Effect of school	*df*	Significant effects	*-2LL*	Effect of school	*df*	Significant effects	*-2LL*	Effect of school
	Base model	13	Grade Age Grade x Age	629.9	0/ 3.695	17	Grade Age Grade x Age	1088.7	0.17/ 4.903	11	Grade	695.7	0.25/ 3.09	3	No significant effects	487.2	1.62/ 10.359
Co-variate	Dutch vocab.	15	Grade Age Grade x Age Dutch Dutch x Age	635.9	0/ 3.478	18	Grade Age Grade x Age Dutch	1089.4	0.12/ 4.823	12	Grade	701.4	0.22/ 3.10	6	Age Dutch Age x Dutch	487.9	2.23/ 9.107
	English vocab.	14	Grade Age Grade x Age	632.7	0/ 3.624	18	Grade Age Grade x Age	1093.8	0.16/ 4.889	12	Grade	702.4	0.26/ 3.09	4	No significant effects	490.3	1.63/ 10.150
	Balance	14	Grade Age Grade x Age	627.2	0/ 3.634	18	Grade Age Grade x Age	1089.1	0.17/ 4.920	12	Grade	696.4	0.25/ 3.11	4	No significant effects	485.95	1.65/ 10.45

N.B.: See next page

Table 3.6
Results of the Multilevel Analyses for Each of the Four Phonological Awareness Measures

		Rhyming				Phoneme blending				Onset phoneme identification				Phoneme deletion			
		df	Significant effects	*-2LL*	Effect of School	*df*	Significant effects	*-2LL*	Effect of school	*df*	Significant effects	*-2LL*	Effect of school	*df*	Significant effects	*-2LL*	Effect of school
	Base model	13	Grade Age Grade x Age	629.9	0/ 3.695	17	Grade Age Grade x Age	1088.7	0.17/ 4.903	11	Grade	695.7	0.25/ 3.09	3	No significant effects	487.2	1.62/ 10.359
Co-variate	Working memory (WM)	14	Grade Age Grade x Age WM	626.5	0/ 3.695	20	Grade Age Grade x Age WM WM x Grade	1081.70	0.17/ 4.605	12	Grade	698.6	0.27/ 3.07	4	WM	480.8	1.46/ 9.452
	Short-term memory (STM)	14	Grade Age Grade x Age STM	624.3	0/ 3.521	18	Grade Age Grade x Age STM	1081.35	0.17/ 4.719	12	Grade STM	698.6	0.27/ 3.07	6	Age STM STMxAge	485.4	1.29/ 10.0

N.B.: See previous page

3.4 Discussion and conclusion

This study had three aims. It aimed to investigate, first, whether bilingualism (positively or negatively) affects the development of phonological awareness, second, how much or how little bilingual experience is needed in order for these developmental differences to occur, and third, whether the relation between bilingualism and phonological awareness is affected by individual differences in children's linguistic and cognitive development. Participants were native Dutch pupils from mainstream Dutch primary schools and early-English schools, as well as Dutch-English bilingual children. We found that differences in phonological awareness between the three groups 1) were restricted to certain tasks, 2) were small when they appeared, 3) did not obtain for children in all grades, and 4) occasionally disappeared when taking individual differences and school effects into account.

The majority of previous studies have shown positive effects of bilingual experience on various tasks of phonological awareness (see Table in Appendix A). Our first hypothesis was therefore that bilingual experience would help the development of phonological awareness. Previous research suggested that the amount of L2 exposure is positively related to the development of phonological awareness. Relating to our first hypothesis, our second hypothesis therefore was that early-English pupils would have more advanced phonological awareness skills than mainstream pupils, but not as advanced as simultaneous bilinguals. Both our first and second hypothesis have to be rejected. For rhyming and phoneme blending, we found no differences between groups at all. For onset phoneme identification and phoneme deletion, there was a small main effect of bilingualism, but pairwise comparisons revealed no significant differences between the three groups. When analysing the data of the mainstream and early-English pupils in a linear mixed-effects model, thereby taking into account accidental differences between schools, the effect of bilingualism was no longer significant. School differences were relatively large, especially for phoneme deletion. One possible explanation for this finding might be that early-English pupils coincidentally attended schools where phonological awareness skills are more promoted than at the schools attended by the mainstream pupils. Unfortunately, we do not have any further information about phonological awareness instruction or tasks at these schools and so we cannot say whether this is indeed the case. For now, for our sample of schools, we have to conclude that there is no significant relation between bilingualism and phonological awareness.

In our main analyses, we found interaction effects of bilingualism and age, both for onset phoneme identification and phoneme deletion. For all three groups there was a positive relation between onset phoneme identification and age, such that older pupils had better scores on this task. This relation was stronger for simultaneous bilinguals than for mainstream and early-English pupils. For simultaneous bilinguals, there was also a positive relation between age and scores on the phoneme deletion task. For mainstream and early-English pupils, no such relation existed. These findings thus show that for simultaneous bilinguals, age is a positive predictor of phonological awareness scores, even after having attended primary school for two or three years. It has previously been suggested that age plays an important role in the development of phonological awareness for monolingual and bilingual children (Branum-Martin et al., 2012; Carroll et al.,

2003). Our findings are also in line with the findings of Janssen et al. (2017), who found that Dutch-Turkish bilingual children's age correlated positively with their scores on an onset phoneme identification task, but that for monolingual Dutch children this was not the case. It may be that age is a proxy for length of exposure to Dutch. Consequently, because bilingual children on the whole have less exposure to Dutch than their monolingual peers, the effect of exposure may pertain longer in the bilingual group than in the native Dutch children.

Based on previous findings (Bruck & Genesee, 1995; Reder et al., 2013), we expected to find an effect of bilingualism especially in the first two years of primary school, but not in the third, when children learn to read and to write. Contrary to our expectations, however, we only found a small effect of bilingualism for the two phonological awareness tasks performed by the pupils in grade 2 and 3 (who are starting to develop their literacy skills), whereas we found no effect of bilingualism on the two tasks performed by mostly illiterate pupils in grades 1 and 2. This result may seem surprising, but we are not the first to find such an effect: previous studies (Bialystok, Luk, & Kwan, 2005; Bialystok et al., 2003; Chen et al., 2010) also showed effects of bilingualism on phonological awareness skills in children who already started literacy instruction. Unlike those previous studies, however, we did not find any differences between groups of children who differed in bilingual experience.

Several reasons may account for the lack of a positive significant effect of bilingualism. Previous research has suggested that children who learn an L2 profit from either learning a language that is phonologically more complex than their L1 (Chen et al., 2010) or less complex than their L1 (Loizou & Stuart, 2005). Since English and Dutch have comparable levels of phonological complexity (Schepens, 2015), it may be that the children in our study have not profited from learning English and Dutch.

Another possible reason for the lack of an effect of bilingualism may be that the children in this study were not proficient enough in their L2. Previous research (Kang, 2012) has suggested that for phonological awareness to accelerate, children should have a certain level of proficiency in their L2. The early-English pupils in this study had significantly higher English vocabulary scores than their mainstream peers, but their level of proficiency may still have been too low for phonological awareness advantages to be detectable. This explanation seems unlikely, however, given that the bilinguals had significantly higher English vocabulary scores than the mainstream and early-English pupils, and did not differ in their level of Dutch vocabulary. Their proficiency in one of their two languages was high enough, and yet they did not show any systematic phonological awareness advantages either.

A more plausible reason for the absence of an effect of bilingualism can be found in the children's level of literacy in both languages. Unfortunately, we do not know to what extent children were exposed to (Dutch) literacy activities at home, or what their level of literacy was in either language. It may be that children from the mainstream or early-English schools were exposed to more literacy activities at home, and/or had a higher level of literacy than the bilingual children. This may in turn have enhanced their phonological awareness skills (Lerner & Lonigan, 2016). If the groups were indeed not comparable in their level of literacy (activities), this

might have obliterated any effect of bilingualism. Future research should therefore examine to what extent children are exposed to literacy activities at home, and how this exposure may influence the relation between bilingualism and phonological awareness.

Our third hypothesis was that the relation between bilingualism and phonological awareness would be affected by taking individual differences in language development and cognitive development into account. More specifically, we expected positive relations between vocabulary development and phonological awareness, and memory development and phonological awareness. This hypothesis was confirmed, but only for Dutch vocabulary and working memory: the main effects of bilingualism on onset phoneme identification and phoneme deletion disappeared when including these covariates. For onset phoneme identification, this happened when including Dutch vocabulary or working memory. For phoneme deletion, this was the case when working memory was included in the model.

Other covariates showed a significant relation with phonological awareness measures, but did not alter the effects found (i.e., those observed before including the covariate). Dutch vocabulary showed a significant effect on rhyming, phoneme blending, and phoneme deletion. This is in line with previous research showing that vocabulary development in the same language as the phonological awareness task is related to scores on the phonological awareness task (Janssen et al., 2017; Sodoro et al., 2002). English vocabulary only showed a significant relation with rhyming. Given that we measured phonological awareness skills in Dutch, and that Dutch is the stronger language for at least early-English and mainstream pupils, this result is not remarkable. If pupils have a low level of proficiency in one language, proficiency in that language may not influence their phonological awareness skills in the other language. Indeed, research with children with Spanish as their L1 who were learners of English as an L2 and had limited proficiency in that language showed that cross-linguistic transfer between Spanish and English phonological awareness took place, but this transfer only occurred in children who were more proficient in Spanish (Atwill et al., 2007). In addition, for proficient L1 speakers, Spanish vocabulary scores correlated with performance on the English phonological awareness task, whereas this was not the case for low-proficient Spanish speakers (Atwill et al., 2007). For Dutch native pupils, it may be the case that when measuring phonological awareness in English, Dutch plays a larger role in performance on the English task.

To reflect the observation that bilingualism is a continuous rather than a categorical variable (Luk & Bialystok, 2013), we also included lexical balance as a covariate. Previous research has shown that children's balance between their proficiency level in their two languages, as a continuous measure of bilingualism, is related to executive functioning skills (Blom, Küntay, Messer, Verhagen, & Leseman, 2014; Goriot et al., 2018). In Goriot et al. (2018) language balance was positively related to children's executive functioning performance, whereas there were no significant group differences between functionally monolinguals and L2 learners. These findings led us to investigate whether a similar relation exists between language balance and the development of phonological awareness, and whether individual differences in language balance may be a more important predictor of phonological awareness skills than groups based on children's bilingual

experiences. Contrary to our expectations, however, language balance operationalised at the lexical level was not significantly related to phonological awareness in this study. The reason for this may be that the nature of the mechanisms that seem to be related to the development of executive functioning and phonological awareness in bilinguals, are different. For executive functions, it has been hypothesized that managing ongoing linguistic competition between two languages places demands on the executive function system, which enhances the development of this system (Green & Abutalebi, 2013). Since bilingual children's phonological awareness development seems not to be related to competition in language activation, it may be that lexical balance does not play a role in the development of phonological awareness.

In addition to language measures, we included measures of short-term and working memory. Previous research has shown that the development of memory is related to the development of phonological awareness (Janssen et al., 2017). In our study, bilinguals performed better on the short-term memory task than mainstream pupils, whereas early-English pupils' scores did not differ from either of the two other groups. This finding is in line with some previous studies which have suggested that bilinguals show developmental advantages in memory, relative to their monolingual peers (Barac, Bialystok, Castro, & Sanchez, 2014). Short-term memory performance showed a significant and positive relation with rhyming, phoneme blending, and onset phoneme identification. We also assessed pupils' working memory development. We found no differences between the three groups on this variable, which is in line with some findings from previous research that also showed no differences between bilingual and monolingual children in working memory (for an overview, see Barac et al., 2014). Despite the absence of group differences, working memory was significantly and positively related to rhyming and phoneme deletion. In the latter case, it was even the only variable in the model that was significant. This shows that working memory plays a large role in performance on a phoneme deletion task. Future research should therefore take working memory skills into account when assessing phonological awareness, and especially when assessing phoneme deletion skills.

We found significant and large effects of grade and age. This is in line with previous research, showing effects of these two factors (Chen et al., 2004, and Janssen et al., 2017, respectively). Clear objectives are formulated for the Dutch educational system about what children in a certain grade should know with respect to phonological awareness (SLO, 2006), and teachers usually pay a great deal of attention to these skills. Children in grade 1 (first year of kindergarten), for example, should be able to separate both words in a compound word, whereas children in grade 2 (second year of kindergarten) should be able to identify the different sounds in one word. It is thus not surprising that children in higher grades perform better than children in lower grades. We also found an interaction effect between grade and age. Grade and age are obviously strongly related to each other, but there is variation in ages within groups. In the Dutch school system, pupils enter primary school as soon as they turn 4, irrespective of the time of year in which their birthday takes place. All 4-year-olds were thus in grade 1, but some of them had been attending school for a longer period than others at the time of testing and therefore older pupils may have profited more from phonological awareness instruction and/or may

have matured more than younger pupils. Consequently, age is, over and above grade, an important variable to take into account when assessing phonological awareness skills, especially in young children.

To conclude, in this study we did not find any convincing effect of bilingualism on phonological awareness. Where we found a positive effect of bilingualism, it was small, and the post-hoc pairwise comparisons between groups showed no significant differences between (functionally) monolingual pupils, L2 learners, and bilinguals. At the same time, however, our study shows that learning two languages has no *negative* effect on phonological awareness skills either. Many parents and teachers have concerns that learning two languages at the same time may be detrimental to pupils' development in at least one of those languages (Goorhuis-Brouwer & de Bot, 2010), and previous research has shown that bilingual pupils can perform less well than their monolingual peers on phonological awareness tasks (Dodd et al., 2010; Janssen et al., 2015; Janssen et al., 2017). We, however, did not find any trace of a negative effect of bilingualism, neither for bilingual Dutch-English children nor for Dutch children learning English as an L2 from a young age. This study shows that children who are learning two languages at the same time have equal phonological awareness skills in the school language as children for whom that school language is the only language they know.

3.5 References

Alloway, T. P., Gathercole, S. E., Kirkwood, H., & Elliott, J. (2008). Evaluating the validity of the Automated Working Memory Assessment. *Educational Psychology*, *28*, 725–734. http://doi.org/10.1080/01443410802243828

Anthony, J. L., & Francis, D. J. (2005). Development of phonological awareness. *Current Directions in Psychological Science*, *14*(5), 255–259. http://doi.org/10.1111/j.0963-7214.2005.00376.x

Atwill, K., Blanchard, J., Gorin, J. S., & Burstein, K. (2007). Receptive vocabulary and cross-language transfer of phonemic awareness in kindergarten children. *The Journal of Educational Research*, *100*(6), 336–346. http://doi.org/10.3200/JOER.6.336-346

Barac, R., Bialystok, E., Castro, D. C., & Sanchez, M. (2014). The cognitive development of young dual language learners: A critical review. *Early Childhood Research Quarterly*, *29*, 699–714. http://doi.org/10.1016/j.ecresq.2014.02.003

Bialystok, E., Luk, G., & Kwan, E. (2005). Bilingualism, biliteracy, and learning to read: Interactions among languages and writing systems. *Scientific Studies of Reading*, *9*, 43–61. http://doi.org/10.1207/s1532799xssr0901_4

Bialystok, E., Majumder, S., & Martin, M. M. (2003). Developing phonological awareness: Is there a bilingual advantage? *Applied Psycholinguistics*, *24*, 27–44. http://doi.org/10.1017.S014271640300002X

Blom, E., Küntay, A. C., Messer, M., Verhagen, J., & Leseman, P. (2014). The benefits of being bilingual: Working memory in bilingual Turkish-Dutch children. *Journal of Experimental Child Psychology*, *128*, 105–119. http://doi.org/10.1016/j.jecp.2014.06.007

Branum-Martin, L., Tao, S., Garnaat, S., Bunta, F., & Francis, D. J. (2012). Meta-

analysis of bilingual phonological awareness: Language, age, and psycholinguistic grain size. *Journal of Educational Psychology*, *104*, 932–944. http://doi.org/10.1037/a0027755

Bruck, M., & Genesee, F. H. (1995). Phonological awareness in young second language learners. *Journal of Child Language*, *22*, 307–324. http://doi.org/10.1017/S0305000900009806

Carroll, J. M., Snowling, M. J., Stevenson, J., & Hulme, C. (2003). The development of phonological awareness in preschool children. *Developmental Psychology*, *39*, 913–923. http://doi.org/10.1037/0012-1649.39.5.913

Chen, X., Wu, X., & Shu, H. (2004). Phonological awareness of bilingual and monolingual Chinese children. *Journal of Educational Psychology*, *96*, 142–151. http://doi.org/10.1037/0022-0663.96.1.142

Chen, X., Xu, F., Nguyen, T.-K., Hong, G., & Wang, Y. (2010). Effects of cross-language transfer on first-language phonological awareness and literacy skills in Chinese children receiving English instruction. *Journal of Educational Psychology*, *102*, 712–728. http://doi.org/10.1037/a0018802

Dunn, L. M., & Dunn, D. M. (2007). *PPVT-4 Manual*. Bloomington: NCS Pearson.

Dunn, L. M., Dunn, L. M., & Schlichting, L. (2005). *Peabody Picture Vocabulary Test-III-NL* (2nd ed.). Amsterdam: Pearson Assessment and Information.

Enever, J., Krikhaar, E., Lindgren, E., Lopriore, L., Lundberg, G., Mihaljevic Djigunovic, J., … Tragant Mestres, E. (2011). *ELLiE: Early Language Learning in Europe*.

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*, *14*, 289–302. http://doi.org/10.1177/1367006910367846

Goriot, C., Broersma, M., McQueen, J. M., Unsworth, S., & Van Hout, R. (2018). Language balance and switching ability in children acquiring English as a second language. *Journal of Experimental Child Psychology*, *173*, 168–186.

Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology*, *25*, 515–530. http://doi.org/10.1080/20445911.2013.796377

Heck, R. H., Thomas, S. L., & Tabata, L. N. (2014). *Multilevel and Longitudinal Modeling with IBM SPSS* (2nd ed.). New York: Taylor & Francis.

Janssen, C., Segers, E., McQueen, J. M., & Verhoeven, L. (2015). Lexical specificity training effects in second language learners. *Language Learning*, *65*, 358–389. http://doi.org/10.1111/lang.12102

Janssen, C., Segers, E., McQueen, J. M., & Verhoeven, L. (2017). Transfer from implicit to explicit phonological abilities in first and second language learners. *Bilingualism: Language and Cognition*, *20*, 795–812. http://doi.org/10.1017/S1366728916000523

Jenniskens, T., Leest, B., Wolbers, M., Bruggink, M., Dood, C., & Krikhaar, E. (2017). *Zicht op vroeg vreemdetalenonderwijs [Insight into early foreign language education]*. Nijmegen.

Kang, J. Y. (2012). Do bilingual children possess better phonological awareness? Investigation of Korean monolingual and Korean-English bilingual children. *Reading and Writing*, *25*, 411–431. http://doi.org/10.1007/s11145-010-9277-4

Kuo, L.-J., & Anderson, R. C. (2010). Beyond cross-language transfer: Reconceptualizing the impact of early bilingualism on phonological awareness. *Scientific Studies of Reading*, *14*, 365–385. http://doi.org/10.1080/10888431003623470

Kuo, L.-J., Uchikoshi, Y., Kim, T.-J., & Yang, X. (2016). Bilingualism and phonological awareness: Re-examining theories of cross-language transfer and structural sensitivity. *Contemporary Educational Psychology*, *46*, 1–9. http://doi.org/10.1016/J.CEDPSYCH.2016.03.002

Lerner, M. D., & Lonigan, C. J. (2016). Bidirectional relations between phonological awareness and letter knowledge in preschool revisited: A growth curve analysis of the relation between two code-related skills. *Journal of Experimental Child Psychology*, *144*, 166–183. http://doi.org/10.1016/j.jecp.2015.09.023

Lesniak, A., Myers, L., & Dodd, B. (2014). The English phonological awareness skills of 5;0–6;0-year-old Polish–English, Portuguese–English bilingual speakers and English monolingual children. *Speech, Language and Hearing*, *17*, 37–48. http://doi.org/10.1179/2050572813Y.0000000029

Loizou, M., & Stuart, M. (2003). Phonological awareness in monolingual and bilingual English and Greek five-year-olds. *Journal of Research in Reading*, *26*, 3–18. http://doi.org/10.1111/1467-9817.261002

Luk, G., & Bialystok, E. (2013). Bilingualism is not a categorical variable: Interaction between language proficiency and usage. *Journal of Cognitive Psychology*, *25*, 605–621. http://doi.org/10.1080/20445911.2013.795574

Marinova-Todd, S. H., Zhao, J., & Bernhardt, M. (2010). Phonological awareness skills in the two languages of Mandarin–English bilingual children. *Clinical Linguistics & Phonetics*, *24*, 387–400. http://doi.org/10.3109/02699200903532508

Reder, F., Marec-Breton, N., Gombert, J.-E., & Demont, E. (2013). Second-language learners' advantage in metalinguistic awareness: A question of languages' characteristics. *British Journal of Educational Psychology*, *83*, 686–702. http://doi.org/10.1111/bjep.12003

Rubin, H., & Turner, A. (1989). Linguistic awareness skills in grade one children in a French immersion setting. *Reading and Writing*, *1*, 73–86. http://doi.org/10.1007/BF00178839

Schepens, J. (2015). *Bridging linguistic gaps: The effects of linguistic distance on the adult learnability of Dutch as an additional language*. Radboud University. Retrieved from http://www.lotschool.nl

SLO. (2006). *TULE-Nederlands kerndoel 11 [TULE-Dutch objective 11]*. Retrieved from http://tule.slo.nl/Nederlands/bestand/P-L11.pdf

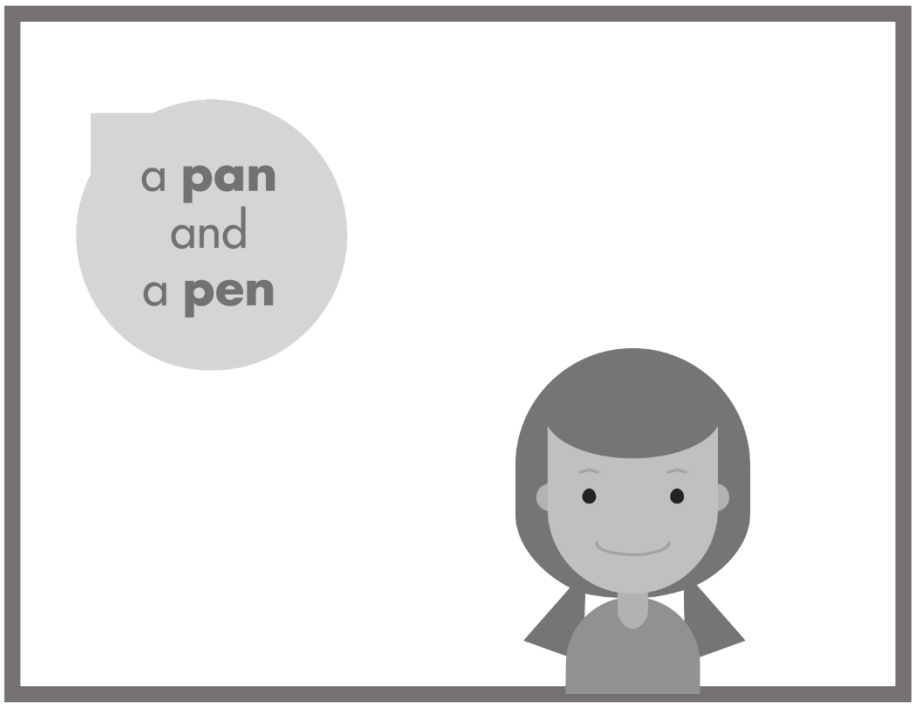
Sodoro, J., Allinder, R. M., & Rankin-Erickson, J. L. (2002). Assessment of phonological awareness: Review of methods and tools. *Educational Psychology Review*, *14*, 223–260. http://doi.org/10.1023/A:1016050412323

Thijs, A., Trimbos, B., Tuin, D., Bodde, M., & de Graaff, R. (2011). *Engels in het basisonderwijs [English in primary education]*. Enschede.

Verhoeven, L. (2005). *Screeningstest voor Taal- en Leesproblemen [Diagnostic Test for Language and Literacy Problems]*. Arnhem: Cito.

Vloedgraven, J., Keuning, J., & Verhoeven, L. (2009). *Screeningsinstrument*

beginnende geletterdheid [Screening instrument for emerging literacy]. Arnhem, the Netherlands: Cito.



Chapter 4: Perception of English phonetic contrasts by Dutch children: How bilingual are early-English pupils?

This chapter is based on:
Goriot, C., McQueen, J. M., Unsworth, S., van Hout, R., & Broersma, M. (2018). Perception of English phonetic contrasts by Dutch children: How bilingual are early-English pupils? Manuscript submitted for publication.

Chapter 4

Perception of English phonetic contrasts by Dutch children: How bilingual are early-English pupils?

Abstract

The aim of this study was to investigate whether early-English education benefits the perception of English phonetic contrasts that are known to be perceptually confusable for Dutch native speakers, comparing Dutch pupils who were enrolled in an early-English programme at school from the age of four with pupils in a mainstream programme with English instruction from the age of 11 on the one hand, and English-Dutch early bilingual children on the other hand. Children were 4-5-year-olds (start of primary school), 8-9-year-olds, or 11-12-year-olds (end of primary school). Children were tested on four contrasts that varied in difficulty: /b/-/s/ (easy), /k/-/g/ (intermediate), /f/-/θ/ (difficult), /ɛ/-/æ/ (very difficult). Bilingual children outperformed the two other groups on all contrasts except /b/-/s/. Early-English pupils did not outperform mainstream pupils on any of the contrasts. This shows that early-English education as it is currently implemented is not beneficial for pupils' perception of non-native contrasts.

4.1 Introduction

When it comes to second language (L2) sound perception, it is often claimed that 'earlier is better' (see Singleton & Ryan, 2004 for a review): becoming able to distinguish certain speech sounds that occur in a non-native but not the native language can be very difficult at a later age. In the Netherlands, for example, the premise of 'the earlier the better' was why the Dutch Education Council advised the Ministry of Education to lower the starting age of foreign-language education, from the age of ten to - preferably - the age of four (Onderwijsraad, 2008). The Council reasoned that older children would be hindered by their first language (L1) when learning a new language, while younger children would not have fully developed their L1 yet and hence could learn a new language more easily and with greater success than older children (Onderwijsraad, 2008). Indeed, just like in many other countries in Europe (Enever, 2013), a growing number of Dutch primary schools now provide early-English education, often from the moment children enter primary school (Nuffic, 2017).

The question that is addressed in the current study is whether this early foreign-language education has a positive influence on children's L2 speech perception. We investigate whether children who receive this kind of education are indeed better able than mainstream pupils to discriminate English phonetic contrasts that are known to be difficult to acquire for Dutch L1 speakers. In addition, we examine whether they can do so as well as bilingual children, that is children who are growing up with both Dutch and English as their native languages.

4.1.1 Foreign speech perception

Initially, infants can perceive phonetic contrasts of all languages, but within the first year of life that ability diminishes while their perception becomes attuned to the L1. Recent evidence shows that infants already have robust knowledge about the

sounds of their L1 by the age of three months (Choi, Cutler, & Broersma, 2017). Around the age of six months old, infants become more sensitive to contrasts in their native language(s), and lose their ability to perceive some of the vowel contrasts of foreign languages (Kuhl, 2004). For consonants, this happens by the time infants are eleven months old (Kuhl, 2004; Werker & Tees, 1984). Infants seemingly effortlessly learn the phonetic contrasts of the language they hear (Kuhl, 2004), and if they receive input in two languages, they will learn the contrasts of both languages during infancy (Bosch & Sebastián-Gallés, 2003b, 2003a; Burns, Yoshida, Hill, & Werker, 2007; Sebastián-Gallés & Bosch, 2009; Sundara & Polka, 2008). This does however not mean that infants immediately categorize all sounds of their native language(s) correctly. Research with monolingual English children has shown that five-year-olds are more likely than nine-year-olds to categorize a foreign vowel as a native one, and that both groups are not as consistent in categorizing native and non-native vowels as adults (Walley & Flege, 1999). As they grow older, monolingual children become increasingly consistent in phonemic categorization, but even at the age of 12 their performance is not adult-like (Hazan & Barrett, 2000).

It is much more difficult to learn to perceive the speech contrasts of a foreign language at a later age (for reviews see Bohn & Munro, 2007; Cutler, 2012; Strange, 1995). For example, previous research has shown that, notoriously, adult Japanese learners of English have difficulty distinguishing between /r/ and /l/ (Aoyama, Flege, Guion, Akahane-Yamada, & Yamada, 2004; Flege, Takagi, & Mann, 1996; Goto, 1971), and Dutch learners of English, despite high proficiency in English as an L2, are not as good as native English speakers in perceiving the difference between /ɛ/ and /æ/ (Broersma, 2005).

Children are better able than adults to learn foreign speech sounds (Aoyama et al., 2004; Baker, Trofimovich, Flege, Mack, & Halter, 2008; Flege, MacKay, & Meador, 1999; Tsukada et al., 2005). For example, Tsukada et al. (2005) compared English speech perception abilities of Korean adults and children between 9 and 17 years old, who had lived in America for either 3 or 5 years. After 3 years of residence, the children's discrimination abilities were already significantly better than the adults'. Adults' ability to discriminate between English contrasts did not differ between the groups that had been in America for a shorter or longer time. Children who had resided in America for 5 years were significantly better able to discriminate English speech contrasts than those who been in America for 3 years, however, resulting in an even greater difference between the adult and child group when comparing the groups who had resided in America for five years.

As the children in the previous study had extensive exposure to the new language, the question remains whether children are also able to learn phonetic contrasts of a foreign language that they learn in a less naturalistic setting, such as in school. McCarthy, Mahon, Rosen, and Evans (2014) investigated four-year-old Sylheti-English sequential bilingual children's perception of the English voicing contrast in plosives when they were in kindergarten and again one year later. Before entering kindergarten, children were exposed to English less than 20% of the time. In kindergarten the sequential bilingual children were not as proficient as their English monolingual peers in differentiating between voiced and voiceless plosives. After one year, however, their performance had significantly improved and matched that of the monolingual group. In contrast, a study with 11-year-old Turkish-German

sequential bilinguals who had been immersed in a German environment from before the age of four showed that their perception of German vowels was significantly poorer than that of German monolingual children (Darcy & Krüger, 2012). Unlike the bilingual children in the McCarthy et al. (2005) study, who were educated in a monolingual programme, the Turkish-German children were educated in a German-Turkish bilingual programme. They thus received less input in their L2 than the Sylheti-English children, but were exposed to it for a longer time.

In both studies (Darcy & Krüger, 2012; McCarthy et al., 2014), children were immersed in their L2 at school, either 50 or 100% of the time. Such extensive L2 exposure is much more than the amount of exposure offered in foreign language programmes at school, such as the ones gaining in popularity in most countries in continental Europe, including the Netherlands (Enever, 2013). Nowadays, almost one in five Dutch schools start their English lessons from the moment children enter primary school at the age of four. The obligatory onset of English education, however, is not until the penultimate year, when children are approximately ten years old (Nuffic, 2017). As children in the Netherlands receive on average 60 minutes of English lessons per week, early-English pupils leave primary school after 320 hours of English education, as opposed to 60 hours (45 minutes per week) for pupils who start in the penultimate grade (at age ten) (Jenniskens et al., 2017). The idea behind early-English education is that it will benefit pupils' English proficiency more than mainstream English lessons, but not at the expense of Dutch. Previous research in the Netherlands has indeed shown beneficial effects of early-English programmes for pupils' English vocabulary (Goriot, Broersma, McQueen, Unsworth, & Van Hout, 2018; Unsworth, Persson, Prins, & de Bot, 2015; van der Leij, Bekebrede, & Kotterink, 2010), grammar (Unsworth et al., 2015), reading and writing abilities (de Graaff, 2015), and word pronunciation (Lobo, 2013). Research on Dutch early-English pupils' abilities to perceive English speech contrasts, however, is not yet available.

International research on early-English pupils' perception of English contrasts is scarce, too. Jost et al. (2015) investigated Swiss-German children's responses to the English contrast /t/-/θ/: once before starting English lessons, and once a year later, after having received one and a half hours of English lessons per week. Children showed a significant increase in accuracy in discriminating between /t/ and /θ/ after one year of English instruction.

With respect to the Dutch situation, early-English pupils necessarily encounter unfamiliar sound contrasts that they have to learn to distinguish. Nevertheless, during English lessons not much attention seems to be given to the perception of these unfamiliar speech sounds. The goals for English education focus on understanding spoken and written English texts, and being able to (and being confident to) communicate in English (Bodde, Schippers, Klein Tank, & van der Linde-Meijerink, 2008). Consequently, the goal of English lessons in Dutch education is mainly oral proficiency, and especially speaking, listening, and vocabulary development (Thijs et al., 2011). Activities for young pupils are aimed at learning English in a playful way, for example by the means of songs and stories. For older pupils, from grade five onwards (8-9-year-olds), content learning and language learning are integrated and English is used as the language of communication during other subject lessons. In practice, teachers speak English

during those lessons, but pupils hardly speak any English to each other or the teacher (Geurts & Hemker, 2013). This raises the question whether the learning activities that pupils in early-English education take part in contribute to the advancement of their perception of L2 speech contrasts. In this study, we investigate whether Dutch children who are exposed to English from a young age in an educational setting are better able to distinguish such English speech contrasts that do not exist in their native language than children in a mainstream program, or not.

4.1.2 Contrasts of interest

Dutch and English phoneme categories differ in various ways. According to the Perceptual Assimilation Model (PAM), the degree to which L2 learners should be able to learn to perceive L2 contrasts depends on the way in which they differ from the L1 phonetic system (Best, 1993, 1995; Best & Tyler, 2007). The focus of this study is on four English phonetic contrasts that are expected to vary in difficulty for native speakers of Dutch, ranging from very easy to very difficult.

In the first contrast of interest, /b/-/s/, the phonemes differ from each other in place, voicing and manner of articulation. Both English and Dutch have a /b/ and /s/ that are pronounced rather similarly in both languages. (Differences include the VOT of /b/, which is negative in Dutch and around 0 for English). In the PAM, such a contrast is called a 'two category' assimilation contrast (Best & Tyler, 2007). According to the predictions of the PAM, when two non-native phonemes are very similar to two phonemes in the native language, listeners will perceptually assimilate these non-native phonemes to the corresponding native categories. Consequently, pupils should be able to easily discriminate between members of such a contrast. The /b/-/s/ pair was thus included in this study as an easy contrast. At the same time, it served as a control contrast, to ensure that children understood the task.

Second, the /k/-/g/ contrast was expected to be of intermediate difficulty. Dutch as well as English has /k/ and both are pronounced rather similarly (but in English with a long positive VOT, and in Dutch with a 0 to short positive VOT). Contrary to English, Dutch does not have a /g/. According to the PAM, the English /k/-/g/ contrast would be a 'category goodness' contrast, where two different L2 phonemes are both mapped onto the same L1 category but one is considered to be a better exemplar of that category than the other. According to the PAM, recognition of an L2 phoneme should be more difficult when two phonemes are mapped onto the same L1 category than when both phonemes are mapped onto two different L1 categories (Best, 1993, 1995; Best & Tyler, 2007). Previous research has shown that, indeed, native Dutch adults find it difficult to identify the English /g/ (Cutler, Weber, Smits, & Cooper, 2004). Further, although most Dutch native adults are able to produce word-initial /g/, substitution with [k] or [x] also occurs (Hamann & De Jonge, 2015; van Bezooijen & Gerritsen, 1994). There are, however, three reasons to expect that the English /k/-/g/ contrast might be easier for Dutch listeners to differentiate than other category goodness contrasts. First, in Dutch, whereas there is no velar voicing contrast (/k/-/g/), bilabial (/p/-/b/) and alveolar plosives (/t/-/d/) are contrasted in voicing. In other words, /g/ is missing from a consonant system which does contain /b/ and /d/, and hence there is a *system gap* (Gussenhoven & Jacobs, 2011). Second, the /g/ does occur in Dutch in loan words (Gussenhoven, 1999; Hamann & De Jonge, 2015; van Bezooijen & Gerritsen, 1994) and, third, it occurs

as an allophone of /k/. The PAM does not explicitly discuss the possible effect of such L1 experience with (features of) sounds that are not present as a phoneme in the L1, and its predictions for category goodness contrasts seem to assume that the L2 learners are completely inexperienced with respect to one of the members of the contrast. This does not hold however for the /g/ in Dutch. The first reason is that Dutch native listeners *do* have experience with the voicing contrast, in other places of articulation. Second, they also have some experience with /g/ specifically, since it occurs in loanwords and as an allophone of /k/. It is therefore expected that Dutch listeners might find it easier to perceive the English /k/-/g/ contrast than a category goodness contrast that they do not have such L1 experience with. Indeed, it has been shown that Dutch listeners tend to identify /g/ more accurately than consonants with features that do not occur at all in Dutch, like /θ/ or /ð/ (Cutler et al., 2004).

The third contrast we included was therefore /f/-/θ/. Just like /k/-/g/, /f/-/θ/ can be considered to be a 'category goodness' contrast, as Dutch has an /f/ fairly similar to the English one, but no /θ/. Previous research has shown that Dutch adults have difficulty identifying the English consonant /θ/. They often perceive /θ/ to be similar to /t/, /s/, or /f/ (Cutler et al., 2004; Hanulíková & Weber, 2012). The predictions of PAM for /k/-/g/ and /f/-/θ/ are similar. We expected the /f/-/θ/ contrast, however, to be more difficult for Dutch learners of English than the /k/-/g/ contrast. First, the lack of /θ/ in Dutch is not a system gap, as Dutch has no dental place of articulation in the consonant system. Furthermore, unlike /g/, /θ/ does not occur in the pronunciation of loanwords or as an allophone. Further the perception of /θ/ is intrinsically difficult; even native speakers of American English found /θ/ to be confusable with /f/ (Cutler et al., 2004; Johnson & Babel, 2010).

The final contrast we included was /ɛ/-/æ/. Whereas the English vowel inventory includes both open midfront unrounded vowels, Dutch only has /ɛ/. The Dutch /ɛ/ lies between the English /ɛ/ and /æ/, and is typically lower than the English /ɛ/. Following the PAM predictions for what is called a 'single category' assimilation contrast, Dutch native speakers should find it very difficult to distinguish between/ɛ/ and /æ/, as shown previously indeed (Broersma, 2005). According to the PAM, both /ɛ/ and /æ/ will be perceived as an exemplar of the Dutch /ɛ/. Given that unlike the /k/-/g/ or /f/-/θ/ contrast, neither member will be a better exemplar of the Dutch category, Dutch learners of English are expected to have the most difficulty distinguishing between both members of this contrast.

Whereas Dutch children learning English may have difficulty distinguishing between the English contrasts mentioned above, that does not imply that they will not be able to learn to distinguish those contrasts at all. Best and Tyler (2007) predict that for category goodness contrasts - when two L2 phonemes map onto the same L1 category while only one L2 phoneme is a good exemplar of the L1 category - learners will eventually form a new category for the other phoneme, and learn to discriminate between the members of the L2 contrast. It may therefore be expected that Dutch children may initially not perceive the difference between /k/ and /g/, or /f/ and /θ/, but that they will learn to distinguish these phonemes after several years of learning English. Our expectation is that Dutch children are able to distinguish between /k/ and /g/ earlier than between /f/ and /θ/, given that they encounter voicing distinctions in the Dutch language too, as well as instances of /g/ in loanwords and as allophones of /k/.

For the single category assimilation contrast, when two L2 phonemes are mapped onto the same L1 category but neither of them is a good exemplar of that L1 category, learners may not overcome their difficulties in discriminating between the two L2 sounds (Best & Tyler, 2007). It is therefore expected that Dutch L2 learners' difficulty to distinguish between /ɛ/ and /æ/ is not limited to the initial stages of L2 acquisition, but may persist even after several years of L2 learning (Broersma, 2005; Cutler et al., 2004).

4.1.3 This study

One of the ideas behind early-English education is that such a programme will benefit pupils' English language skills, while at the same time their Dutch language skills develop just like those of their monolingually educated peers. The main research question of this study was whether early-English pupils are better able to discriminate between English phonemes than pupils from mainstream schools, and whether early-English pupils' performance is comparable to that of children growing up with both English and Dutch as their home languages.

The participants differed in their experience with English. The bilingual children grew up with both English and Dutch as their native languages. The early-English and mainstream pupils were raised in Dutch. Both are expected to be exposed to English via media, as English is very present in Dutch media (Kuppens, 2010). For example, English movies are not dubbed but rather subtitled. For early-English pupils, English lessons started at the start of primary school (i.e., at the age of four). For mainstream pupils, English lessons started near the end of primary school (at the age of ten). Since English lessons generally do not last more than one hour per week, even with out-of-school exposure to English, pupils had much more limited exposure to English than the children that were being raised bilingually. This study examines whether limited L2 exposure in a non-naturalistic setting is related to improved perception of L2 speech contrasts, and if so, how long it takes for such an improvement to be detectable. By investigating to what extent children are able to learn to discriminate non-native speech sounds under such conditions, this study will thus contribute to knowledge about the plasticity of the speech perception system.

We investigated speech perception abilities in three age groups: children who had just started primary school (4-5 year olds), children who were halfway through primary school (8-9 year olds), and children who were at the end of primary school (11-12 year olds). As is often the case, age at testing and amount of L2 exposure are intertwined (Muñoz, 2008); we do not attempt to separate the two from each other nor do we want to make any claims about whether starting at an earlier age is better for phonological learning. Age of onset is the same for all the early-English children participating in this study (i.e., 4 years). The goal of including pupils from different age (at testing) groups was to investigate the effect of limited English instruction after several years. Since early-English pupils generally have one hour of English lessons per week, the 4-5-year-old mainstream and early-English pupils were expected to be comparable in amount of instruction (i.e., none to very limited amount of instruction). After five years, 8-9-year-old early-English pupils would have had an estimated 200 hours of English instruction as opposed to no instruction for their peers in mainstream schools. By the end of primary school, 11-

12-year-old early-English pupils would have had an estimated 320 hours of English education, as opposed to 60 hours for mainstream pupils.

Three research questions were investigated. The first one was whether early-English pupils outperform mainstream pupils on the perception of the non-native contrasts, and whether bilingual children outperform early-English pupils. We hypothesized that the bilingual children should outperform the two other groups, and if anything, the early-English group should outperform the mainstream pupils.

Second, we asked whether older pupils have better speech perception abilities than younger pupils. We hypothesized that on all contrasts and for all three groups of children, older children would perform better than younger children. This was expected because older children generally show better phoneme categorization than younger children (Hazan & Barrett, 2000; Walley & Flege, 1999), and moreover the older groups in this study have had more exposure than the younger groups.

Third, we examined whether the four different speech contrasts investigated here differed in difficulty level for native speakers of Dutch. The hypothesis was that the degree of similarity between the English and Dutch phoneme inventory should predict the perceptual difficulty, especially for the mainstream and early-English children. Following the predictions based on Best and Tyler (2007) discussed earlier, we expected pupils to find the /b/-/s/ contrast relatively easy, the /k/-/g/ contrast more difficult, the /f/-/θ/ contrast even more difficult, and the /ɛ/-/æ/ contrast the most difficult.

4.2 Method

4.2.1 Participants

Three groups of children participated: a control group (N = 48), a group of early-English pupils (N = 64), and a group of bilingual children (N = 48; see also Table 4.1). The Dutch primary school system consists of eight grades. The control group included children attending a mainstream Dutch school in which English lessons started in grade six or seven (when pupils are nine or ten years old; two schools). Early-English pupils attended a school in which English lessons started in the first grade (kindergarten, i.e., when pupils are four years old; three schools). All early-English schools had a certificate from an independent organization that they taught English for at least 60 minutes per week, and that teachers had at least a B2- (high intermediate) level of English (except for writing, for which B1 [low intermediate level] had been deemed sufficient by the certification organization). The schools had been early-English schools for at least eight years, such that the children who were now in the final grade had started their English education when they entered primary school. Schools were recruited via telephone. If they were interested in participating, they received a document with more information about the study. Schools voluntarily agreed to participate.

Bilinguals were children who had at least one parent who was a native speaker of English. All bilingual children had started to learn Dutch before the age of four, for example because they went to Dutch day care. Three bilingual children were exposed to one additional language (beyond Dutch and English), and one child to two additional languages.

Children had no known hearing or developmental disorders. All parents gave informed consent for participation, and were also asked to fill in a questionnaire about out-of-school exposure to English.

4.2.2. Instruments

XAB non-word discrimination task. Participants were presented with an XAB task in which three non-word stimuli were presented. Children had to indicate which of these stimuli (the second or third) matched the first stimulus (X). The XAB task was presented as a game (Zhou, 2015)**,** programmed with Presentation software (version 14.7) from Neurobehavioral Systems. The task consisted of 64 trials, which were administered in two sessions. The first session was preceded by an explanation and six practice trials. The second session was preceded by four practice trials. In the practice trials, a cartoon of one large and two smaller dinosaurs showed the large dinosaur saying a bisyllabic non-word, not containing the contrasts of interest, and subsequently, the two smaller dinosaurs each saying a non-word, one of which was the same as that spoken by the larger dinosaur, and one different. The child was instructed to indicate with a key on the keyboard of the laptop which smaller dinosaur correctly repeated the larger dinosaur. The button 'A' corresponded to the left animal, and the button 'L' to the right one. Buttons were marked with stickers, to remind children of the response buttons. Children received feedback on their performance. If they pressed the wrong button, the dinosaur that corresponded to the correct button started crying. The large dinosaur would say that the child pressed the wrong button and encouraged the child to press the correct button. If children pressed the correct button, the corresponding dinosaur would jump up and down while throwing around confetti. The large dinosaur would also verbally confirm that the correct button was pressed.

In the experimental trials in which the four contrasts were tested, the cartoon of the three animals covered two-thirds of the screen, and a cartoon of a small animal on a staircase covered the other one-third (see Figure 4.1). For the first two blocks of each session the animals were dinosaurs, and for the last two blocks of each session they were pandas. In the experimental trials, children no longer received feedback. As an incentive, once children pressed a button, the animal on the staircase would jump up one step, no matter if the child performed correctly. In both the practice and the experimental trials, it was always the left animal who first imitated the larger animal. Children could press a button after the second small animal had spoken.

Figure 4.1. Screenshot of an experimental trial in the XAB task.

Stimuli. The stimuli consisted of eight minimal pairs of non-words. Every minimal pair was presented eight times during the experiment. The stimuli were disyllabic non-words that were phonotactically legal in English (and Dutch). The consonant pairs /b/-/s/, /k/-/g/ and /f/-/θ/ were presented in two different VCV carrier sequences (/əCi/ and /əCa/). The vowel contrast /ɛ/-/æ/ was presented in two different VCVC carrier sequences (/əpVp/ and /ətVt/).

Recordings of the stimuli were made by three female native speakers of Standard American English, each to represent the voice of one designated animal. For each speaker, each stimulus was recorded four times while they read the stimuli one by one in random order, in a clear citation style. Recordings were made in a soundproof booth. The sampling rate at recording was 44.1 kHz.

The 64 trials were presented in eight blocks of eight items each. Each contrast appeared twice in each block. Trials were presented in pseudo-random order: No more than one trial targeting the same contrast and no more than two trials with the same carriers followed each other, and the same animal did not say the correct answer more than three times in a row. The number of times each of the smaller animals correctly repeated the larger animal was counterbalanced across the two smaller animals within blocks.

Vocabulary. To investigate to what extent children had knowledge of both languages, and to examine to what extent the three type-of-English-acquisition groups differed from each other in their knowledge of these languages, English and Dutch vocabulary were measured. English vocabulary was measured with the PPVT-4 (Dunn & Dunn, 2007) and Dutch vocabulary with the PPVT-III-Dutch (Dunn, Dunn, & Schlichting, 2005). The test-retest reliability coefficients for children aged between 4;0 and 13;0 years are very high, ranging from .91 to .94 for the English version (Dunn & Dunn, 2007), and between .91 and .96 for the Dutch version (Dunn et al., 2005). In both tests, the child is presented with a spoken word

and is asked to indicate the corresponding picture out of a set of four. The English test consists of 228 items, the Dutch one of 204. Items are grouped in sets of 12. The rules for administration as given in the manual were followed. Testing stopped if children made more than the maximum number of errors in one set (eight for the English and nine for the Dutch PPVT). The score was computed as the number of correctly performed items.

Intelligence. To control for possible differences in intelligence, the 'Matrix Reasoning' subtest of the Wechsler Nonverbal Scale of Ability was administered (Wechsler & Naglieri, 2008). The subtest consists of 41 matrices of which one piece is missing. The child is asked to indicate the correct piece amongst four or five alternatives. Testing is stopped when four or five errors are made in five consecutive items. The total number of correct items determines the score.

4.2.3 Procedure

Children were tested individually in a quiet room, either at school (early-English and mainstream pupils) or at home (bilingual children). For the participants tested at school, the tasks were presented in two sessions of 25 minutes each. In the first session, the PPVT-4 was administered first, followed by the first part of the XAB task (4 blocks, 32 trials). Session two started with the second part of the XAB task, followed by Matrix Reasoning and then the PPVT-III-NL. The bilingual children took part in a larger study in which multiple tasks were administered which are not reported here. For them, like for the early-English and mainstream children, the XAB task was always administered at the end of the first session and at the beginning of the second session, the PPVT-4 in the first session, and the PPVT-III-NL in the final session. For all (mainstream, early-English, and bilingual) children except four, the two sessions were administered on two separate days, which were on average 8 days apart (SD = 11; of those four exceptional children, three did both sessions on the same day, and one participated only in the first session). For five bilingual children, the matrix reasoning task was not (or not correctly) administered, for three bilingual children, the PPVT-4, and for two bilingual children the PPVT-III-NL was not (correctly) administered. Scores for these children on these tasks are missing. Responses to all tasks were registered on a laptop (XAB task) and/or noted down by the experimenter (PPVTs and Matrix Reasoning).

4.3 Results

4.3.1 Data screening XAB task and differences in background variables

Each individual participant's response pattern was investigated. We removed data where children either consistently pressed the left or the right button, or pressed the two buttons in strict alternation. First, we removed all the data of children who showed such a pattern throughout the entire task (n = 9; 2 early bilinguals, 4 early-English, and 3 mainstream pupils; all were 4-5-year-olds). Next, we removed data of individual sessions that showed such a pattern: The first session was removed for one participant from a mainstream school, and the second session for three early-English and two mainstream-school pupils, all in the 4-5 year old group. After that, we removed data of individual blocks with such a response pattern. In total, 23 blocks were removed, from seven early-English and nine mainstream-school pupils, one in the 11-12 year old group, two in the 8-9 year old

group, and 13 in the 4-5 year old group. Next, we removed trials on which participants had RTs that were more than 2.5 *SD*s above their own mean RT. In total 13.4% of the data were removed. Next, some pupils had a proportion correct of less than .60 on the perceptually easy /b/-/s/ contrast. As this was taken as an indication that those children either did not understand the task or were not concentrating, all data of those children were removed as well (9 children; eight 4-5-year-olds, one 11-year-old; five boys, four girls, two mainstream pupils, five early-English pupils, two bilingual children). For the remaining participants, average proportions correct (and *SD*s) are shown in Table 4.1.

Information on age, and mean scores on all background measures are shown in Table 4.2. For the bilingual children, the table also contains information on how often (in % of time) bilingual children were spoken to in English at home (as opposed to Dutch), calculated as the percentage of time English was used by parents during the hours parents and children were together. We also calculated how often (again in % of time) this English came from a native speaker parent (as opposed to a non-native speaker).

To assess whether the groups were comparable on all measures besides the variables of interest (i.e., the phonetic contrasts), ANOVAs were done with Age (4-5-, 8-9-, and 11-12-year-olds) and Type of English acquisition (mainstream, early-English, bilingual) as independent variables. We used partial eta-squared as a measure for effect size, and followed the general rule of thumb that effects between .02 and .13 are small, those between .14 and .26 are medium, and values of .26 and higher are large (Cohen, 1988).

For the dependent variable Age, there was a significant interaction effect between Age and Type of English acquisition ($F(4, 146) = 5.28$, $p = .001$, $\eta_p^2 = .126$, $R^2_{\text{model}} = .952$), Tukey post hoc analyses showed that 8-9-year-old bilinguals were on average younger than mainstream and early-English pupils in the same age group, and 11-12-year-old bilinguals were younger than mainstream pupils in the same age group. For English vocabulary, there was a main effect of Type of English acquisition ($F(2,148) = 118.56$, $p < .001$, $\eta_p^2 = .616$, $R^2_{\text{model}} = .739$). Tukey post hoc tests showed that, as to be expected, bilinguals obtained higher scores than mainstream and early-English pupils. Mainstream and early-English pupils did not differ from each other. As expected, all groups were comparable on Dutch vocabulary ($R^2_{\text{model}} = .861$), and intelligence ($R^2_{\text{model}} = .561$; all *p*s > .05).

The groups were also compared on out-of-school exposure to English. Only a limited number of parents completed the questionnaire. Therefore, Age was not included in the analyses, and an ANOVA with Type of English acquisition as the only independent variable was performed. As expected, the weekly number of hours of out-of-school exposure to English did not differ among the groups ($F(2,83) = 0.07$, $p > .05$, $R^2_{\text{model}} = -.022$).

Table 4.1
Percentage Correct on Each of the Contrasts, for Each Type of English acquisition Separately

			4-5-year-olds			8-9-year-olds			11-12-year-olds		
			Mainstream	Early-English	Bilingual	Mainstream	Early-English	Bilingual	Mainstream	Early-English	Bilingual
		N	15	21	12	15	19	18	18	24	18
Contrast (prop. correct)	/b/-/s/	*M*	.81	.84	.89	.96	.98	.96	.99	.98	1.0
		(SD)	(.13)	(.13)	(.11)	(.06)	(.04)	(.06)	(.03)	(.04)	(.00)
	/k/-/g/	*M*	.54	.60	.68	.71	.78	.90	.80	.84	.94
		(SD)	(.14)	(.18)	(.09)	(.12)	(.15)	(.13)	(.13)	(.11)	(.07)
	/θ/-/f/	*M*	.47	.56	.60	.63	.55	.66	.59	.61	.69
		(SD)	(.10)	(.13)	(.12)	(.14)	(.10)	(.14)	(.18)	(.16)	(.17)
	/ɛ/-/æ/	*M*	.57	.53	.68	.52	.55	.71	.51	.47	.81
		(SD)	(.12)	(.12)	(.12)	(.10)	(.05)	(.20)	(.09)	(.08)	(.17)

Table 4.2
Age and Scores on all Measures, for Each Type of English acquisition Separately

		4-5-year-olds			8-9-year-olds			11-12-year-olds		
		Mainstream	Early-English	Bilingual	Mainstream	Early-English	Bilingual	Mainstream	Early-English	Bilingual
English vocabulary (raw score, max. 228)	*N*	15	21	12	15	19	16	18	24	17
	M	20.87	26.57	84.83	76.47	72.00	135.19	114.33	108.71	155.59
	(SD)	(14.64)	(13.04)	(18.38)	(17.76)	(21.51)	(16.79)	(27.33)	(27.54)	(14.41)
Dutch vocabulary (raw score, max. 204)	*N*	15	21	12	15	19	17	18	24	16
	M	69.67	73.52	73.00	113.33	110.16	114.00	132.56	134.79	142.81
	(SD)	(10.63)	(13.56)	(11.72)	(6.70)	(6.47)	(11.10)	(12.20)	(9.84)	(11.92)
Intelligence (raw score, max. 41)	*N*	15	21	11	15	19	16	18	24	15
	M	11.87	11.62	11.36	20.47	19.11	19.56	21.72	22.83	23.33
	(SD)	(4.12)	(2.82)	(3.56)	(3.62)	(3.23)	(3.93)	(5.10)	(4.19)	(5.09)
Age (years)	*N*	15	17	12	15	19	18	18	24	18
	M	4.6	4.9	5.1	9.4	9.2	8.5	12.3	12.1	11.1
	(SD)	(0.5)	(0.3)	(0.4)	(1.0)	(0.5)	(1.0)	(0.5)	(0.5)	(.8)
Home exposure to English media (hours per week)	*N*	3	15	10	4	11	17	4	6	16
	M	4.00	7.93	12.77	5.63	11.82	11.56	29.75	25.46	16.03
	(SD)	(1.00)	(5.75)	(6.33)	(3.47)	(7.97)	(10.64)	(41.62)	(11.33)	(8.99)

N. B.: See next page

Table 4.2
Age and Scores on all Measures, for Each Type of English acquisition Separately

		4-5-year-olds			8-9-year-olds			11-12-year-olds		
		Mainstream	Early-English	Bilingual	Mainstream	Early-English	Bilingual	Mainstream	Early-English	Bilingual
Home exposure to English media (hours per week)	*N*	3	15	10	4	11	17	4	6	16
	M (SD)	4.00 (1.00)	7.93 (5.75)	12.77 (6.33)	5.63 (3.47)	11.82 (7.97)	11.56 (10.64)	29.75 (41.62)	25.46 (11.33)	16.03 (8.99)
% exposure to English at home (from parents, either native or non-native speaker of English)	*N*	-	-	10	-	-	18	-	-	16
	M (SD)	-	-	58.8 (24.3)	-	-	51.4 (14.8)	-	-	53.1 (19.1)
% exposure to English from native speaker parent(s)	*N*	-	-	10	-	-	18	-	-	15
	M (SD)	-	-	84.2 (21.9)	-	-	80.6 (23.7)	-	-	77.3 (23.8)

N. B.: See previous page

4.3.2 Perception of phonetic contrasts

The sensitivity measure *d'* with a correction for near-perfection (MacMillan & Creelman, 1991) was calculated for each contrast and for each participant separately. One member of each contrast pair was taken as the target (/b/, /k/, /f/, and /ɛ/, respectively). In case the child heard this sound and pressed the correct button, that was counted as a 'hit', if the incorrect button was pressed, that was counted as a 'miss'. If the child heard the other sound (/s/, /g/, /θ/, or /æ/, respectively) and she pressed the correct button, this was counted as a 'correct rejection, if the wrong button was pressed, it was counted as a 'false alarm'. The *d'* was then calculated as the difference between z-transformed hits and false alarms. A *d'* of 0 indicates that performance is at chance level, and a higher *d'* value indicates a better ability to discriminate between the two members of the phonetic contrast. Mean values for *d*' are presented in Figure 4.2.

An ANOVA with Contrast, Age and Type of English acquisition as fixed factors, and *d'* scores as dependent variable was performed. There were significant main effects for all fixed factors (Contrast: $F(4,604) = 317.64$, $\eta_p^2 = .612$; Type of English acquisition: $F(2,604) = 54.78$, $\eta_p^2 = .154$; Age: $F(2,604) = 72.59$, $\eta_p^2 = .194$; all ps < .001, $R^2_{\text{model}} = .691$), and significant interactions between Age and Contrast ($F(6,604) = 12.09$, $p < .001$, $\eta_p^2 = .107$), and Type of English acquisition and Contrast ($F(4,604) = 6.32$, $p < .001$, $\eta_p^2 = .059$). The three-way interaction was also significant ($F(12,604) = 1.97$, $p = .025$, $\eta_p^2 = .038$).

To follow up on this three-way interaction, we performed the same analysis but without Age for each of the age groups separately (see Table 4.3). The results of all three analyses showed main effects of Type of English acquisition and Contrast, and for 8-9- and 11-12-year-olds a significant interaction effect. Post hoc tests (Tukey HSD) showed that in all three age groups, all English-acquisition-type groups obtained the highest *d'* scores on the /b/-/s/ contrast, followed by the /k/-/g/ contrast, which in turn had higher scores than the other two contrasts. In both the 4-5- and 8-9-year-olds, scores on the /f/-/θ/ and /ɛ/-/æ/ contrasts did not significantly differ. Significant differences among the English-acquisition-type groups emerged for the 11-12-year-olds for the comparison between the /f/-/θ/ and /ɛ/-/æ/ contrasts: For the 11-12-year-old mainstream pupils, scores on the /f/-/θ/ and /ɛ/-/æ/ contrasts did not significantly differ. For the 11-12-year-old early-English pupils, scores on /f/-/θ/ were significantly higher than for /ɛ/-/æ. For the 11-12-year-old bilinguals, to the contrary, scores on /ɛ/-/æ/ were significantly higher than for /f/-/θ/.

To investigate whether there were effects of bilingualism or Age on the *d'* of the different contrasts, four separate ANOVAs were conducted, one for each contrast. In each of these analyses, Type of English acquisition and Age were fixed factors, and the *d*' on each of the contrasts was the dependent variable. The results are shown in Table 4.4.

There was a main effect of Type of English acquisition for all contrasts except /b/-/s/. Bilinguals outperformed both mainstream and early-English pupils on all contrasts except the easy control contrast /b/-/s/, although for the /ɛ/-/æ/ contrast there was a significant interaction with Age; the post hoc tests revealed that only older (i.e., 8-9- and 11-12-year-old) bilinguals significantly outperformed both other groups, whereas the bilinguals in the 4-5 year old group outperformed only the

early-English and not the mainstream pupils. Early-English pupils never scored significantly different from mainstream pupils.

Further, there was a significant main effect of Age, showing that younger pupils performed significantly less well than older pupils, for all contrasts except /ɛ/-/æ/. On the /k/-/g/ contrast, all three Ages significantly differed from each other. On the /b/-/s/ and /f/-/θ/ contrast the 8-9- and 11-12-year-olds significantly outperformed the 4-5-year-olds, but the former two groups did not significantly differ from each other[1,2]

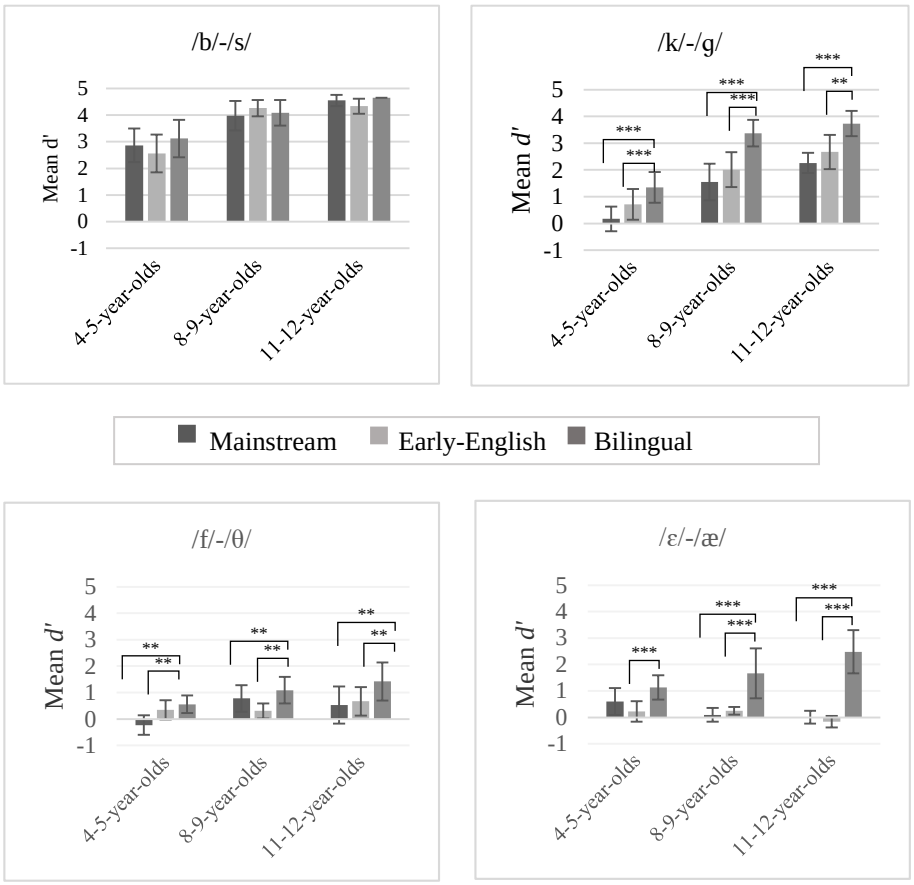


Figure 4.2. Mean values (with *SD*) for *d',* for each group separately. Pairwise significant differences (Tukey post hoc tests) are indicated with $^{*}p < .05$, $^{**}p < .01$, $^{***}p \leq .001$.

Table 4.3
ANOVAs on d' Scores by Contrast (Within-Subject) and Type of English Acquisition (Between-Subject)

	4-5-year-olds				8-9-year-olds				11-12-year-olds			
	df	*F*	η_p^2	Post hoc	*df*	*F*	η_p^2	Post hoc	*df*	*F*	η_p^2	Post hoc
Intercept	1	189.68***	.513		1	717.54***	.785		1	1250.49***	.846	
Contrast	3	48.28***	.446	b>k>f,e	3	123.34***	.654		3	195.05***	.720	
Type of English acquisition	2	8.77***	.089	B> EE, M	2	17.18***	.149	B> EE, M	2	38.19***	.251	B> EE, M
Contrast x Category	6	0.90			6	3.89**	.106	b>k>f, e (all groups)	6	6.94***	.154	b>k>f,e (M) b>k>f>e (EE) b>k>f<e (B)
	180				196				228			

N.B. η_p^2: < .02-.13 small, .14-.26 medium, >.26 large; M = Mainstream, EE = Early-English, B = Bilingual

Table 4.4.
ANOVAs with Type of English Acquisition and Age as Fixed Factors and d' for Each Contrast as Dependent Variable

		/b/-/s/			/k/-/g/			/f/-/θ/			/ɛ/-/æ/		
	df	*F*	η_p^2	Post hoc	*F*	η_p^2	Post hoc	*F*	η_p^2	Post hoc	*F*	η_p^2	Post hoc
Type of English acquisition	2	1.68	.032		21.31***	.258	M,EE<B	6.12**	.087	M,EE<B	47.10***	.384	
Age	2	54.99***	.435	4-5<8-9,11-12	47.77***	.387	4-5<8-9<11-12	5.55**	.061	4-5<8-9, 11-12	.024	.001	
Age x Category	4	1.29	.033		0.49	.013		1.03	.034		4.66**	.110	4-5: EE<B 8-9: M, EE<B 11-12: M,EE<B
Error	151												
R^2_{model}		.427			.472			.119			.398		

*p < .05; **p < .01; ***p < .001
N.B. η_p^2: < .02-.13 small, .14-.26 medium, >.26 large; M = Mainstream, EE = Early-English, B = Bilingual

4.4 Discussion

The aim of this study was to investigate whether Dutch children receiving early-English education were better able to discriminate between English phonetic contrasts than children attending mainstream Dutch primary schools, and whether they were as good as children growing up bilingually in Dutch and English. This question was investigated in three different age groups: 4-5-year-olds who had just started primary school, children who were in the middle of primary school (8-9-year-olds), and 11-12-year-old children who were at the end of primary school.

The first research question was whether the three English-acquisition-type groups of children (mainstream, early-English, bilingual) differed in their abilities to perceive English speech sounds. The hypothesis was that bilingual children might outperform both early-English and mainstream pupils. For the latter two groups, we hypothesized that if anything, the early-English pupils might perform better on the speech perception task than the mainstream pupils. The first part of this hypothesis was confirmed, the second was not. That is, the bilinguals showed greater sensitivity than the two other groups on all the contrast pairs except the easy control contrast /b/-/s/ (i.e., /k/-/g/, /f/-/θ/, /ɛ/-/æ/) but the early-English pupils did not show better performance than the mainstream pupils. The second hypothesis was that there would be age differences in performance on the different contrasts. In line with previous research (Hazan & Barrett, 2000), we found that children between 4 and 12 years old show an increase in phoneme categorization performance. This effect was larger for the /b/-/s/ than for the /k/-/g/ contrast, and smallest for the /f/-/θ/, as shown by a large, medium, and small effect size respectively.

The youngest group (4-5-year-olds) always showed less sensitivity to the contrasts than the older pupils, even to the easy /b/-/s/ contrast. This may be explained by developmental differences in phonemic categorization. It may also be that younger children have difficulty remaining concentrated during the whole task, resulting in sub-optimal performance, even on an easy contrast.

The 11- to 12-year-old children did not perform better than 8-to-9-year-old children on the /f/-/θ/ contrast, whereas they did for the /k/-/g/ contrast. The PAM predictions for category goodness contrast pairs are that L2 learners learn to hear the differences between the two phonemes after some experience with the L2 (Best & Tyler, 2007). It may be that the relatively small amount of time that mainstream and early-English pupils are exposed to English is enough for an initial growth in sensitivity towards the /f/-/θ/ contrast, but not enough for a further improvement. On the other hand, the oldest group of bilingual children did not perform significantly better than the 8-9-year-old group either. Even adult English native speakers find /f/ and /θ/ to be perceptually confusable (Cutler et al., 2004; Johnson & Babel, 2010). Therefore, the difficulties that the early-English pupils experienced with this contrast are hardly surprising. In fact, it could be that between 8-9 years and 11-12 year, perception of this contrast does not measurably develop.

For the /ɛ/-/æ/ contrast, mainstream and early-English pupils of all ages performed at chance level and showed no improvement at all with age. According to the predictions outlined in the PAM-L2 model (Best & Tyler, 2007), both members of such a single category contrast should be perceived as a similarly good or poor exemplar of the closest native phoneme and therefore learners are not likely to learn to distinguish between the two members (Best & Tyler, 2007). We proposed that /ɛ/

and /æ/ should be perceived as an exemplar of the Dutch /ɛ/, resulting in pupils of all ages performing at chance level. Our results seem to confirm this reasoning.

Our third hypothesis was that the early-English and mainstream pupils' perceptual difficulty would vary across the phonetic contrasts tested and depend on the degree of similarity of those phonemes in English and Dutch. This hypothesis was based on Best and Tyler (2007) who predict that L2 learners will have difficulty perceiving non-native speech contrasts, especially when one or both members are very similar but not identical to a native phonetic category. Since the bilingual children were native speakers of both Dutch and English, we predicted that their perception should generally be more accurate. We focused on four contrasts, which we expected to vary in difficulty for the early-English and mainstream pupils: /b/-/s/ (easy), /k/-/g/ (intermediate), /f/-/θ/ (hard), and /ɛ/-/æ/ (very hard). Our hypothesis that perceptual difficulty would vary across the phonetic contrasts was confirmed. All children (mainstream, early-English, and bilingual) performed best on the /b/-/s/ contrast. All Type-of-English-acquisition groups performed relatively well on the /k/-/g/ contrast too, although scores were significantly lower than on the /b/-/s/ contrast. We expected the k/-/g/ contrast to be of intermediate difficulty, as Dutch, first, contrasts between voiceless-voiced plosives at other places of articulation and second, contains /g/ in loan words and third, contains it as an allophone of /k/. Confirming our hypothesis that L2 acquiring pupils would have less difficulty with the /k/-/g/ contrast than with the /f/-/θ/ or ɛ/-/æ/ contrast, mainstream and early-English pupils performed worst on these two latter contrasts. For the oldest group of bilingual children, the performance pattern was different: they performed worse on /ɛ/-/æ/ than on /k/-/g/, but they had even more difficulty with /f/-/θ/.

The PAM-L2 model does not differentiate between novel L2 sounds that listeners are entirely unfamiliar with, and sounds that they do have some L1 experience with, because they represent a system gap such that they are familiar with relevant features, or because they occur in the L1 in loan words or as allophones. In this study, we tested perception of two contrasts of the category goodness type, /k/-/g/ and /f/-/θ/. Our results, with better performance for /k/-/g/ than for /f/-/θ/, imply that L2 learners have less difficulty distinguishing between two members of a contrast (here /k/-/g/) when they have L1 experience with the type of contrast (in this case concerning stop voicing) or with the 'absent' phoneme itself (because it occurs in L1 as a marginal phoneme and an allophone), than between those of a category goodness contrast that does not have this benefit of L1 familiarity (here /f/-/θ/). Another explanation for the difference between pupils' performance on the /k/-/g/ contrast and the /f/-/θ/ contrast could be that the difference between /f/-/θ/ is intrinsically difficult, as shown by the bilinguals' relatively low performance on this contrast. Further research on adult and child L2 learners' performance on different types of category goodness contrasts is needed to confirm whether contrasts such as the /k/- /g/ in this study are treated differently in L2 learning.

The overall aim of this study was to investigate if children who are learning English by means of an early-English educational programme outperform their peers who are not enrolled in such a programme. Early-English pupils did not outperform mainstream pupils on any of the non-native contrasts. Note that problems with performance on these contrasts cannot be due to the difficulty level of the task itself:

Even children in the youngest group were capable of performing the task, as indicated by their performance on the easy /b/-/s/ contrast. There are several explanations possible for these outcomes. First of all, it may be that the limited exposure to English at school is simply not enough for early-English pupils to learn the difference between English speech sounds, or for pupils' knowledge of English in general to develop.

Second, it could be that the exposure to English that children get at home is more important than the input at school. The available parental questionnaire data show that both mainstream and early-English pupils had considerable exposure to English media at home, and that early-English pupils had more exposure to English out of school than in school. Unfortunately, information about out-of-school exposure was not available for all children. Exploratory partial correlation analyses between out-of-school exposure and performance on each of the contrast pairs while controlling for age revealed no significant relations between the two variables (for /b/-/s/: r = -.266; /k/-/g/: r = .109; /f/-/θ/: r = .005; /ɛ/-/æ/: r = -.007; all ps > .05). Although for a correlation analysis to be more informative, these results suggest that out-of-school exposure to English does not contribute to pupils' perception of English contrasts.

An alternative explanation for the lack of a difference between the early-English and mainstream pupils' results could be that type of exposure that pupils receive at school does not contribute to the perception of L2 speech contrasts. As outlined in the introduction, English education in Dutch schools is generally aimed at improving pupils' oral proficiency, and specifically at improving their speaking and listening skills and their vocabulary knowledge (Thijs et al., 2011). Its goal is furthermore to advance pupils' understanding of spoken English and written English texts, and for them to become able (and confident) to communicate in English (Bodde et al., 2008). As being able to hear and produce the differences between English speech sounds is not one of the goals for English education, explicit instruction on perception or pronunciation of English phonemes appears not to be part of the educational curriculum. In addition, English lessons are generally provided by the regular classroom teacher (Jenniskens et al., 2017). Although the required English proficiency level is B2 (intermediate), teachers' proficiency levels vary largely, with a substantial number of teachers not obtaining the required level (Jenniskens et al., 2017). As a consequence, teachers may not correctly produce (some of the) English speech sounds. We did not measure the teachers' ability to produce the speech contrasts in question, so we do not know whether pupils received the correct input. Given that Dutch speakers of English have difficulty with pronouncing /θ/ (Wester, Gilbers, & Lowie, 2007), /ɛ/, and /æ/, and, albeit to a lesser extent, /g/ (Hamann & De Jonge, 2015), it may thus well be that teachers were not able to produce these sounds. They may thereby providing children with Dutch-accented English, in which these sounds were not pronounced correctly, and limiting the possibilities for pupils to implicitly learn to distinguish between those speech sounds.

The results of this study are not in line with the premise that early-English education has a positive effect on pupils' English language knowledge. We did not find any advantages for the early-English group compared to the mainstream group in their perception of English speech sounds. We also measured pupils' English

vocabulary, to examine whether early-English pupils' English vocabulary is greater than that of mainstream pupils. Contrary to what might be expected, we did not find advantages for the early-English group in this domain either. These outcomes is important for policy makers, who base the implementation of English educational programmes on such premises. Such premises are often based on findings that show that children who are growing up bilingually in naturalistic settings successfully master two languages (Muñoz, 2008). Growing up with two languages is however a fundamentally different way of language learning than learning a new language at school while one language has already largely developed (Muñoz, 2008). This study shows that early-English education, as it is currently implemented in the Dutch school system, may have very limited effects on pupils' abilities to perceive differences between English speech contrasts. Early-English education might be beneficial to pupil's perception of English speech contrasts if teaching perception and pronunciation of non-native phonemes receives a more prominent role in the curriculum. Research with adults has shown that for those learning an L2 after their L1 has already been largely established, both perception and pronunciation teaching may benefit the acquisition of non-native speech contrasts (Lee & Lyster, 2016; Saito, 2013), even when the teacher is not a native speaker herself (Levis, Sonsaat, Link, & Barriuso, 2016).

As our study was not longitudinal, we cannot rule out the possibility that early-English pupils older than four but younger than eight (i.e., between the ages at which we tested them) have an advantage in vocabulary or in the perception of the /f/-/θ/ contrast that the mainstream pupils later catch up on. This cross-sectional study nevertheless provides a first and important exploration of the question whether early-English programmes are beneficial for pupils' perception of English speech contrasts. Because we assessed pupils who were at the end of primary school, we could show that even after eight years of English lessons, pupils are not better able to perceive English speech contrasts that do not exist in Dutch than pupils who started English education little more than a year before being tested. By investigating multiple contrasts that varied in difficulty, we have shown that young learners of English, both in early-English and in mainstream education, did manage to learn to distinguish the difference between members of some non-native phoneme pairs, but only those that were easy or of moderate difficulty.

4.4.1 Conclusion

This study shows that early-English education is not beneficial to non-native speech perception: Dutch children who get a maximum of two hours of English lessons per week from the moment they enter primary school at the age of four are not better able to perceive the difference between members of English contrasts than children who start in the penultimate grade of primary school. Children growing up bilingually with Dutch and English were however better at hearing the differences in all phonetic contrasts except the easy control contrast than the children who did not grow up bilingually at home. Starting to acquire English in an instructed setting at a young age thus appears not to be beneficial for learning to perceptually distinguish non-native contrasts. For pupils to learn to perceive non-native speech contrasts, perception (and pronunciation) instruction should get a more prominent role in the early-English curriculum.

4.5 Notes

[1] The same analyses without the children who were exposed to three or four languages revealed the same pattern of results, for all contrasts.

[2] When the analyses for *d'* were performed again with the children who had <60% correct on the /b/-/s/ contrast, the results largely remained the same, with only some minor shifts in the outcomes regarding the age differences.

4.6 References

Aoyama, K., Flege, J. E., Guion, S. G., Akahane-Yamada, R., & Yamada, T. (2004). Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese /r/ and English /l/ and /r/. *Journal of Phonetics, 32*, 233–250. https://doi.org/10.1016/S0095-4470(03)00036-6

Baker, W., Trofimovich, P., Flege, J. E., Mack, M., & Halter, R. (2008). Child–adult differences in second-language phonological learning: The role of cross-language similarity. *Language and Speech*, *51*(4), 317–342. https://doi.org/10.1177/0023830908099068

Best, C. T. (1993). Emergence of language-specific constraints in perception of non-native speech: A window on early phonological development. In B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. McNeilage, & J. Morton (Eds.), *Developmental Neurocognition: Speech and Face Processing in the First Year of Life. NATO ASI Series (Series D: Behavioural and Social Sciences), vol 69.* (pp. 289–304). Dordrecht: Springer. https://doi.org/10.1007/978-94-015-8234-6_24

Best, C. T. (1995). A direct relaist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171–204). Timonium, MD: New York Press.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In M. J. Munro & O.-S. Bohn (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 13–34). Amsterdam: John Benjamins. https://doi.org/10.1075/lllt.17.07bes

Bodde, M., Schippers, S., Klein Tank, M., & van der Linde-Meijerink, G. (2008). *TULE-Engels. Inhouden en activiteiten bij de kerndoelen van 2006 [TULE-English. Content and activities of the objectives of 2006]*. Enschede. Retrieved from http://tule.slo.nl

Bohn, O.-S., & Munro, M. J. (2007). *Language experience in second language speech learning: In honor of James Emil Flege*. Amsterdam: John Benjamins.

Bosch, L., & Sebastián-Gallés, N. (2003a). Language experience and the perception of a voicing contrast in fricatives: Infant and adult data. *Proceedings of the 15th International Conference of Phonetic Sciences*, 1987–1990.

Bosch, L., & Sebastián-Gallés, N. (2003b). Simultaneous bilingualism and the perception of a language-specific vowel contrast in the first year of life. *Language and Speech*, *2*, 217–243.

Broersma, M. (2005). Perception of familiar contrasts in unfamiliar positions. *Journal of the Acoustical Society of America*, *117*, 3890–3901. https://doi.org/10.1121/1.1906060

Burns, T. C., Yoshida, K. A., Hill, K., & Werker, J. F. (2007). The development of

phonetic representation in bilingual and monolingual infants. *Applied Psycholinguistics*, *28*, 455–474. https://doi.org/10.1017.S0142716407070257

Choi, J., Cutler, A., & Broersma, M. (2017). Early development of abstract language knowledge: Evidence from perception–production transfer of birth-language memory. *Royal Society Open Science*, *4*(160660), 1–14. https://doi.org/10.1098/rsos.160660

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: L. Lawrence Earlbaum Associates. Retrieved from http://imaging.mrc-cbu.cam.ac.uk/statswiki/FAQ/effectSize

Cutler, A. (2012). *Native listening: Language experience and the recognition of spoken words*. Cambridge: The MIT Press.

Cutler, A., Weber, A., Smits, R., & Cooper, N. (2004). Patterns of English phoneme confusions by native and non-native listeners. *The Journal of the Acoustical Society of America*, *116*, 3668–3678. https://doi.org/10.1121/1.1810292

Darcy, I., & Krüger, F. (2012). Vowel perception and production in Turkish children acquiring L2 German. *Journal of Phonetics*, *40*, 568–581. https://doi.org/10.1016/J.WOCN.2012.05.001

de Graaff, R. (2015). Vroeg of laat Engels in het basisonderwijs: Wat levert het op? [Early or late English in primary education: Does it matter?]. *Levende Talen Tijdschrift*, *16*(2), 3–15.

Dunn, L. M., & Dunn, D. M. (2007). *PPVT-4 Manual*. Bloomington: NCS Pearson.

Dunn, L. M., Dunn, L. M., & Schlichting, L. (2005). *Peabody Picture Vocabulary Test-III-NL* (2nd ed.). Amsterdam: Pearson Assessment and Information.

Enever, J. (2013). *Contemporary societal perspectives on English in Europe*. Chile.

Flege, J. E., MacKay, I. R. A., & Meador, D. (1999). Native Italian speakers' perception and production of English vowels. *The Journal of the Acoustical Society of America*, *106*, 2973–2987. https://doi.org/10.1121/1.428116

Flege, J. E., Takagi, N., & Mann, V. (1996). Lexical familiarity and English-language experience affect Japanese adults' perception of /r/ and /l/. *The Journal of the Acoustical Society of America*, *99*, 1161–2711. https://doi.org/10.1121/1.421225

Geurts, B., & Hemker, B. (2013). *Balans van het Engels aan het einde van de basisschool 4 [Balance of English at the end of primary school. Outcomes of the 4th evaluation in 2012]*. Arnhem. Retrieved from https://www.cito.nl/-/media/Files/kennis-en-innovatie-onderzoek/ppon/cito_ppon_balans_52.pdf?la=nl-NL

Goriot, C., Broersma, M., McQueen, J. M., Unsworth, S., & Van Hout, R. (2018). Language balance and switching ability in children acquiring English as a second language. *Journal of Experimental Child Psychology*, *173*, 168–186.

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds "L" and "R." *Neuropsychologia*, *9*, 317–323.

Gussenhoven, C. (1999). Illustrations of the IPA: Dutch. In *Handbook of the International Phonetic Association* (pp. 74–77). Cambridge: Cambridge University Press.

Gussenhoven, C., & Jacobs, H. (2011). *Understanding phonology* (3rd ed.). London: Hodder Education.

Hamann, S. R., & De Jonge, A. (2015). Eliciting the Dutch loan phoneme /g/ with

the Menu Task. In *Proceedings of the 18th International Congress of Phonetic Sciences Glasgow*. Glasgow: The University of Glasgow.

Hanulíková, A., & Weber, A. (2012). Sink positive: Linguistic experience with th substitutions influences nonnative word recognition. *Attention, Perception, & Psychophysics*, *74*, 613–629. https://doi.org/10.3758/s13414-011-0259-7

Hazan, V., & Barrett, S. (2000). The development of phonemic categorization in children aged 6–12. *Journal of Phonetics*, *28*, 377–396. https://doi.org/10.1006/JPHO.2000.0121

Jenniskens, T., Leest, B., Wolbers, M., Bruggink, M., Dood, C., & Krikhaar, E. (2017). *Zicht op vroeg vreemdetalenonderwijs [Insight into early foreign language education]*. Nijmegen.

Johnson, K., & Babel, M. (2010). On the perceptual basis of distinctive features: Evidence from the perception of fricatives by Dutch and English speakers. *Journal of Phonetics*, *38*(1), 127–136. https://doi.org/10.1016/J.WOCN.2009.11.001

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience*, *5*, 831–843. https://doi.org/10.1038/nrn1533

Kuppens, A. H. (2010). Incidental foreign language acquisition from media exposure. *Learning, Media and Technology*, *35*, 65–85. https://doi.org/10.1080/17439880903561876

Lee, A. H., & Lyster, R. (2016). Effects of different types of corrective feedback on receptive skills in a second language: A speech perception training study. *Language Learning*, *66*, 809–833. https://doi.org/10.1111/lang.12167

Levis, J. M., Sonsaat, S., Link, S., & Barriuso, T. A. (2016). Native and Nonnative Teachers of L2 Pronunciation: Effects on Learner Performance. *TESOL QUARTERLY*, *50*. https://doi.org/10.1002/tesq.272

Lobo, V. R. (2013). *Teaching L2 English at a very early age : A study of Dutch schools*. Radboud University. Retrieved from https://www.lotpublications.nl/teaching-l2-english-at-a-very-young-age-teaching-l2-english-at-a-very-young-age-a-study-of-dutch-schools

MacMillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. Cambridge: Cambridge University Press.

McCarthy, K. M., Mahon, M., Rosen, S., & Evans, B. G. (2014). Speech perception and production by sequential bilingual children: A longitudinal study of voice onset time acquisition. *Child Development*, *85*, 1965–1980. https://doi.org/10.1111/cdev.12275

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

Nuffic. (2017). Locaties vvto-scholen [Locations early foreign language schools]. Retrieved August 28, 2017, from https://www.nuffic.nl/primair-onderwijs/talenonderwijs/vroeg-vreemdetalenonderwijs-vvto/locaties-vvto-scholen

Onderwijsraad. (2008). *Vreemde talen in het onderwijs [foreign languages in education]*. The Hague: OBT bv.

Saito, K. (2013). The acquisitional value of recasts in instructed second language speech learning: Teaching the perception and production of English /ɹ/ to adult

Japanese learners. *Language Learning Research Club*, *63*, 499–529. https://doi.org/10.1111/lang.12015

Sebastián-Gallés, N., & Bosch, L. (2009). Developmental shift in the discrimination of vowel contrasts in bilingual infants: Is the distributional account all there is to it? *Developmental Science*, *12*, 874–887. https://doi.org/10.1111/j.1467-7687.2009.00829.x

Singleton, D., & Ryan, L. (2004). *Language acquisition: The age factor* (2nd Ed.). Clevedon: Multilingual Matters.

Strange, W. (1995). Speech perception and linguistic experience: Issues in cross-language research. *Phonetica*, *56*, 105–107. https://doi.org/10.1159/000028444

Sundara, M., & Polka, L. (2008). Discrimination of coronal stops by bilingual adults: The timing and nature of language interaction. *Cognition*, *106*(1), 234–258. https://doi.org/10.1016/J.COGNITION.2007.01.011

Tsukada, K., Birdsong, D., Bialystok, E., Mack, M., Sung, H., & Flege, J. (2005). A developmental study of English vowel production and perception by native Korean adults and children. *Journal of Phonetics*, *33*, 263–290. https://doi.org/10.1016/j.wocn.2004.10.002

Unsworth, S., Persson, L., Prins, T., & de Bot, K. (2015). An investigation of factors affecting early foreign language learning in the Netherlands. *Applied Linguistics*, *36*, 527–548. https://doi.org/10.1093/applin/amt052

van Bezooijen, R., & Gerritsen, M. (1994). De uitspraak van uitheemse woorden in het Standaard-Nederlands: Een verkennende studie [The pronunciation of foreign words in Standard Dutch: An explorative study]. *De Nieuwe Taalgids*, *87*, 145–161.

van der Leij, A., Bekebrede, J., & Kotterink, M. (2010). Acquiring reading and vocabulary in Dutch and English: The effect of concurrent instruction. *Reading and Writing*, *23*, 415–434. https://doi.org/10.1007/s11145-009-9207-5

Walley, A. C., & Flege, J. E. (1999). Effect of lexical status on children's and adults perception of native and non-native vowels. *Journal of Phonetics*, *27*, 307–332. Retrieved from http://www.idealibrary.comon

Wechsler, D., & Naglieri, J. A. (2008). *WNV NL. Wechsler Nonverbal Scale of ability. Nederlandstalige bewerking.* Amsterdam: Pearson.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behaviour and Development*, *7*, 49–63.

Wester, F., Gilbers, D., & Lowie, W. (2007). Substitution of dental fricatives in English by Dutch L2 speakers. *Language Sciences*, *29*, 477–491. https://doi.org/10.1016/J.LANGSCI.2006.12.029

Zhou, W. (2015). *Assessing birth language memory in young adoptees*. Radboud University Nijmegen. Retrieved from http://repository.ubn.ru.nl/bitstream/handle/2066/143200/143200.pdf?sequence=1

Chapter 5: Using the Peabody Picture Vocabulary Test in L2 children and adolescents: Effects of L1.

This chapter is based on:
Goriot, C., van Hout, R., Broersma, M., Lobo, V., McQueen, J. M., & Unsworth, S. (2018). Using the Peabody Picture Vocabulary Test in L2 children and adolescents: Effects of L1. *International Journal of Bilingual Education and Bilingualism, Advance online publication*. doi: 10.1080/13670050.2018.1494131

Chapter 5

Using the Peabody Picture Vocabulary Test in L2 children and adolescents: Effects of L1

Abstract

This study investigated to what extent the Peabody Picture Vocabulary Test (PPVT-4) is a reliable tool for measuring vocabulary knowledge of English as a second language (L2), and to what extent L1 characteristics affect test outcomes. The PPVT-4 was administered to Dutch pupils in six different age groups (4 to 15 years old) who were or were not following an English educational programme at school. Our first finding was that the PPVT-4 was not a reliable measure for pupils who were correct on maximally 24 items, but it was reliable for pupils who performed better. Second, both primary-school and secondary-school pupils performed better on items for which the phonological similarity between the English word and its Dutch translation was higher. Third, young unexperienced L2 learners' scores were predicted by Dutch lexical frequency, while older more experienced pupils' scores were predicted by English frequency. These findings indicate that the PPVT may be inappropriate for use with L2 learners with limited L2 proficiency. Furthermore, comparisons of PPVT scores across learners with different L1s are confounded by effects of L1 frequency and L1-L2 similarity. The PPVT-4 is however a suitable measure to compare more proficient L2 learners who have the same L1.

5.1 Introduction

Vocabulary tests are frequently used in research on monolinguals, and early bilinguals and second language (L2) learners, both to measure children's vocabulary development in a specific language, and to evaluate their overall language abilities (e.g., Bialystok, Luk, Peets, & Yang, 2010; Dongsun, Yoon, & Jiyeon, 2016; Poarch & van Hell, 2012b). As previously noted (Gathercole et al., 2008), most often these tests have been developed for use with native speakers of the language being assessed, and therefore have been normed on a monolingual population. The way in which learners acquire an L2, however, by definition differs from how they acquire a first language (L1), in terms of for example context or age. The use of an L1 vocabulary test with an L2 population may therefore be problematic: Various L1 and L2 factors may have an influence on test outcomes, such as linguistic overlap between the L1 and L2 or how frequently young L2 learners encounter certain words. The question we address in this paper is to what extent an L1 vocabulary test can be reliably used with young L2 learners.

For children acquiring English as an L1, many language proficiency tests are available (for a discussion see Gathercole et al., 2008), one of the most commonly-used being the Peabody Picture Vocabulary Test (PPVT), a receptive vocabulary test (Dunn, 1959). This test has been widely used with monolingual English speakers, and with bilingual children in non-instructed settings (see for example Bialystok, Luk, Peets, & Yang, 2010 for a meta-analysis). The use of the PPVT (or its British-English equivalent the BPVS) in L2 acquisition contexts is

widespread, too (see for example Cohen, 2016; Crevecoeur, Coyne, & McCoach, 2014; Dahl & Vulchanova, 2014; Dongsun, Yoon, & Jiyeon, 2016; Jensen, 2017; Leśniewska & Pichette, 2016).

One of the domains in which the PPVT is frequently used with L2 learners is the domain of early foreign-language education. Also, in various European countries, using English as the language of instruction in addition to the official language in education has gained in popularity in recent years, both in primary (Huang, 2016) and secondary education (Nikula, 2017). Many researchers have investigated whether these educational programmes lead to gains in English receptive vocabulary. Such studies have been conducted in many countries, including Belgium (Buyl & Housen, 2014), Finland (Merisuo-Storm, 2007), France (Cohen, 2016), Germany (Steinlen & Piske, 2013), the Netherlands (Admiraal, Westhoff, & de Bot, 2006; Lobo, 2013; Unsworth, Persson, Prins, & de Bot, 2015; van der Leij, Bekebrede, & Kotterink, 2010), Norway (Dahl & Vulchanova, 2014), and Spain (Jimenez Catalan & Terrazas Gallego, 2005). Many of them made use of the PPVT or the BPVS, either in its original form (Buyl & Housen, 2014; Cohen, 2016; Dahl & Vulchanova, 2014; Unsworth et al., 2015; van der Leij et al., 2010) or back-translated from the Dutch version of the test to English (Lobo, 2013). The PPVT was not designed as a task for L2 learners, however, and the pupils in those studies learned English in a context differing from the one in which L1 learners learned it, at an age at which their L1 had already developed substantially. Different factors, such as the frequency with which certain words are used in the L2 or the linguistic overlap between the L1 and L2, may play a role in their L2 vocabulary development and influence their scores on the PPVT. In this study, we investigate the extent to which, given these factors, the PPVT is suitable for use with young Dutch pupils learning English as an L2 via an educational programme.

5.1.1 The PPVT

The PPVT was originally developed as a measure of verbal intelligence (Dunn, 1959). The English version of the PPVT is currently in its fourth edition. This edition consists of 228 items grouped in 19 sets of 12 items each, arranged in order of decreasing frequency, with increasing difficulty being assumed. An item consists of four full-colour pictures. The participant has to select the picture that best matches the orally presented word. The items include verbs, adjectives, and nouns. The words belong to one of 20 different content categories, like animals, actions, or emotions. According to the guidelines in the manual, the start set is dependent on the age of the participant. If the participant gives two or more incorrect responses in one set, an easier set is presented. The basal set is determined as the set in which maximally one incorrect response is given. After selection of the basal set, the main phase of testing begins. Testing ends when the participant makes more than seven incorrect responses within a single set, thereby reaching the ceiling set. The PPVT-4 was normed on a sample of 3540 people, representative for the population of the United States. Both split-half and test-retest reliability were consistently high, with coefficients higher than .90 for all age groups (Dunn & Dunn, 2007).

It is not surprising that the use of the PPVT is so widespread: it requires no literacy skills or oral response, there is minimal risk for stress or perceived failure, it is appropriate for use with participants aged two and older, and the coloured pictures

are perceived as appealing to children. The authors of the test consider it as a useful measure for the assessment of "the extent and nature of a person's knowledge of standard American English words" (Dunn & Dunn, 2007, p. 3), even for individuals whose L1 is not English (Dunn & Dunn, 2007). However, as noted above, the PPVT was not developed for L2 learners. Its use for such a population may therefore not be completely unproblematic. A potential problem involves the possible interplay between the L1 and the L2, and this may influence outcomes (Wood & Pena, 2015).

5.1.2 L1 effects when investigating L2 vocabulary

A common view in the L2 acquisition literature is that L2 learners do not begin from scratch when they start building up a lexicon in the new language. Instead, they rely on the knowledge and the concepts that they have in their L1 (Dijkstra & Van Heuven, 2002; Kroll & Stewart, 1994). One aspect of children's L1 that may affect their vocabulary acquisition and, in particular, their performance on the PPVT is L1 word frequency. The structure of the PPVT-4 is based on the idea that some words are more frequent than others, and that children will acquire more frequent words before less frequent words and hence are more familiar with more frequent words (Dunn & Dunn, 2007). However, word familiarity may be different for L2 learners than for native speakers. Wood and Pena (2015) showed that the difficulty level of the items in the PPVT-4, as determined by their order in the test, was positively related to children's error scores, but this relation was stronger in English monolingual children than in Spanish L2 learners of English. Differences in word familiarity between L1 and L2 learners may be especially more likely for children who have limited L2 exposure, as is the case for Dutch children who are exposed to English either at school or via media (Lindgren & Muñoz, 2013). Words that are frequent in English as an L1 are not necessarily frequent in English as an L2, but frequency measures for English as an L2 are not available. It is therefore difficult to determine to which specific words children are exposed, and with what frequency. While we thus have no suitable measure of the frequency with which words may have been experienced in *English* in the sample of Dutch children we test, we investigate whether performance on the PPVT depends on *Dutch* frequency. It is however likely that children who start acquiring a new language at school at a time that they already have acquired their first language will do so by making use of their L1 lexicon. Consequently, the frequency of the words in the L1 may be a more important predictor of word knowledge in the L2 (English) than the L2 frequency itself.

Children may also rely on their knowledge of the L1 by recognizing the similarities between words in the L1 and the L2. This might be particularly helpful for cognates, which in this study are defined as words that show semantic overlap between two languages, as well as large similarities in spelling and/or sounds. Translation pairs that show orthographic and phonetic similarities but that have different meanings (i.e., 'false friends') are not included in this study. It is known that meaning and form overlap helps children derive the meaning of a word (Pérez, Peña, & Bedore, 2010; Potapova et al., 2016), even when children have limited exposure to the L2 (Bosma, Blom, Hoekstra, & Versloot, 2016). Children have also been found to process cognates faster than non-cognates (Brenders, van Hell, & Dijkstra, 2011; Poarch & van Hell, 2012a). Just like adults (Dijkstra, Miwa,

Brummelhuis, Sappelli, & Baayen, 2010), children seem to show a gradual cognate facilitation effect: they are more likely to know the meaning of an identical cognate item than of a non-identical cognate, although they are also able to derive the meaning of non-identical cognates (Bosma et al., 2016). Older children are better at recognizing cognates than younger children, especially if the items are non-identical cognates (Bosma et al., 2016). Since the number of cognates is greatest in closely related language pairs (Schepens, Dijkstra, Grootjen, & van Heuven, 2013), and Dutch and English are both Germanic languages that are known to share a large proportion of cognates (Broersma, 2009; Schepens et al., 2013), it is likely that the PPVT-4 will contain relatively many English words of which the phonological similarity is close to their Dutch translation equivalent. When administering the PPVT-4 to native Dutch children it is thus to be expected that they may well benefit from these items.

Indeed, it has been noted that the PPVT-4 contains cognates for Dutch-speaking children (Lobo, 2013; Unsworth et al., 2015). The researchers mentioned that this may have helped monolingual children with word association and recognition. Previous research has also shown that the PPVT-3 and PPVT-4 contain Spanish-English cognates (Potapova et al., 2016; Wood & Pena, 2015), and that Spanish-English bilingual children perform better on cognates than on non-cognates (Potapova et al., 2016). Furthermore, a recent study with adult L2 learners showed that the PPVT-4 contains more French-English than Polish-English cognates, and hence French L1 speakers obtained higher scores on the PPVT-4 than Polish L1 speakers (Leśniewska, Pichette, & Béland, 2018). Similarly, in a large-scale experiment which compared foreign-language learners in seven European countries (Lindgren & Muñoz, 2013) children living in a country in which the language of schooling was linguistically close to English (i.e., Swedish and Dutch) performed better on English listening and reading tasks than children living in a country where the language of schooling was less close to English (i.e., Croatian and Polish). Whilst that experiment did not use the PPVT, it does suggest that children might rely on cognate knowledge when being tested in their L2. All these studies thus suggest that when investigating L2 English vocabulary using the PPVT children's scores might be influenced by cognates.

5.1.3 The current study

Any measure of children's L2 vocabulary may thus be influenced by the fact that the frequency of the L2 words does not have to be the same as the frequency of the words in that language as an L1, and by the degree to which translation equivalents of the items in a test overlap in phonological form. Research on these issues is however limited. The current study addresses these issues by investigating the extent to which word frequency and phonological overlap between item-translation pairs predict the performance on the PPVT-4 by Dutch children who have limited exposure to English. The goal of this study more generally was to investigate how suitable the PPVT-4 is for measuring vocabulary knowledge of Dutch children and adolescents at different stages of learning English as an L2.

We conducted three experiments. In Experiment 1, in order to investigate lexical frequency, we translated all items of the PPVT-4 from English to Dutch, and examined if Dutch and English lexical frequencies are correlated. We also

investigated cognate status. Contrary to previous studies (Bosma et al., 2016; Pérez et al., 2010; Potapova et al., 2016; Wood & Pena, 2015), we used a continuous measure instead of an arbitrary cut-off point to determine phonological similarity between pairs of English items and their translations. We expected the frequencies to be closely related to each other, since previous research has shown that even unrelated languages show considerable overlap in frequency (Moscoso del Prado Martín, Bertram, Häikï, Schreuder, & Baayen, 2004). Since English and Dutch are linguistically close, we expected the similarity of the item-translation pairs to be high.

In Experiments 2 and 3, we investigated whether word frequencies and phonological similarity measures collected in Experiment 1 predicted respectively primary-school and secondary-school pupils' scores on the PPVT-4. The children were learning English via an educational programme and/or were exposed to English via media. In both experiments, we investigated, as in previous studies (Buyl & Housen, 2014; Cohen, 2016; Dahl & Vulchanova, 2014; Lobo, 2013; Unsworth et al., 2015; van der Leij et al., 2010), whether pupils who attended an English programme at their school and those who did not differ in their performance on the PPVT-4.

Experiments 2 and 3 were used to test three hypotheses. The first hypothesis was that the reliability of the PPVT-4 would increase when administering it to pupils who had more experience with English. Across the board, the older pupils are expected to have more experience with English than the younger pupils, and therefore to reach higher sets of the PPVT-4. For the youngest pupils, on the other hand, testing may regularly stop after the first few sets, resulting in a floor effect. Therefore, the test may not reliably differentiate between young pupils' vocabulary abilities.

Our second hypothesis was that children would perform better on English items that are more similar in form to their Dutch translations. Furthermore, since previous research has shown that children are better at recognizing similarities between words as they become older (Bosma et al., 2016), we expected that the relation between form similarity and performance would get stronger in older pupils.

Third, we hypothesized that there would be a decreasing L1 frequency effect and an increasing L2 frequency effect as pupils get older. We expected that for the younger pupils in particular, L1 (Dutch) frequency would affect their performance, because of their limited experience with English. Pupils in Dutch early-English primary schools generally receive English lessons for maximally one hour per week (Jenniskens et al., 2017) and mainly in an educational setting, and hence there is more room for Dutch than for English frequency to play a role. We expected that for older children, English frequency would become a more important predictor of vocabulary performance, since older pupils have more experience with and exposure to English. They may be exposed to English more, both outside school (Lindgren & Muñoz, 2013), and at school: in mainstream education, secondary-school pupils receive between two and four hours of English lessons per week, and in bilingual education 50% of the lessons is in English (approximately 10 hours per week for pupils following pre-university training) (EP-Nuffic, n.d.).

5.2 Experiment 1: Lexical frequencies and cognate status

The aim of this experiment was to investigate similarities and differences between pairs of English items and their Dutch translations, in terms of both lexical frequency and cognate status (operationalised as phonological similarity). Our expectation was that the lexical frequencies would be rather close (Moscoso del Prado Martín et al., 2004). We also expected that many pairs would show phonological overlap, given the West-Germanic origin of both languages.

5.2.1 Method

5.2.1.1 Word frequency

Two online corpora were used to obtain word frequencies per million words: the SUBTLEX-US corpus (van Heuven, Mandera, Keuleers, & Brysbaert, 2014) and the SUBTLEX-NL corpus (Keuleers, Brysbaert, & New, 2010) for English and Dutch words, respectively. The US corpus contains 51 million words, and the NL corpus 44 million. The advantage of the SUBTLEX corpora over traditional written corpora is that they are based on film and television subtitles, and thus on spoken language. Frequency estimates based on spoken language seem to be more accurate than estimates based on written language for explaining language processing in children (Brysbaert & New, 2009; Keuleers et al., 2010).

5.2.1.2 Translations

We translated the items in the PPVT-4 from English to Dutch, making use of the Longman Dictionary of Contemporary English for Advanced Learners (Longman, 2012) and the online version of the Van Dale Dutch-English translation dictionary (Albers, 2015). One translation was chosen, optimising three criteria: the match with the target picture in the PPVT-4, the closeness of the corresponding meaning of the Dutch word, and the similarity between the frequency of the Dutch translation and the English word. In the PPVT-4, verbs are presented in their –ing form. Such verb forms, expressing ongoing action, are uncommon in Dutch. We therefore used root forms in English and Dutch verb pairs, for example, ‘jump’ and its Dutch equivalent *‘spring’*.

5.2.1.3 Phonological transcriptions

Phonological transcriptions for all items were retrieved using the Longman Pronunciation Dictionary (Wells, 2008). For Dutch, the ‘Uitspraakwoordenboek’ (Heemskerk & Zonneveld, 2000) and the ‘Van Dale Middelgroot Woordenboek’ dictionary (Albers, 2015) were used. We used the X-SAMPA system for these transcriptions.

5.2.1.4 Objective phonological similarity

We calculated a normalized Levenshtein Distance (LD) in order to determine to what extent word pairs were similar, following Schepens et al. (2013). Because children are presented with the oral and not the written form of the items in the PPVT-4, we chose to focus on the phonological LD. The distance between an English item and its Dutch translation was calculated as the minimum number of insertions, deletions, and substitutions required to go from one to the other in X-SAMPA notation. All changes were given a weight of 1. For example, for the

English item <ankle> and its Dutch translation <enkel>, the phonological difference between English [{Nkl] and Dutch [ENk@l] is 2. Following Schepens et al. (2013) we subtracted the normalized distance from 1 to obtain the phonological similarity:

$\text{PhonSim} = 1 - \frac{\text{distance}}{\text{length}}$.

Length was operationalized as the segmental length of the longest word, either the English one or its Dutch translation. The outcome ranges between 0 and 1, where 0 means that there is no overlap between the two strings (completely dissimilar), and 1 means that the strings are identical (completely similar). In case of the example given above, the length of the longest word (the number of segments) was 5, and therefore the phonological distance was calculated as $1 - \frac{2}{5} = .60$. If no translation was available (as was the case for three low frequent words), PhonSim was set to 0 (no overlap).

5.2.1.5 Subjective phonological similarity

In addition to the objective similarity measure, a subjective measure was determined for the first 168 items in the PPVT-4. This allowed us to examine whether the two measures correlate with each other (cf. Potapova et al., 2016). Participants were 25 native speakers of Dutch (M_{age} = 24.4; SD_{age} = 4.8). They were recruited at Radboud University (Nijmegen, the Netherlands). All participants were volunteers with no known hearing or visual disorders, and gave written consent before taking part in the experiment. They were rewarded with five Euros for their participation.

The similarity rating task was programmed in WebExp2, an application for online experiments developed at Radboud University. Every participant was presented with all 168 item-translation pairs of the PPVT-4, in random order. On each trial, participants were presented with the written English and Dutch form, and the recorded English spoken form that was used in the PPVT-4. Participants were asked to rate the similarity of the English word form and their Dutch translation on a seven-point scale, ranging from *(1) completely different* to *(7) completely similar*. Participants were not presented with the spoken Dutch form to keep the procedure similar to that used in the administration of the PPVT-4. Participants were explicitly asked to pay attention only to the phonetic overlap. The intraclass correlation coefficient was high (*ICC* = .993), showing that the raters strongly agreed in their ratings.

5.2.2 Results and Discussion

5.2.2.1 Word frequencies

Figure 5.1 shows the average log frequency per million words of the English items and that of their Dutch translations, per set. As expected from how the PPVT-4 was developed, the frequency of the English words declines in the higher sets - although not consistently so. In some sets (e.g., Set 5), the discrepancy between the English and the Dutch frequency is comparatively large, whereas it is minimal in others (e.g., Set 9).

In Figure 5.2, we plotted the English and Dutch log frequencies against each other for all pairs, to examine how strongly they are related. Although there is an overall positive correlation between English and Dutch frequency (r = .783, p < .001), several pairs clearly deviate from this tendency. Discrepancies were largest when words had only one or a few meanings in one language and multiple meanings in the other (see Figure 5.2 for examples).

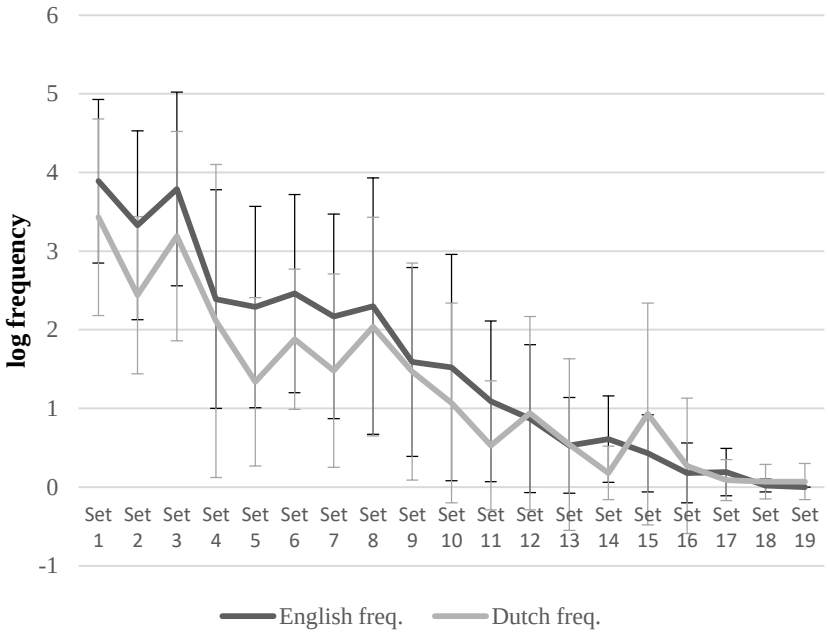


Figure 5.1. Experiment 1: Average English and Dutch log word frequency per set with their standard errors.

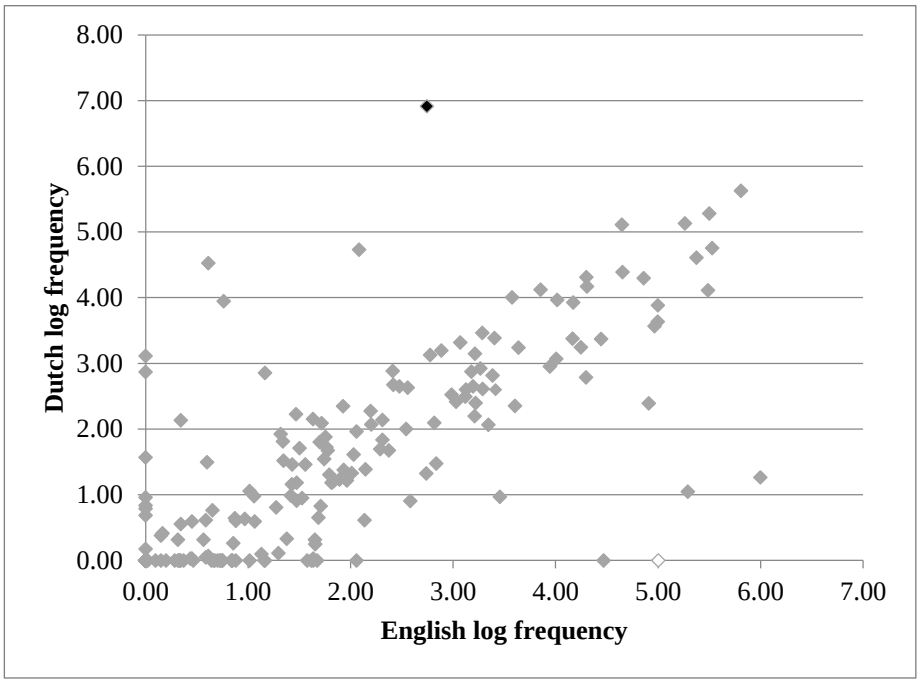


Figure 5.2. Experiment 1: The relation between English and Dutch log frequency. The word 'net' (depicted by the black dot), for example, has a single predominant meaning in English (something that is made of openwork fabric), and many high-frequency meanings in Dutch (in addition to something made of openwork fabric, it also means 'network', 'tidy', 'decent', and 'exactly'). In the case of 'sort' (depicted by the open dot), which we translated with 'sorteer', it is the other way around.

5.2.2.2 Phonological similarity

None of the translation pairs were identical, but 15 pairs had a PhonSim of .71 or higher. Figure 5.3 shows the average PhonSim per set. E.g., the average PhonSim of set 6 is relatively high, whereas that of set 7 is relatively low.

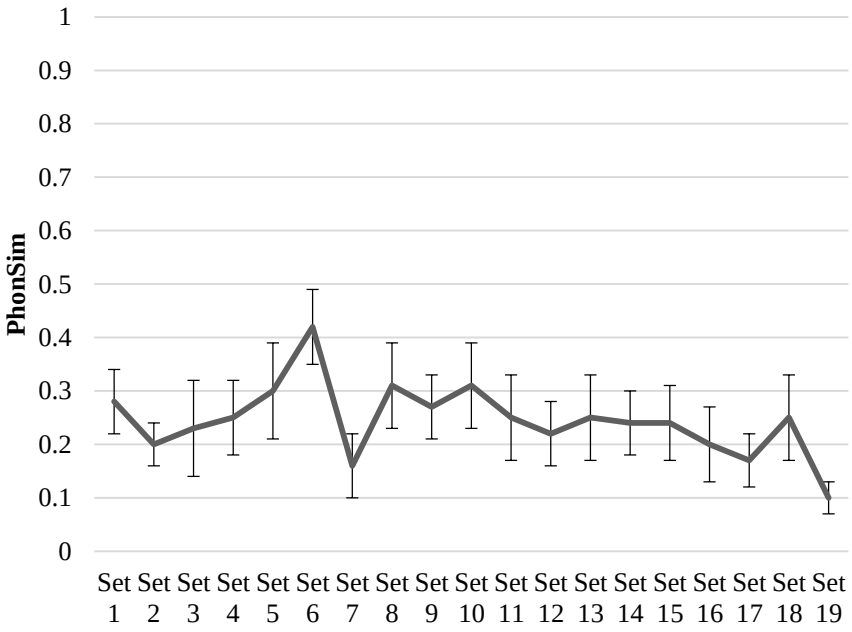


Figure 5.3. Experiment 1: Average PhonSim per set, and their standard errors.

To assess the validity of the objective measure of form overlap, PhonSim, the correlation between the PhonSim and the subjective similarity judgments was determined. The positive and high correlation found between objective and subjective phonological similarity (r = .800, p < .001) implies that PhonSim gives a useful, valid measure to identify phonological overlap between words.

In summary, the results of Experiment 1 show that a considerable number of pairs showed substantial phonological overlap. Furthermore, the frequency of the English and Dutch pairs are highly correlated, but crucially not identical, and there is variation in the similarity of the English and Dutch frequency of the translation pairs.

5.3 Experiment 2: primary-school pupils

The aim of this experiment was to investigate whether the PPVT-4 is a reliable test when using it with young L2 learners, and to what extent the lexical frequencies and cognate similarity between the English words and their Dutch equivalents found in Experiment 1 could explain primary-school pupils' test performance. Our first hypothesis was that the reliability of the PPVT-4 would increase when administering it to pupils who had more experience with English. Our second hypothesis was that children would perform better on items with more form overlap with their Dutch translations. The third hypothesis was that L1 frequency would positively influence these young pupils' performance.

5.3.1 Method

5.3.1.1 Participants

Four early-English primary schools that had at least eight years of experience with teaching English participated and four mainstream (i.e. monolingual) schools matched on area (urbanized or rural), neighbourhood (in terms of average income), religious denomination and educational philosophy. Data were collected from 204 typically developing pupils in one of three age groups: 4-to-5 years old (first grade), 8-to-9 years old (fifth grade), or 11-to-12 years old (final grade). An additional 37 pupils participated, but their data were removed from the analyses for one of the following reasons: they had another home language than Dutch (N = 23), they did not complete the PPVT (N = 8), were learning another language by means of an extracurricular programme (N = 1), they were diagnosed with dyslexia (N = 4), or they clearly had trouble concentrating during testing (N = 1).

The Dutch school system has eight grades, of which the first two are comparable to kindergarten. After the final grade (i.e., grade 8), pupils transfer to secondary school. Pupils are on average 12 years old by then. Pupils from mainstream schools started their English education in the penultimate grade (i.e., at around age 10), and would have had approximately 60 hours of English by the end of primary school. Pupils from early-English schools had had about 320 hours of English education by the end of primary school (Jenniskens et al., 2017). In both types of schools, all the rest of the teaching time is in Dutch. Parents gave informed consent for participation and were also asked to complete a questionnaire about out-of-school exposure to English; the response rate was relatively low (37.2%). Table 5.1 provides the number of girls/boys, exact ages and out-of-school exposure for the type of school and age groups. An independent samples t-test revealed that there was no difference between early-English and mainstream pupils in age ($t(202) = 0.208$; $p > .05$) or in out-of-school exposure to English ($t(74) = -1.11$; $p > .05$), but given the low response rate to the questionnaire these results should be interpreted with caution. The PPVT data we present here were previously reported as part of the findings from a larger battery of tests (Goriot, Broersma, McQueen, Unsworth, & Van Hout, 2018).

Table 5.1
Experiment 2: Number of Girls, Boys, Mean Age and Mean Out-of-school Exposure to English per Age Group

	Mainstream schools			Early-English schools		
	4-5 year-olds	8-9 year-olds	11-12 year-olds	4-5 year-olds	8-9 year-olds	11-12 year-olds
N	38	34	26	40	38	28
Girls (N)	21	17	16	23	19	19
Boys (N)	17	17	10	17	19	9
Age (*M, SD*)	4.90 (0.27)	9.08 (0.40)	12.15 (0.53)	4.81 (0.35)	8.95 (0.41)	11.97 (0.38)
Out-of-school exposure to English in hours per week *(M, SD)*	5.09 (6.08)	9.27 (6.35)	16.10 (14.11)	9.14 (7.97)	7.44 (6.41)	22.50 (16.43)

5.3.1.2 Instruments

PPVT-4. All children took form A of the Peabody Picture Vocabulary Task – 4th edition (Dunn & Dunn, 2007). Pictures were presented on a computer screen. The computer played the accompanying recording, consisting of one word per trial, recorded in a soundproof booth, pronounced in isolation, in clear citation style by a male native speaker of UK English. Administration rules as stated in the manual were followed. Since items in the higher sets (set 15 and higher) were responded to by only a small number of children, too little data were available from these sets to include them in the analyses. We therefore decided to analyse only the first 14 sets, that is, 168 items in total.

5.3.1.3 Procedure

The PPVT-4 was administered as part of a larger test battery. All children were tested individually in a quiet room at their school. The administration of the complete test battery took place in two sessions, each of which lasted approximately 30 minutes. The PPVT-4 was administered in the first session. Administration took between 5 and 20 minutes.

5.3.1.4 Analysis

First, Cronbach's alpha was computed, to determine the reliabilities for each of the age groups. Cronbach's alpha shows whether all items in one or multiple sets measure the same construct. Second, to examine whether frequency and form overlap played a role in pupils' performances on the PPVT-4 we performed generalized linear mixed-effects model analysis (packages lme4, lmerTest, in platform R, version 3.4.1).

5.3.2 Results

5.3.2.1 Reliability

We computed pupils' raw score on the PPVT-4, shown in Table 5.2 for pupils from early-English and mainstream schools separately. For 4-5- and 8-9-year-

olds the difference in English vocabulary scores between mainstream and early-English pupils is small. For 11-12-year-olds, the difference is larger.

Table 5.2
Experiment 2: Vocabulary Scores by Age Group and Type of Education

		4-5 year-olds		8-9 year-olds		11-12 year-olds	
		Main-stream	Early-English	Main-stream	Early-English	Main-stream	Early-English
N		38	40	33	38	22	21
PPVT-4 raw score	*M (SD)*	13.4 (8.3)	18.2 (14.8)	53.5 (22.7)	56.1 (16.8)	84.1 (21.3)	101.1 (21.0)

Table 5.3 shows Cronbach's alpha for the different age groups separately. We calculated Cronbach's alpha over a cumulative number of items, starting with the pupils that only completed the first set of items, thereafter also including pupils that completed the second set, and so on. In this way, we can see what the reliability is at lower proficiency levels, and whether reliability systematically increases when more sets are included. Indeed, as predicted, the higher reliability scores are found in the older age groups, particularly in the 11-12-year-olds. The largest increases in reliability values are found when including more sets for the 4-5-year-olds. Below set 4, the reliability scores are low (<.600) to medium (<.800), meaning that the PPVT-4 does not produce reliable scores ($\alpha \geq .900$; McNamara, 2000) in the lowest scoring group of L2 learners.

Table 5.3
Experiment 2: Cronbach's Alpha for the Different Age Groups and Sets

	4-5 year-olds		8-9 year-olds		11-12 year-olds	
	Alpha	Cumulative N (pupils)	Alpha	Cumulative N (pupils)	Alpha	Cumulative N (pupils)
Set 1	-4.085	5	-	-	-	-
Set 1 – 2	.431	31	-	-	-	-
Set 1 – 3	.784	46	-	1	-	-
Set 1 – 4	.834	66	.875	3	-	-
Set 1 – 5	.863	73	.888	11	.904	2
Set 1 – 6	.886	75	.846	15	.904	2
Set 1 – 7	.916	78	.807	38	.904	2
Set 1 – 8			.816	41	.903	4
Set 1 – 9			.873	48	.879	5
Set 1 – 10			.909	56	.956	9
Set 1 – 11			.944	66	.940	23
Set 1 – 12			.948	69	.933	29
Set 1 – 13			.950	70	.955	35
Set 1 – 14			.954	71	.962	43

Table 5.4 shows that percentages correct for all age groups are positively correlated with cognate status as measured by PhonSim: the closer the English word and its Dutch translation are, the larger the percentage correct, and thus the easier the item. Frequencies in Dutch and English are also positively related to percentages correct, except in the youngest age group.

Table 5.4
Experiment 2: Correlations between Phonological Similarity, Frequencies, and Percentage Correct

		1	2	3	4	5	6
1.	PhonSim	1.					
2.	Freq. EN	.037	1.				
3.	Freq. NL	.094	.754**	1.			
4.	Overall Percentage correct	.569**	.139	.233**	1.		
5.	Percentage correct age 4-5	.490**	-.047	.111	.724**	1.	
6.	Percentage correct age 8-9	.476**	.187*	.251**	.825**	.696**	1.
7.	Percentage correct age 11-12	.475**	.346**	.377**	.907**	.493**	.757**

*$p < .05$, **$p < .01$

5.3.2.2 Generalized Linear Mixed Effects Model Analysis

We conducted a generalized linear mixed-effects model analysis with binomial responses (incorrect coded as 0 and correct coded as 1) as dependent variable, Item number, Subject, and School as random effects, and Type of Education, Age Group, English Frequency, Dutch Frequency, PhonSim, and the interactions between Age Group and all other variables as fixed effects. Continuous variables (English Frequency, Dutch Frequency, and PhonSim) were mean-centered.

The results, displayed in Table 5.5, show that there is a main effect of Age Group: the 8-9- and 11-12-year-olds performed better than the 4-5-year-olds. There is also a positive main effect of Dutch frequency. Early-English pupils tended to perform better on the PPVT-4 than mainstream pupils, resulting in a marginally significant main effect of Type of Education (p = .058). The effects of English frequency and PhonSim were different for 8-9-year-olds and 11-12-year-olds than for 4-5-year-olds (see Figure 5.4 and 5.5, respectively), as shown by the interaction effects between Frequency and Age Group, and PhonSim and Age Group.

Table 5.5
Experiment 2: Parameter Estimates from the Model for Primary-School Pupils' Vocabulary Scores

	Fixed effects		
Parameters	Estimate	SE	Z-value
Intercept	-1.83	0.21	-8.80***
8-9 year-olds[a]	1.34	0.18	7.60***
11-12 year-olds[a]	2.07	0.19	10.80***
Type of Education[b]	0.33	0.18	1.89(*)
English frequency	-0.14	0.14	-1.01
Dutch frequency	0.32	0.15	2.17*
PhonSim	1.45	0.59	2.45*
8-9 year-olds×Type of Education	-0.32	0.21	-1.52
11-12 year-olds×Type of Education	0.16	0.23	0.72
8-9 year-olds×En. Freq.	0.21	0.08	2.60**
11-12 year-olds×En. Freq.	0.39	0.08	4.63***
8-9 year-olds×Dutch Freq.	0.04	0.08	0.49
11-12 year-olds×Dutch Freq.	0.06	0.08	0.71
8-9 year-olds×PhonSim	2.34	0.32	7.24***
11-12 year-olds×PhonSim	3.01	0.36	8.26***

[a] The 4-5 year-olds are the reference group, [b] Mainstream education is the reference group (*)p < .06, *p < .05, **p < .01, ***p < .001

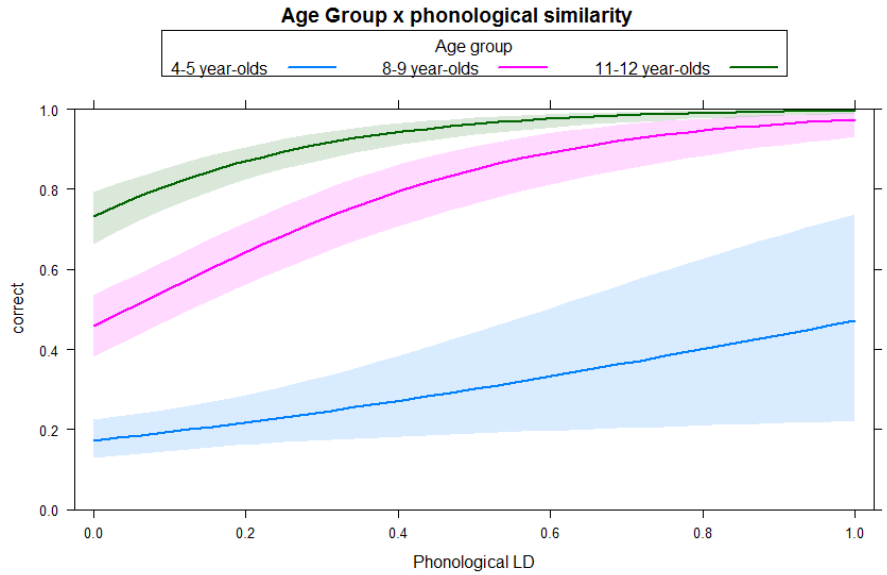


Figure 5.4. Experiment 2: Relation between PhonSim and PPVT scores for the different age groups.

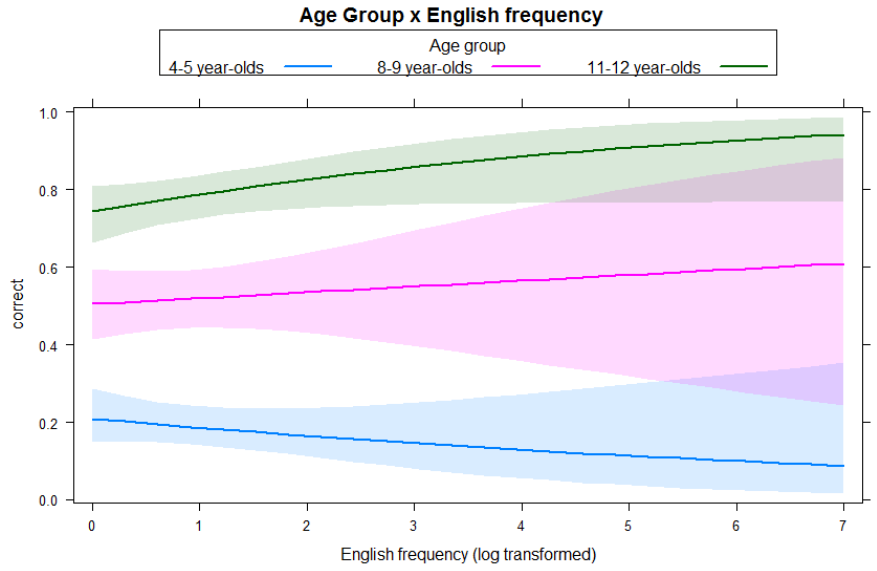


Figure 5.5. Experiment 2: Relation between English frequency and vocabulary scores, for the different age groups.

To gain insight in the effects of frequencies and cognate status on the performance of pupils of different ages, we conducted follow-up analyses for the three age groups separately. For each of the age groups, we fitted the same model as

reported previously (but without Age Group as a predictor). The results are shown in Table 5.6. Type of Education only showed a significant main effect for 11-12-year-olds: early-English pupils performed better than mainstream pupils. Similarly, English frequency was a significant and positive predictor of correct responses in the highest age group only. Dutch Frequency on the other hand was a positive and significant predictor of correct responses in the 8-9-year-olds and the 11-12-year-olds; in the youngest group, Dutch Frequency showed a trend towards significance (p = .057). Phonological distance was a positive and significant predictor in all three age groups.

Table 5.6
Experiment 2: Parameter Estimates for the Models for Primary-School Pupils from Different Age Groups

	4-5 year-olds			8-9 year-olds			11-12 year-olds		
	Fixed effects			Fixed effects			Fixed effects		
Parameters	Estimate	SE	Z-value	Estimate	SE	Z-value	Estimate	SE	Z-value
Intercept	-1.28	0.27	-4.39***	-0.37	0.17	-2.26*	0.29	0.25	1.67
Type of Education	0.29	0.24	1.23	0.01	0.12	0.12	0.56	0.27	2.12*
English frequency	-0.12	0.14	-0.91	0.04	0.13	0.31	0.30	0.15	1.97*
Dutch frequency	0.24	0.13	1.90(*)	0.34	0.13	2.55*	10.44	0.16	2.71**
PhonSim	2.84	0.58	4.88***	3.68	0.54	6.85***	4.41	0.65	6.75***
AIC	3546.6			7214.5			5692.9		

(*)p < .10, *p < .05, **p < .01, ***p < .001

5.3.3 Discussion

Confirming our first hypothesis, this experiment showed that the reliability of the PPVT-4 was quite low for pupils who completed only the first three sets (59% of the 4-5-year-olds), but became increasingly higher for pupils who reached the higher sets. Second, we found that, confirming our second hypothesis, pupils performed better on items that sounded more similar to their Dutch translations. Third, we found that pupils performed better on items that were more frequent in Dutch, whereas English frequency was a positive predictor of 11-12-year-olds' performance only. These findings partly confirm our third hypothesis, namely that for young pupils Dutch frequency is an important predictor of performance. Contrary to the hypothesis, for 4-5-year-olds, Dutch frequency only played a marginally significant role. We will elaborate on this finding in the general discussion.

Given that most of the children did not respond to any items from set 15 onwards, it was not possible to draw any conclusions about the performance on later items in the test, or on the possible frequency and similarity effects of those items. To investigate what the role of L1 effects in L2 vocabulary testing is in older children who have more experience with the English language, as well as to investigate the quality of the remaining items in the PPVT-4, we collected data from secondary-school pupils and ran the same analyses.

5.4 Experiment 3: secondary-school pupils

We replicated the primary-school experiment (Experiment 2) with secondary-school pupils. Based on the results of Experiment 2, our first hypothesis was that the reliability of the PPVT-4 would be high when administering it to secondary-school pupils. The second hypothesis was that, again, pupils' performance would be positively influenced by phonological similarities between item-translation pairs. Finally, our third hypothesis was that, contrary to primary-school pupils, the English frequency of the items would positively influence the performance of secondary-school pupils, whereas the influence of Dutch frequency would decrease.

5.4.1 Method

5.4.1.1 Participants

One school participated that had provided both a bilingual and a mainstream (i.e. monolingual) curriculum for six years. Data were collected from 152 pupils, who were in the first year (12-13-year-olds; 14 female, 21 male), second year (13-14-year-olds; 27 female, 26 male), or third year (14-15-year-olds; 35 female, 32 male). All pupils were in the pre-university track (Dutch: "VWO"), which upon completion gives admission to university education (the Dutch educational system is selective, with three secondary-school tracks). Pupils had been following either the mainstream or the bilingual curriculum since the start of secondary school (see Table 5.7): those in the mainstream curriculum followed their lessons in Dutch, but had lessons on English as a foreign language for 150 (12-13-year-olds) or 120 minutes per week (13-14- and 14-15-year-olds). Pupils in the bilingual curriculum received half of their subject lessons (approximately 800 minutes per week) in English. All pupils gave informed consent for participation.

Table 5.7
Experiment 3: Number of Pupils in the Mainstream and the Bilingual Curriculum

Year	Mainstream curriculum	Bilingual curriculum
1	15	20
2	2	51
3	18	46
Total	35	117

5.4.1.2 Procedure

Pupils were tested individually in a quiet room in the school during school time. All pupils performed the PPVT-4, which was part of a larger test battery (but a different test battery than in Experiment 2). The administration procedure was slightly different from the procedure in Experiment 2: all pupils started with the first set instead of with the age-appropriate set. Contrary to Experiment 2, for part of the pupils, printed pictures were shown, and words were read by the experimenter. The other pupils were presented with a computerized version of the PPVT-4 similar to the one used in Experiment 2. Similar to Experiment 2, testing stopped when pupils made eight or more errors, as indicated in the manual. Administration of the PPVT-4 took approximately 20 minutes.

5.4.1.3 Analysis

In a similar fashion to Experiment 2, we examined Cronbach's alpha and the percentage correct to investigate the reliability of the different sets, and the difficulty of the items. Again, a generalized linear mixed-effects analysis was conducted in order to investigate effects of L1 and of Type of Education on performance.

5.4.2 Results

Figure 5.6 shows the scores on the PPVT-4, both for primary-school (Experiment 2) and secondary-school pupils (Experiment 3). It shows that, descriptively, early-English pupils have higher scores than mainstream pupils, and older pupils have higher scores than younger pupils. Table 5.8 shows that the reliability of the sets is excellent in all three secondary-school years.

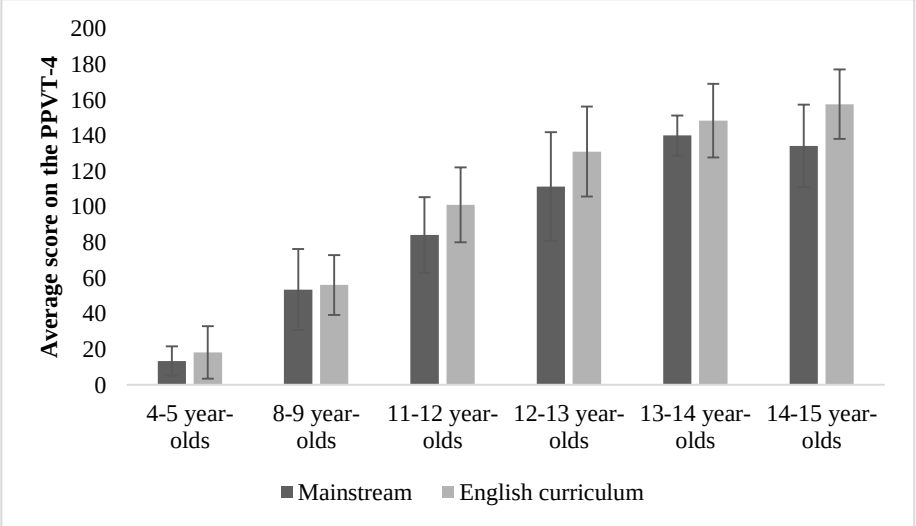


Figure 5.6. Experiments 2 and 3: Average scores (raw number) on the PPVT-4 with standard deviations.
Note: In the 13-14 year-olds, the mainstream group consists of only two pupils.

Table 5.8
Experiment 3: Cronbach's Alpha for the Different Age Groups and Sets

	12-13 year-olds (Year 1)		13-14 year-olds (Year 2)		14-15 year-olds (Year 3)	
Set	Alpha	Cumulative N (pupils)	Alpha	Cumulative N (pupils)	Alpha	Cumulative N (pupils)
Set 1	-	-	-	-	-	-
Set 1 – 2	-	-	-	-	-	-
Set 1 – 3	-	-	-	-	-	-
Set 1 – 4	-	-	-	-	-	-
Set 1 – 5	-	-	-	-	-	-
Set 1 – 6	-	-	-	-	-	-
Set 1 – 7	-	-	-	-	-	-
Set 1 – 8	-	-	-	-	-	-
Set 1 – 9	-	-	-	-	-	-
Set 1 – 10	.968	2	-	-	-	-
Set 1 – 11	.956	9	.942	4	.943	3
Set 1 – 12	.952	10	.942	4	.913	4
Set 1 – 13	.946	15	.942	6	.940	6
Set 1 – 14	.942	21	.950	9	.932	10
Set 1 – 15	.963	28	.950	30	.943	32
Set 1 – 16	.964	29	.942	40	.942	42
Set 1 – 17	.971	33	.946	49	.955	54
Set 1 – 18	.974	35	.950	51	.960	61
Set 1 – 19			.956	53	.967	64

Figure 5.7 shows the percentage correct per set, for each age group, again both for primary-school (Experiment 2) and secondary-school pupils (Experiment 3). For the secondary-school pupils (12- to 15-year-olds), the first sets seem to be easy with performance at (or near) ceiling. The percentage correct declines steeply in the later sets.

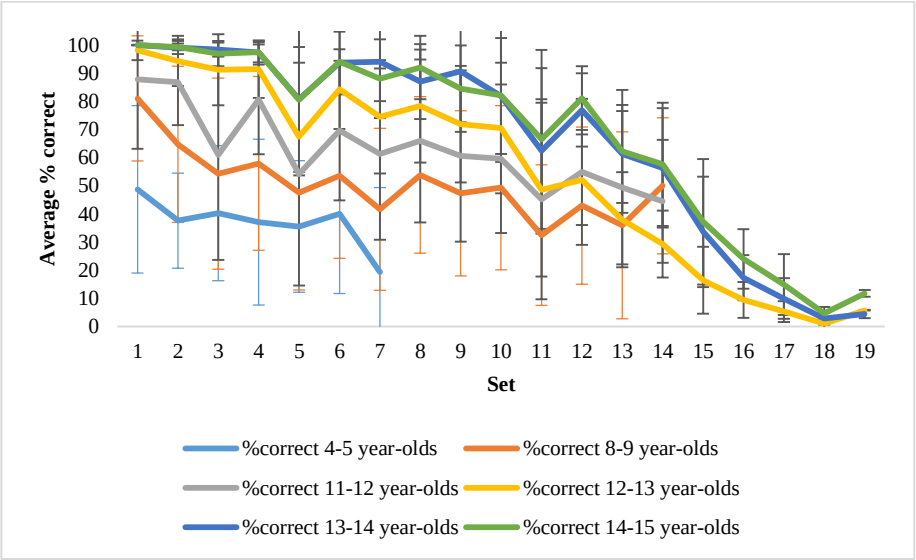


Figure 5.7. Experiments 2 and 3: Percentages correct per set, for each age group (with *SD*s).

We correlated the percentages correct for the different year groups with PhonSim, the Dutch frequencies, and the English frequencies (Table 5.9). The results were highly similar to the results in the younger children: Percentage correct correlated significantly with PhonSim and with frequencies in English and in Dutch.

Table 5.9
Experiment 3: Correlations between Phonological Similarity, Frequencies, and Percentage Correct

	1	2	3	4	5	6
1. PhonSim	1					
2. Freq. EN	.124	1				
3. Freq. NL	.132*	.759***	1			
4. Overall percentage correct	.236***	.634***	.543***	1		
5. Percentage correct 12-13 year-olds	.256***	.661***	.580***	.970***	1	
6. Percentage correct 13-14 year-olds	.210**	.615***	.534***	.998***	.949***	1
7. Percentage correct 14-15 year-olds	.227***	.601***	.520***	.982***	.940***	.986***

*p < .05, **p < .01, ***p < .001

5.4.2.1 Generalized Linear Mixed-Effects Analysis

We performed a general linear mixed-effects analysis on pupils' performance (0 as incorrect and 1 as correct) on the items of the PPVT-4, with Age Group, Type of Education, English Frequency, Dutch Frequency, PhonSim, and the interactions between Age Group and the other variables. Random slopes were included at the Subject and Item level. Pupils in the 13-14-year-old age group were left out of the analysis, because of the low number of participants (N = 2) in the

mainstream curriculum. The results, in Table 5.10, show that there are main effects of all independent variables except for Dutch Frequency. Older pupils perform better on the items in the PPVT-4 than younger pupils. Pupils perform better on items that are more frequent in English, and on items that sound more similar to their Dutch translations. Pupils in the bilingual curriculum outperformed pupils in the mainstream curriculum; the differences between pupils from the mainstream and the bilingual curriculum were smaller in 12-13- than in 14-15-year-olds (see Figure 5.8), resulting in a significant interaction.

Table 5.10
Experiment 3: Parameter Estimates for the Model for Secondary-school Pupils' Vocabulary Scores

Parameters	Fixed effects		
	Estimate	SE	Z-value
Intercept	0.02	0.14	0.11
14-15 year-olds[a]	0.85	0.07	11.68***
Type of Education[b]	0.68	0.07	9.79***
English frequency	1.10	0.14	7.98***
Dutch frequency	0.20	0.15	1.37
PhonSim	4.70	0.59	7.94***
14-15 year-olds×Type of Education	0.41	0.09	4.47***
14-15 year-olds×En. Freq.	0.03	0.05	0.60
14-15 year-olds×Dutch Freq.	0.07	0.05	1.28
14-15 year-olds×PhonSim	0.10	0.21	0.47
AIC	14718.2		

[a] The 12-13 year-olds are the reference category, [b] Mainstream education is the reference category
***$p < .001$

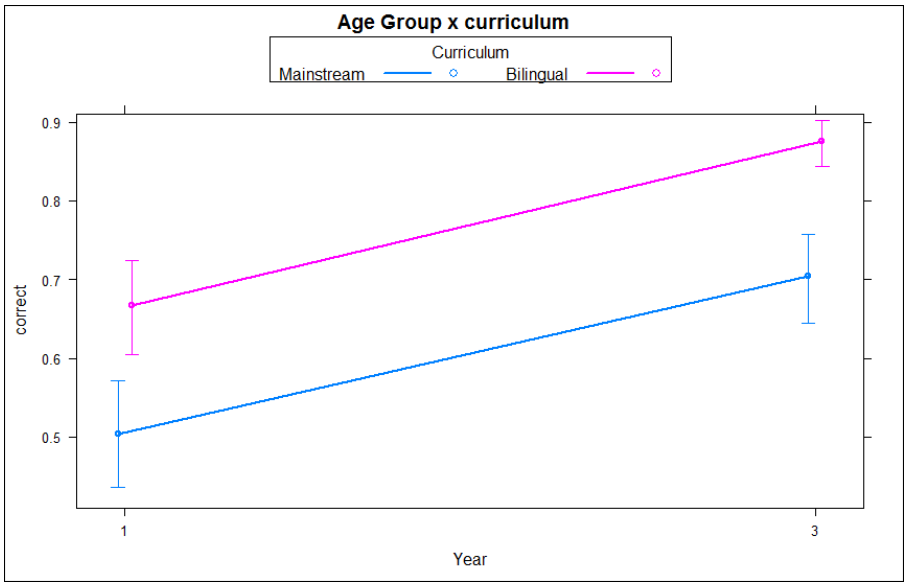


Figure 5.8. Experiment 3: The relation between age and correct scores, for pupils from the two curricula.

To investigate the effects of the independent variables on the performance of pupils in the different age groups separately, we performed the same analysis again but now for each age group separately (without Age Group as a predictor). The outcomes are shown in Table 5.11. For 14-15-year-old pupils, there was a main effect of Type of Education in favour of pupils in the bilingual curriculum. For the 12-13-year-olds this effect was only marginally significant (p = .057). In both age groups, English frequency and PhonSim were significant and positive predictors of correct responses. Dutch frequency was never significant, although it seemed that the oldest pupils performed better on items that were more frequent in Dutch (p = .093).

Table 5.11
Experiment 3: Parameter Estimates for the Models for Secondary-School Pupils from Different Age Groups

	12-13 year-olds Fixed effects			14-15 year-olds Fixed effects		
Parameters	Estimate	SE	Z-value	Estimate	SE	Z-value
Intercept	0.05	0.35	0.16	1.04	0.32	3.21**
Type of Education	0.79	0.42	1.90(*)	1.33	0.33	4.05***
English frequency	1.33	0.16	8.24***	1.36	0.18	7.73***
Dutch frequency	0.24	0.17	1.38	0.32	0.19	1.68(*)
PhonSim	5.52	0.69	7.95***	5.64	0.76	7.42***
AIC	5188.5			7957.7		

(*)$p < .10$, *$p < .05$, **$p < .01$, ***$p < .001$

5.4.3 Discussion

Confirming our first hypothesis that the reliability of the PPVT-4 would be high when administering it to secondary-school pupils who have more experience with English, we found that the reliability of the PPVT-4 was always ≥.900. Our second hypothesis, that pupils would perform better on English items that are phonologically closer to their Dutch translations, was also confirmed. Our third hypothesis was that English frequency would positively influence secondary-school pupils' performance, and that, contrary to primary-school pupils, the role of Dutch frequency would be decreasing. The results confirmed the hypothesis in the sense that pupils performed better on items that were more frequent in English, while Dutch frequency was not a significant predictor of pupils' performance. Contrary to our expectations Dutch frequency did however show a trend towards significance in the oldest age group. We will come back to this finding in the general discussion.

5.5 General Discussion

The first aim of this study was to examine the reliability of the PPVT-4 as a tool for measuring L2 vocabulary knowledge of Dutch learners of English. The second aim was to examine whether L1 characteristics, in particular lexical frequency and cognate similarity (operationalised by phonological similarity) of the item-translation pairs, could affect scores on the PPVT-4 when using it to measure L2 vocabulary. We investigated this question by administering the PPVT-4 to primary-school and secondary-school pupils of different age groups who were learning English as an L2. We investigated the technical quality of the test, and the characteristics of the individual items. We also examined the relation between L1 and L2 word frequency and cognate status on the one hand, and primary and secondary-school pupils' vocabulary scores on the other.

Because English and Dutch are both Germanic languages, we expected that there would be relatively strong form overlap in the English-Dutch item-translation pairs. We also expected that Dutch and English lexical frequency of the PPVT-4 items would be closely related to each other (Moscoso del Prado Martín et al., 2004). The results of Experiment 1 showed that this was indeed the case: English and Dutch lexical frequency correlated positively and highly with each other, but were not identical. Furthermore, there was ample variation in the similarity of the two items' frequencies across translation pairs. There was also substantial phonological similarity between the English items in the PPVT-4 and their Dutch translations, thereby confirming previous results that showed that English and Dutch share many cognates (Schepens et al., 2013).

Our first hypothesis was that the PPVT-4 would be a more reliable vocabulary measure when administering it to pupils who have more experience with English than when administering it to less experienced pupils. The latter group was expected to complete only the first few sets of the PPVT-4, and we expected that the test would not reliably differentiate between pupils' scores based on a limited number of items. Indeed, for 4-5-year-olds who only made it to the first three sets, Cronbach's alpha was low ($\alpha < .600$) to medium ($\alpha < .800$). The test reached an acceptable reliability level ($\alpha > .900$; McNamara, 2000) when pupils completed more sets.

Our second hypothesis was that phonological similarity between the English words and their Dutch translations would positively influence pupils scores on the PPVT-4, as a cognate effect had already been shown for Spanish children (Potapova et al., 2016). Moreover, we expected that this effect would be larger for older pupils, as previous research had shown that older Dutch-Frisian bilingual children are better at recognizing overlap between words in two languages than younger children (Bosma et al., 2016). Extending previous research (Bosma et al., 2016; Pérez et al., 2010; Potapova et al., 2016), we showed that both for younger and older pupils, phonological similarity between English and Dutch words was a positive predictor of pupils' performance. As expected, this effect was larger for older pupils than for younger pupils. This suggests that older pupils may be more able than younger pupils to make use of phonological similarities between item-translation pairs for the comprehension of L2 words.

We assessed phonological rather than orthographic similarity between the English and Dutch words. This means that the extent to which orthographic similarity predicts pupils' L2 vocabulary scores remains unknown. We believe that our choice to investigate phonological similarity is theoretically well-grounded, for two reasons. The first is that in the PPVT-4, children are typically presented with the words orally. The second reason is that pupils are probably more familiar with the oral than with the written form of words, if they are familiar with the written forms at all: the focus of English lessons in primary school is mostly on oral skills (Jenniskens et al., 2017; Thijs, Trimbos, Tuin, Bodde, & de Graaff, 2011), and, in secondary school, pupils receive content and language integrated learning (CLIL) contexts in the bilingual programme, in which the focus is on meaning and not form (Dalton-Puffer, 2011). In other words, whilst we cannot rule out a possible influence of orthographic similarity in addition to phonological similarity, it is unlikely that the former would be greater than the latter.

Our third hypothesis was that the influence of L1 (Dutch) frequency would decrease in older, more experienced pupils, while the influence of L2 (English) frequency on pupils' scores would increase. Even within early-English educational programmes, primary-school pupils are generally exposed to English for maximally one hour per week (Jenniskens et al., 2017). Since pupils would not have had enough English exposure for English frequency to matter a great deal, we expected that pupils would mainly rely on their knowledge of the items in Dutch, and thus the Dutch frequency of the words would be a better predictor of their vocabulary scores. Secondary-school pupils would be more exposed to English, both inside and outside the school (Lindgren & Muñoz, 2013), and hence English frequency may play a more important role in this group. Our hypothesis was confirmed: Dutch frequency was positively related to primary-school pupils' performance on the PPVT-4, whereas for secondary-school pupils it was not. English frequency was only a significant predictor of scores for the oldest (i.e., 11-12-year-old) primary-school pupils and for the pupils in secondary school. Wood and Pena (2015) showed that the relation between children's errors and difficulty level of the items in the PPVT-4 (as measured by the ordering of the items), was stronger for English L1 children than for Spanish children learning English as an L2. We extended these findings by showing that English L2 learners' performance on the PPVT seems to depend, in the least experienced group, on the items' frequency in their L1, rather than on the frequency in the L2. For 4-5-year-olds, who are in this study also the least experienced, Dutch frequency was however only marginally significant. It is possible that because these pupils completed relatively few items, there may not have been enough variation in the Dutch lexical frequencies to show a significant relation with their performance on the PPVT in English. Furthermore, in the oldest, and in this study the most experienced group, pupils seemed to perform better on items that had a higher frequency in Dutch, although this relation was not significant. It may be that as these children encountered the low-frequent words at the end of the test, their performance depended on whether they knew the word in Dutch in the first place.

We also asked whether pupils following an English curriculum at school would differ in their performance on the PPVT-4 from their peers who followed the mainstream curriculum. We indeed found a difference in favour of the pupils in the English curriculum, but only in the older pupils. For the 4-5-year-olds and 8-9-year-olds, there was no significant difference between pupils from the two types of education. Several reasons could account for the absence of this difference. The first reason could be that there really are no differences between the groups. Although previous research has shown that early-English pupils in kindergarten (Unsworth et al., 2015) and in the final grade of primary school (de Graaff, 2015) outperformed their peers from mainstream schools as a group, it has also been shown that the performance scores of the groups overlapped, such that individual pupils from mainstream schools outperformed individuals from early-English schools on English proficiency tests (De Graaff, 2015). Note that in the latter study, pupils were not tested on vocabulary but on spelling, listening and reading skills, and on 'use of English'. Furthermore, those two previous studies have shown that, besides early-English versus mainstream education, the development of English is related to other factors, such as the English proficiency level of the teacher, the amount of English

input in school (Unsworth et al., 2015), and out-of-school-exposure to English (de Graaff, 2015). It may be the case that the early-English participants in our study got less than one hour of English input in school or that they were educated by a teacher with a moderate proficiency level of English, two factors that have been shown to result in lower vocabulary scores compared to pupils who receive more input in English, or who are educated by a (near-)native speaker of English (Unsworth et al., 2015). In addition, the response rate to the parental questionnaire was very low, making it hard to draw conclusions about out-of-school exposure to English. It may be the case that the mainstream pupils received more out-of-school exposure to English than the early-English pupils, which may have compensated for the lack of English instruction at school.

The second reason for the absence of a significant difference between pupils from the mainstream and early-English curriculum might be the low reliability of the PPVT-4, especially in the 4-5 year olds. In previous research, differences between early-English and mainstream kindergartners were small (Unsworth et al., 2015). It may thus be that there are in fact differences between the pupils of the two types of education in our study, too, but that these subtle differences do not always show when administering the PPVT-4 as outlined in the test manual. We therefore suggest, when administering the PPVT, to start with an earlier set than the age-appropriate set.

5.5.1 Implications for research and practice

Our study has shown that L1 lexical frequency and cognate similarity (operationalised as phonological similarity) influenced Dutch pupils' performance on the PPVT-4 as a measure of L2 English vocabulary. This is reminiscent of findings with adults that show that acquiring a new language is easier when that language is close to your native language (Schepens, van der Slik, & van Hout, 2016), and that linguistic similarity helps learners to derive the meaning of foreign words (Pérez et al., 2010; Potapova et al., 2016). Related languages overlap by definition, and are thus very likely to contain translation equivalents that also share aspects of their form. Previous research with adults has already shown that L2 learners with different mother tongues obtain significantly different scores, depending on the number of cognates between their L1 and the words in the PPVT (Leśniewska et al., 2018). Dutch and English are relatively close to each other, having a relatively large number of cognates (Lindgren & Muñoz, 2013; Schepens et al., 2013). Since effects of L1-L2 similarities are different for different L1s, researchers should therefore be cautious in comparing receptive English vocabulary knowledge of children across L1s, as is sometimes done (Enever et al., 2011; Lindgren & Muñoz, 2013; Steinlen & Piske, 2013). Our findings suggest that Dutch children will perform better on the PPVT-4 than children with a non-Germanic language as mother tongue because of the higher proportion of cognate items for the Dutch children.

When investigating the question whether children following an English programme at school have better knowledge of English vocabulary than pupils who do not follow such a programme, researchers should ideally make use of a curriculum-independent vocabulary test that is able to capture English vocabulary knowledge that pupils learned in school. Since such a test does not exist, in any case

not for the Dutch context, the PPVT-4 is often used to answer the question whether early-English pupils have a better developed English vocabulary than pupils enrolled in mainstream schools. We investigated whether it is suitable to use the PPVT-4 to answer this question. As many as 31 out of 78 (40%) 4-5-year-old participants completed no more than the first two sets of the PPVT-4. The reliability of these two sets was very low, which suggests that the PPVT-4 may not be suitable for use with unexperienced (or younger) L2 learners. The higher values for the older groups suggest that the test is more reliable when using it with more experienced (or older) L2 learners. Nevertheless, effects that were unaccounted for in the design of the test, such as cognate status and L1 word frequency, still play an important role in these groups. Researchers should thus be cautious in interpreting the results.

This study had a cross-sectional design. Including pupils from six different age groups has provided us with insight in what factors may play a role in L2 vocabulary testing at different ages. We cannot be certain, however, about the (causal) developmental pattern of the L1 and L2 factors at play. A longitudinal experiment would provide more insight in the relation between the extent to which Dutch and English lexical frequency and cognate status predict performance on the PPVT-4 as pupils grow older. Further, including a group of very young learners from a bilingual programme may reveal whether the test becomes a more reliable tool to test English vocabulary in young learners when they have more knowledge of English.

5.6 Conclusion

We investigated the use of the PPVT-4 as a measure of receptive English vocabulary in L2 learners. We found that in young primary-school pupils, the frequency of the Dutch translations of the English items as opposed to the frequency of the English test words themselves positively related to their performance on the test. Both primary and secondary-school pupils performed better on English items that were phonetically closer to their Dutch translations. These findings indicate that pupils' L1 plays a role when assessing vocabulary in the L2. Researchers should be aware of these influences, especially when comparing pupils with different mother tongues. Nevertheless, the PPVT-4 seems to be a suitable curriculum-independent instrument for the relative ranking on L2 English vocabulary size of more experienced L2 learners with the same mother tongue.

5.7 References

Admiraal, W., Westhoff, G., & de Bot, K. (2006). Evaluation of bilingual secondary education in The Netherlands: Students' language proficiency in English. *Educational Research and Evaluation*, *12*, 75–93. http://doi.org/10.1080/13803610500392160

Albers, M. (2015). *Van Dale middelgroot woordenboek Nederlands* (Second). Utrecht: VBK Media.

Bialystok, E., Luk, G., Peets, K. F., & Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition*, *13*, 525–531. http://doi.org/10.1017/S1366728909990423

Bosma, E., Blom, E., Hoekstra, E., & Versloot, A. (2016). A longitudinal study on the gradual cognate facilitation effect in bilingual children's Frisian receptive

vocabulary. *International Journal of Bilingual Education and Bilingualism. Advance Online Publication.* http://doi.org/10.1080/13670050.2016.1254152

Brenders, P., van Hell, J. G., & Dijkstra, T. (2011). Word recognition in child second language learners: Evidence from cognates and false friends. *Journal of Experimental Child Psychology*, *109*, 383–396. http://doi.org/10.1016/J.JECP.2011.03.012

Broersma, M. (2009). Triggered Codeswitching between Cognate Languages. *Bilingualism: Language and Cognition*, *12*, 447–462. http://doi.org/10.1017/S1366728909990204

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*, 977–990. http://doi.org/10.3758/BRM.41.4.977

Buyl, A., & Housen, A. (2014). Factors, processes and outcomes of early immersion education in the Francophone community in Belgium. *International Journal of Bilingual Education and Bilingualism*, *17*, 178–196. http://doi.org/10.1080/13670050.2013.866627

Cohen, C. (2016). Relating input factors and dual language proficiency in French-English bilingual children. *International Journal of Bilingual Education and Bilingualism*, *19*, 296–313. http://doi.org/10.1080/13670050.2014.982506

Crevecoeur, Y. C., Coyne, M. D., & McCoach, D. B. (2014). English language learners and English-only learners' response to direct vocabulary instruction. *Reading & Writing Quarterly*, *30*, 51–78. http://doi.org/10.1080/10573569.2013.758943

Dahl, A., & Vulchanova, M. (2014). Naturalistic acquisition in an early language classroom. *Frontiers in Psychology*, *5*, 329. http://doi.org/10.3389/fpsyg.2014.00329

Dalton-Puffer, C. (2011). Content-and-language integrated learning: From practice to principles? *Annual Review of Applied Linguistics*, *31*, 182–204. http://doi.org/10.1017/S0267190511000092

de Graaff, R. (2015). Vroeg of laat Engels in het basisonderwijs: Wat levert het op? [Early or late English in primary education: Does it matter?]. *Levende Talen Tijdschrift*, *16*(2), 3–15.

Dijkstra, T., Miwa, K., Brummelhuis, B., Sappelli, M., & Baayen, H. (2010). How cross-language similarity and task demands affect cognate recognition. *Journal of Memory and Language*, *62*, 284–301. http://doi.org/10.1016/J.JML.2009.12.003

Dijkstra, T., & Van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, *5*, 175–197. http://doi.org/10.1017/S1366728902003012

Dongsun, Y., Yoon, S., & Jiyeon, L. (2016). Receptive vocabulary assessment in Korean-English bilingual children: cross-linguistic investigations. *Communication Sciences and Disorders*, *21*, 131–138. http://doi.org/10.12963/csd.14186

Dunn, L. M. (1959). *Peabody Picture Vocabulary Test*. Minneapolis: American Guidance Service.

Dunn, L. M., & Dunn, D. M. (2007). *PPVT-4 Manual*. Bloomington: NCS Pearson.

Enever, J., Krikhaar, E., Lindgren, E., Lopriore, L., Lundberg, G., Mihaljevic Djigunovic, J., … Tragant Mestres, E. (2011). *ELLiE: Early Language Learning in Europe*.

EP-Nuffic. (n.d.). Standaard tweetalig onderwijs Engels havo/vwo [Standard bilingual education English havo/vwo].

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*, 678–720. http://doi.org/10.1080/13670050802149283

Goriot, C., Broersma, M., McQueen, J. M., Unsworth, S., & Van Hout, R. (2018). Language balance and switching ability in children acquiring English as a second language. *Journal of Experimental Child Psychology*, *173*, 168–186.

Heemskerk, J. S., & Zonneveld, W. (2000). *Uitspraakwoordenboek*. Houten: Het Spectrum.

Huang, B. H. (2016). A synthesis of empirical research on the linguistic outcomes of early foreign language instruction. *International Journal of Multilingualism*, *13*(3), 257–273. http://doi.org/10.1080/14790718.2015.1066792

Jenniskens, T., Leest, B., Wolbers, M., Bruggink, M., Dood, C., & Krikhaar, E. (2017). *Zicht op vroeg vreemdetalenonderwijs [Insight into early foreign language education]*. Nijmegen.

Jensen, S. (2017). Gaming as an English language learning resource among young children in Denmark. *Calico Journal*, *34*(1), 1–19. http://doi.org/10.1558/cj.29519

Jimenez Catalan, R. M., & Terrazas Gallego, M. (2005). The receptive vocabulary of English foreign language young learners. *Journal of English Studies*, *5–6*, 173–191.

Keuleers, E., Brysbaert, M., & New, B. (2010). SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles. *Behavior Research Methods*, *42*, 643–650. http://doi.org/10.3758/BRM.42.3.643

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*, 149–174. http://doi.org/10.1006/jmla.1994.1008

Leśnieuwska, & Pichette, F. (2016). Songs vs. stories: Impact of input sources on ESL vocabulary acquisition by preliterate children. *International Journal of Bilingual Education and Bilingualism*, *19*, 18–34. http://doi.org/10.1080/13670050.2014.960360

Leśniewska, J., Pichette, F., & Béland, S. (2018). First language test bias? Comparing French-speaking and Polish-speaking participants' performance on the Peabody Picture Vocabulary Test. *Canadian Modern Language Review*, *74*, 27–52. http://doi.org/10.3138/cmlr.3670

Lindgren, E., & Muñoz, C. (2013). The influence of exposure, parents, and linguistic distance on young European learners' foreign language comprehension. *International Journal of Multilingualism*, *10*, 105–129. http://doi.org/10.1080/14790718.2012.679275

Lobo, V. R. (2013). *Teaching L2 English at a very early age : A study of Dutch schools*. Radboud University. Retrieved from

https://www.lotpublications.nl/teaching-l2-english-at-a-very-young-age-teaching-l2-english-at-a-very-young-age-a-study-of-dutch-schools

Longman. (2012). *Longman Dictionary of Contemporary English for Advanced Learners* (6th ed). Pearson Education.

McNamara, T. F. (Timothy F. (2000). *Language testing*. Oxford: Oxford University Press.

Merisuo-Storm, T. (2007). Pupils' attitudes towards foreign-language learning and the development of literacy skills in bilingual education. *Teaching and Teacher Education*, *23*, 226–235. http://doi.org/10.1016/j.tate.2006.04.024

Moscoso del Prado Martín, F., Bertram, R., Häikï, T., Schreuder, R., & Baayen, R. H. (2004). Morphological family size in a morphologically rich language: The case of Finnish compared to Dutch and Hebrew. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 1271. http://doi.org/http://dx.doi.org/10.1037/0278-7393.30.6.1271

Nikula, T. (2017). CLIL: A European approach to bilingual education. In N. Van Deusen-Scholl & S. May (Eds.), *Second and Foreign Language Education. Encyclopedia of Language and Education* (3rd ed., pp. 111–124). Cham: Springer. http://doi.org/10.1007/978-3-319-02246-8_10

Pérez, A. M., Peña, E. D., & Bedore, L. M. (2010). Cognates facilitate word recognition in young Spanish-English bilinguals' test performance. *Early Childhood Services*, *4*(1), 55–67. http://doi.org/10.1044/1092-4388(2013/12-0056)

Poarch, G. J., & van Hell, J. G. (2012a). Cross-language activation in children's speech production: Evidence from second language learners, bilinguals, and trilinguals. *Journal of Experimental Child Psychology*, *111*, 419–438. http://doi.org/10.1016/J.JECP.2011.09.008

Poarch, G. J., & van Hell, J. G. (2012b). Executive functions and inhibitory control in multilingual children: Evidence from second-language learners, bilinguals, and trilinguals. *Journal of Experimental Child Psychology*, *113*, 535–551. http://doi.org/10.1016/j.jecp.2012.06.013

Potapova, I., Blumenfeld, H. K., & Pruitt-Lord, S. (2016). Cognate identification methods: Impacts on the cognate advantage in adult and child Spanish-English bilinguals. *International Journal of Bilingualism*, *20*, 714–731. http://doi.org/10.1177/1367006915586586

Schepens, J., Dijkstra, T., Grootjen, F., & van Heuven, W. J. B. (2013). Cross-language distributions of high frequency and phonetically similar cognates. *PLoS ONE*, *8*(5). http://doi.org/10.1371/journal.pone.0063006

Schepens, J., van der Slik, F., & van Hout, R. (2016). L1 and L2 distance effects in learning L3 Dutch. *Language Learning*, *66*, 224–256. http://doi.org/10.1111/lang.12150

Steinlen, A., & Piske, T. (2013). Academic achievement of children with and without migration backgrounds in an immersion primary school: A pilot study. *Zeitschrift Für Anglistik Und Amerikanistik*, *61*, 215–244. http://doi.org/10.1515/zaa-2013-0303

Thijs, A., Trimbos, B., Tuin, D., Bodde, M., & de Graaff, R. (2011). *Engels in het basisonderwijs [English in primary education]*. Enschede.

Unsworth, S., Persson, L., Prins, T., & de Bot, K. (2015). An investigation of factors

affecting early foreign language learning in the Netherlands. *Applied Linguistics*, *36*, 527–548. http://doi.org/10.1093/applin/amt052

van der Leij, A., Bekebrede, J., & Kotterink, M. (2010). Acquiring reading and vocabulary in Dutch and English: The effect of concurrent instruction. *Reading and Writing*, *23*, 415–434. http://doi.org/10.1007/s11145-009-9207-5

van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *The Quarterly Journal of Experimental Psychology*, *67*, 1176–1190. http://doi.org/10.1080/17470218.2013.850521

Wells, J. C. (2008). *Longman Pronunciation Dictionary* (3rd ed). Pearson Education.

Wood, C., & Pena, V. (2015). Lexical considerations for standardized vocabulary testing with young Spanish-English speakers. *Contemporary Issues in Communication Science and Disorders*, *42*, 202–215. http://doi.org/1092-5171/15/4202-0202

Chapter 6: Discussion and conclusion

Chapter 6

Discussion and conclusion

The overall aim of this thesis was to shed light on the possible effects of early-English education on pupils' cognitive and linguistic development, and more specifically to investigate whether the limited bilingual experience offered in such programmes may change children's development in three different domains: executive functions, phonological awareness, and English speech perception. In addition, I also investigated whether L1 factors play a role in L2 vocabulary assessment when using a frequently used vocabulary measure: the PPVT-4. To investigate each of these questions, I compared groups of early-English pupils with groups of mainstream pupils, and in some studies also with simultaneous Dutch-English bilingual children. To investigate whether the relation between learning two languages and development was different for pupils of different ages, and with different levels of bilingual experience, pupils always came from three different age groups. In this closing chapter, the main findings of these studies, and implications for future research as well as for educational practice will be discussed.

6.1 Summary of the results

6.1.1 Executive functions

The main research questions in Chapter 2 were first, whether early-English pupils differ from mainstream pupils in the development of executive functions, and second, whether the balance between Dutch and English vocabulary plays a role in the development of executive functions. I administered four executive functioning tasks (measuring switching, inhibition, verbal and non-verbal working memory) to mainstream and early-English pupils of three different age groups: 4-5 year-olds (grade 1, start of primary school), 8-9 year-olds (grade 5, middle of primary school), and 11-12 year-olds (grade 8, end of primary school). The results of this study showed that there were no differences between early-English and mainstream groups in executive functions. Lexical balance, however, was positively related to both mainstream and early-English pupils' scores on the switching task. In other words, pupils with more balanced English and Dutch lexicons showed better switching performance than pupils with less balanced lexicons.

The results of this study are in line with those of previous studies, which found that more balanced bilingual children show better performance on executive functioning tasks than less balanced bilingual children (Blom, Küntay, Messer, Verhagen, & Leseman, 2014; Thomas-Sunesson, Hakuta, & Bialystok, 2018; Vega & Fernandez, 2011). These findings have been attributed to the fact that bilingual children, and especially L2 learners, have to constantly monitor competing activation between their two languages, a process that is argued to foster the development of executive functions (Blom et al., 2014; Gathercole et al., 2014; Vega & Fernandez, 2011). My findings mirror those of Vega & Fernandez (2011), who also observed a significant relation between language balance and switching but not inhibition. It may be that a relation between language balance and inhibition does not exist, or that it only shows when using a non-timed task, as young children tend to rush their responses on a timed task, resulting in sub-optimal performance

(Davidson, Amso, Anderson, & Diamond, 2006). With respect to working memory, the findings of this thesis partly reflect those of Blom et al. (2014), who also found no relation between language balance and non-verbal working memory; they did however find a relation with verbal working memory. As the participants in the study by Blom et al. were more experienced L2 learners than the participants in this research, I hypothesized that a relation between language balance and verbal working memory exists, but only in L2 learners who are more proficient in the L2.

There were no group differences between mainstream and early-English pupils' development of executive functions, a finding which was as expected for younger pupils, but not for older (eight-grade) pupils. Previous studies already suggested that a certain threshold of exposure to the L2 should be met before the development of L2 learners' executive functions starts to differ from that of (more) monolingual children. Carlson and Meltzoff (2008) for example, found that pupils enrolled in immersion education for only half a year did not differ from monolingual children in the development of executive functions, whereas simultaneous bilingual children did. Purić, Vuksanović, and Chondrogianni (2017) also found that a group with more exposure to the L2 (5 hours per day) outperformed a group with less exposure to the L2 (1.5 hours per day) on a working memory task. For the eighth-grade pupils, I expected that although the amount L2 exposure per week was low (approximately 60 minutes), after eight years of English education, the amount of exposure (320 hours) would have been enough for developmental differences to appear. For executive functions, this was not the case. However, the early-English pupils in final grade did obtain higher English vocabulary scores and comparable Dutch vocabulary scores compared to pupils in mainstream education. This suggests that early-English education indeed has an influence on pupils' development, at least in the domain of English vocabulary. In other words, devoting only a minimal part of the teaching time (15% at most) to English lessons may thus foster pupils' English vocabulary knowledge, but it is not enough to foster differences in pupils' development of executive functions.

The focus of Chapter 2 was the comparison between mainstream and early-English pupils. The bilingual experience of both groups was very limited. To know whether children with more bilingual experience show more developed executive functions than the groups with limited bilingual experience, a group of experienced (simultaneous) bilinguals should be included. Although this was my original intention, the nature of the research project restricted the possibilities to do so and the bilingual data was therefore not reported in Chapter 2. Furthermore, since all bilingual children had to be tested individually, and after school time, I was forced to test this group over a much longer period of time than the early-English children, and consequently the dataset for the simultaneous bilingual group was completed at a very late stage in the PhD project. Consequently, only preliminary results can be presented at this point in time. Data from 74 simultaneous Dutch-English bilingual children (4-5-year-olds: n = 24; 8-9-year-olds: n = 26; 11-12-year-olds: n = 24) were gathered. The same tasks were used, except for the inhibition task, which could not be included due to technical errors with the equipment. Children were tested at home, following a comparable procedure to that for the pupils tested at schools (described in Chapter 2). The data from the simultaneous bilingual children were combined with those of the mainstream and early-English pupils. The descriptive

statistics are presented in Table 6.1. The main findings were as follows: the bilingual children of all three age groups outperformed the mainstream and early-English pupils on the verbal working memory task, and the 11-12 year-old bilingual group also showed significantly better performance on the switching task. Bilingual children's performance on the non-verbal working memory task did not differ from that of the two other groups, but the results showed a trend in the expected direction (p = .071). In addition, as was also the case for mainstream and early-English pupils, lexical balance was positively and significantly related to bilingual's performance on the switching task, but it showed no significant relation to performance on the verbal and non-verbal working memory tasks.

The results that I found in these preliminary analyses are partially the same as previous findings with 6-year-old Turkish Dutch bilingual children (Blom et al., 2014). Just like it was found in the preliminary analyses presented here, Blom et al. also showed an advantage for bilingual children over monolingual children on a backwards digit recall task assessing verbal working memory, and a trend in the same direction on an odd-one-out task assessing non-verbal working memory. Contrary to Blom et al., however, I did not find a relation between language balance and verbal working memory skills. Whereas these contradicting findings may seem surprising at first, this is in line with the largely mixed results from previous studies on bilingual children's working memory skills, suggesting that more research is needed in order to draw firm conclusions (Barac, Bialystok, Castro, & Sanchez, 2014).

Taken together, the findings of Chapter 2 and the findings of the additional analyses suggest that some threshold level of bilingual experience should be reached before differences in development of executive functions become apparent. Receiving English lessons from age four for less than two hours per week is not enough to reach this threshold. At the same time, my findings also show that both in unexperienced and more experienced bilinguals, more balanced lexicons are associated with better switching skills.

Table 6.1
Raw Scores on Executive Functioning Measures, the Intelligence Measure, and Vocabulary Measures for Mainstream (Control), Early-English, and Bilingual Groups

			4-5 year-olds			8-9 year-olds			11-12 year-olds		
			Control	Early-English	Bilingual	Control	Early-English	Bilingual	Control	Early-English	Bilingual
Switching (accuracy, max. 24)		*N*	29	30	16	33	37	24	26	28	23
		M	18.45	18.10	17.50	19.70	19.32	20.54	20.77	21.25	22.83
		(SD)	(1.76)	(1.47)	(3.58)	(3.03)	(3.09)	(2.92)	(2.86)	(2.29)	(1.27)
Inhibition	Congruent (accuracy, max. 120)	*N*	36	35	-	34	37	-	25	27	-
		M	105.39	105.97		108.06	107.51		113.56	113.41	
		(SD)	(11.15)	(8.71)		(5.78)	(6.74)		(4.08)	(4.30)	
	Incongruent (accuracy, max. 120)	*M*	99.56	90.54	-	99.97	99.43	-	105.76	105.00	-
		(SD)	(13.77)	(13.77)		(9.20)	(9.26)		(9.26)	(8.20)	
	Simon Effect (RT)	*M*	64.00	72.03	-	39.81	45.51	-	36.51	39.77	-
		(SD)	(41.04)	(60.93)		(18.71)	(0.94)		(20.03)	(17.35)	
Verbal working memory (accuracy, max. 162)		*N*	37	39	23	32	37	25	26	28	24
		M	4.08	3.46	6.13	12.38	11.73	14.40	16.38	15.07	16.63
		(SD)	(3.44)	(2.61)	(3.84)	(3.70)	(3.44)	(4.37)	(4.22)	(3.64)	(5.19)
Nonverbal working memory (accuracy, max. 36)		*N*	37	39	24	32	37	26	26	28	24
		M	11.62	10.28	9.92	20.00	19. 89	22.96	25.00	23.93	25.58
		(SD)	(3.56)	(2.80)	(3.22)	(4.02)	(3.96)	(6.43)	(4.72)	(4.40)	(6.46)
Intelligence (raw score, max. 62 for 4-5 year-olds, max. 73 for older pupils)		*N*	38	40	22	34	38	23	26	28	22
		M	21.39	21.63	20.77	32.26	30.66	32.09	36.96	36.57	37.32
		(SD)	(5.24)	(5.83)	(6.05)	(4.26)	(4.04)	(4.91)	(5.36)	(4.61)	(7.01)

N.B.: See next page

Table 6.1
Raw Scores on Executive Functioning Measures, the Intelligence Measure, and Vocabulary Measures for Mainstream (Control), Early-English, and Bilingual Groups

		4-5 year-olds			8-9 year-olds			11-12 year-olds		
		Control	Early-English	Bilingual	Control	Early-English	Bilingual	Control	Early-English	Bilingual
Dutch vocabulary (raw score, max. 204)	*N*	38	40	24	34	38	26	26	28	22
	M	72.05	72.58	65.50	110.68	111.13	115.23	133.65	134.86	142.36
	(SD)	(10.28)	(12.21)	(15.91)	(7.53)	(7.56)	(12.61)	(10.14)	(9.80)	(10.70)
English vocabulary (raw score, max. 228)	*N*	38	40	24	34	38	25	26	28	24
	M	13.39	18.18	81.25	55.94	56.13	135.12	93.77	111.50	156.75
	(SD)	(8.33)	(10.47)	(18.97)	(26.67)	(16.77)	(15.48)	(32.00)	(27.14)	(13.13)
Balance	*N*	38	40	24	34	38	25	26	28	22
	M	-1.97	-1.65	0.22	-0.89	-0.84	0.16	-0.52	-0.33	0.10
	(SD)	(0.61)	(0.56)	(0.37)	(0.44)	(0.32)	(0.10)	(0.29)	(0.22)	(0.10)

N.B.: See previous page

6.1.2 Phonological awareness

The research question addressed in Chapter 3 was whether bilingualism helps or hinders the development of phonological awareness, and whether the degree of bilingualism modulates this effect. Phonological awareness tasks were administered to pupils who were in first, second, or third grade of primary school. I showed that the bilingual groups did not differ in their performance on a rhyming task, nor in performance on a phoneme blending task. Effects of bilingualism were found on an onset phoneme determination task, and a phoneme deletion task. Contrary to what was expected on the basis of previous findings (Bruck & Genesee, 1995; Reder, Marec-Breton, Gombert, & Demont, 2013), the effect of bilingualism appeared on tasks that were administered to pupils who were already (starting to become) literate in Dutch. Despite this general effect of bilingualism, there appeared to be no robust pairwise differences between the three groups. Moreover, the effect of bilingualism disappeared when controlling for pupils' level of Dutch vocabulary or their working memory abilities. The effect of bilingualism also disappeared when comparing only mainstream and early-English pupils' scores to each other, thereby controlling for the fact that pupils were nested within schools. I therefore hypothesized that school practices rather than bilingualism may account for differences in the development of phonological awareness.

Previous studies on the relation between bilingualism and phonological awareness yielded mixed results (see Appendix A). Some studies showed positive relations between bilingualism and phonological awareness (Bialystok, Luk, & Kwan, 2005; Chen, Xu, Nguyen, Hong, & Wang, 2010; Kang, 2012; Kuo & Anderson, 2010; Marinova-Todd, Zhao, & Bernhardt, 2010). Others showed that bilingual children perform better than monolingual children on some phonological awareness tasks, while it is the other way around on other tasks (Bialystok, Majumder, & Martin, 2003; Bruck & Genesee, 1995; Chen, Wu, & Shu, 2004; Kuo, Uchikoshi, Kim, & Yang, 2016; Loizou & Stuart, 2003; Rubin & Turner, 1989). More recently, other studies have even found only negative effects of bilingualism on phonological awareness (Janssen, Segers, McQueen, & Verhoeven, 2015, 2017; Lesniak, Myers, & Dodd, 2014). The findings reported in this thesis are in line with those studies that show that bilingualism does not influence the development of phonological awareness (Bialystok et al., 2003; Loizou & Stuart, 2003; Reder et al., 2013). Since I did not measure pupils' exposure to literacy activities, nor their level of literacy, it remains unknown whether such differences may have influenced my findings.

Rather than bilingualism, other factors were at play in performance on the phonological awareness tasks. Both pupils' grade (i.e., the grouping at the level of school cohort), and their chronological age (i.e., their actual age at the level of months) were important predictors of all four phonological awareness skills. Chronological age played an especially important role for the pupils in the lower grades. I attributed this to the fact that all pupils enter primary school when they turn four years old, which may be earlier in the school year for some pupils than for others. Consequently, the pupils that have entered primary school at an earlier point may have benefited more from the educational instruction and/or are more cognitively matured by the time of testing. Chronological age was also a more important predictor for bilinguals' scores than for the two other types of pupils'

scores on the onset phoneme determination and phoneme deletion task. I attributed this to the fact that older bilingual children have had more experience with the Dutch language. Since the mainstream and early-English pupils were being raised in Dutch, they would have reached this point of experience earlier in time. Dutch vocabulary, working memory, and short-term memory were also related to pupils' performance on phonological awareness tasks. For phoneme deletion, working memory scores even appeared to be the sole statistically significant predictor. Differences in English vocabulary were not related to phonological awareness skills, except for rhyming for which it was positively related to performance.

The findings of this chapter show that, rather than bilingualism, individual differences in age, language development, and working memory are related to children's performance on phonological awareness tasks. Future research should therefore take these individual differences into account when assessing phonological awareness skills. This thesis shows that learning two languages at the same time, either at school with minimal exposure to one language or in daily life with considerably more exposure to both languages, is not beneficial for children's phonological awareness, but at the same time, it also does not hinder it either. Pupils who are in the first three grades of primary school, and who are attending Dutch mainstream schools, early-English schools, or are being raised bilingually, all show the same rate of development of phonological awareness skills.

6.1.3 Speech perception

Chapter 4 focused on the question whether early-English pupils are better than mainstream pupils at distinguishing English phonetic contrasts, and whether they are as good at it as simultaneous English bilingual children. This question was investigated with an XAB task, in which children had to say whether stimulus A or B matched stimulus X. I tested children on four phonetic contrasts, varying in their level of difficulty for native speakers of Dutch: /b/-/s/ (easy control contrast), /k/-/g/ (intermediate difficulty), /f/-/θ/ (hard), /ɛ/-/æ/ (very hard). The major finding was that early-English pupils were not better than mainstream pupils at distinguishing any of the speech contrasts, while bilinguals outperformed both mainstream and early-English pupils on all contrasts except the easy control contrast.

I attributed the finding that early-English pupils did not outperform mainstream pupils on non-native contrasts to the fact that explicit instruction of English phonemes seems to receive little attention in the educational curriculum. In addition, teachers may be providing children with Dutch-accented English input in which the speech sounds are not correctly produced, thereby limiting pupils' opportunities to implicitly learn to distinguish between different English speech sounds.

I chose the four contrasts on the basis of the Perceptual Assimilation Model for L2 learners, which distinguishes different types of contrasts that have different levels of difficulty for non-native speakers (Best & Tyler, 2007). Two phonetic contrasts that PAM considers a Category Goodness type were included: /k/-/g/ and /f/-/θ/. PAM does not differentiate between predictions for /k/-/g/ and /f/-/θ/, while one (/k/-/g/) entails a sound that learners have some L1 familiarity with (/g/) and the other does not. This study shows that L2 learners have less difficulty with a Category Goodness contrast that involves a sound which is a system gap, occurs in

loan words, or occurs as an allophone, than with a Category Goodness contrast for which that is not the case (here: /f/-/θ/). Future research is needed to confirm whether this holds for other sound contrasts, and for other languages.

I also found that older pupils were better at distinguishing speech contrasts than younger pupils, in all three groups (mainstream, early-English, bilingual). I thereby confirmed the findings of previous research that showed that children between 4 and 12 years of age are increasingly consistent in phonemic categorization (Hazan & Barrett, 2000). Importantly, this did not hold for the /ɛ/-/æ/ contrast. Older bilingual children performed better on this contrast than younger bilingual children, but this was not true for the mainstream and early-English pupils. These latter two groups always performed at chance level, at all ages. According to the PAM, English /ɛ/ and /æ/ should both be assimilated to the closest Dutch phoneme /ɛ/, but as neither of the English phonemes is a good exemplar of the Dutch phoneme, learners will likely have a hard time distinguishing between the two (Best & Tyler, 2007).

In summary, the results of Chapter 3 show that early-English pupils are not better able to distinguish between English speech sounds than mainstream pupils, whereas bilingual children outperformed both other groups. This shows that the plasticity of the speech perception system is limited, as minimal input in the L2 is not enough to perform at the same level as children growing up with two languages.

6.1.4 Using the PPVT to measure L2 vocabulary development

The fifth chapter revolved around the question of whether the PPVT-4 is a reliable test to assess English vocabulary in learners of English as an L2, and whether Dutch word frequency and cognate status of items influence pupils' scores on this test. The PPVT-4 was administered to primary- and secondary-school pupils aged from four to fifteen years old. Three major findings came out of this study. First, the PPVT-4 appeared not to be a reliable test for unexperienced L2 learners, namely those who were only able to complete the first three sets (36 items) of the test. It became a more reliable test when administered to more experienced pupils, which in this study were pupils in the highest grade of primary school, and those in secondary school. Second, L1 (Dutch) frequency of the translation equivalents of the items in the test was a predictor of young pupils' scores, whereas for older pupils, L2 (English) frequency was a significant predictor. Third, for pupils of all ages, cognate status was an important predictor of their scores, meaning that pupils were more likely to respond correctly to those items in the test of which the Dutch translation equivalent was phonetically similar to the English word.

Despite the fact that the lexical frequency of the English items and their Dutch translation equivalents were highly correlated, unexperienced pupils' scores depended on the Dutch and not the English frequency. This was attributed to the fact that pupils' exposure to English was very limited, and hence there was not much room for English frequency to play a role. For more experienced pupils, who were expected to get more exposure to English inside as well as outside the school environment (Lindgren & Muñoz, 2013), English frequency mattered more than Dutch frequency.

English and Dutch are both Germanic languages and they share many cognates. Previous research (Bosma, Blom, Hoekstra, & Versloot, 2016; Pérez,

Peña, & Bedore, 2010; Potapova, Blumenfeld, & Pruitt-Lord, 2016) already showed that children can make use of such similarities between words. My study showed that although even four-year-olds are able to make use of similarities between English and Dutch for L2 word comprehension, older children do this to a higher degree than younger children.

In addition to assessing the suitability of the PPVT-4 for young English learners, I also used the PPVT-4 to investigate whether English programmes at school are effective. I found that pupils enrolled in an English-language curriculum outperformed pupils in a mainstream curriculum, but only at the end of primary school, and in secondary school. In other words, in the earlier years, early-English and mainstream pupils obtained comparable scores. Three possible reasons were given to account for this finding. First, it could be that the two groups do actually differ in vocabulary development, but that these differences do not show when the PPVT-4 is administered as outlined in the manual. Inexperienced L2 learners usually only respond to relatively few items, resulting in a floor effect. It is probable that the test does not differentiate between inexperienced pupils' scores. Given that my study showed that the PPVT-4 is not a reliable measure for inexperienced L2 learners, this is a plausible reason. I therefore suggest starting with an earlier set than the age-appropriate set when administering the PPVT-4. Second, although not in line with the results of previous studies (Unsworth, Persson, Prins, & de Bot, 2015; van der Leij, Bekebrede, & Kotterink, 2010), it could be that the two groups simply do not differ in English vocabulary. One reason for the absence of group differences could be that the early-English pupils were exposed to English for less than 60 minutes per week, or were educated by a teacher with a B level only, both factors which have been associated with slower English vocabulary development when compared with pupils who get more input or input by a more proficient teacher (Unsworth et al., 2015).

Two other important recommendations for research practice come from the findings of Chapter 5. First, this study showed that the test is not reliable when administering it to an unexperienced group of L2 learners. Therefore, the test should not be used in very inexperienced groups. Second, the PPVT-4 contains many items that are cognates to Dutch learners of English, but it may contain less cognates to pupils with other (non-Germanic) mother tongues. Researchers should thus be careful in administering the test in cross-sectional research which compares learners of English who have different L1s.

6.1.5 Dutch and English vocabulary development

Although it was not one of the main questions of the thesis, pupils' vocabulary development in English and Dutch was investigated in every study. This was mainly done to investigate whether early-English pupils differed from their mainstream peers in the development of vocabulary in both languages. The expectation was that early-English pupils would have Dutch vocabulary levels that are comparable to that of their peers, but would have larger vocabularies in English. This expectation was based on previous findings. Research with 4- and 5-year-olds (Goorhuis-Brouwer & de Bot, 2010) and with 8-year-olds (van der Leij et al., 2010) showed that early-English education had no negative influence on the development of Dutch, and that pupils' Dutch vocabulary was comparable to that of mainstream

pupils of the same age. Prior research also showed that early-English programmes seem effective, since early-English pupils had better knowledge of English vocabulary than pupils who were not enrolled in such a programme (Lobo, 2013; Unsworth et al., 2015; van der Leij et al., 2010). Those previous studies focused mostly on children who had just started primary school, or who were still in the lower grades of primary school. Only one previous study (de Graaff, 2015) investigated final-grade pupils' knowledge of English, but that study did not investigate pupils' Dutch or English vocabulary knowledge.

Across the various studies included in this thesis, I investigated Dutch and English vocabulary development of pupils in grade 1 (Chapter 2, 3, and 4), grades 2 and 3 (Chapter 3), and grades 5 and grade 8 (Chapter 2 and 4). The (combined) results of these investigations show a clear picture for pupils' Dutch vocabulary level. In every study and for pupils in all grades, I found that early-English pupils' Dutch vocabulary level was comparable to that of their peers attending mainstream schools. This finding confirms and extends the findings of previous studies, showing that also after several years of English education, when pupils are in the final grade of primary school, English lessons do not adversely affect their Dutch vocabulary development, which is a concern that many parents and teachers still have (Goorhuis-Brouwer & de Bot, 2010).

The picture for pupils' English vocabulary development is more mixed. Positive effects of early-English education on pupils' English vocabulary level were found for first-grade pupils (but only in the study reported in Chapter 3), for second- and third-grade pupils, and for eighth-grade pupils (but only in the study reported in Chapter 2). No positive effects for fifth-grade pupils were found. One possible reason for the mixed effects may be because the quality of the English lessons differed. Some of the participating schools had a certificate for their English lessons, others did not. Such a certificate is issued by one of three independent organisations upon meeting the requirements with regards to the quality standards for early-English education as described by Nuffic (Nuffic, n.d.). Having a certificate was however not predictive of pupils' English vocabulary scores: For pupils in grade 1, the schools in Chapter 2 had no certificate, the ones in Chapters 3 and 4 did have a certificate; only in Chapter 3 higher English vocabulary scores were found for early-English than for mainstream pupils. For the pupils in grade 8, schools in Chapter 2 had no certificate, whereas schools in Chapter 4 did; only pupils from the former (counterintuitively) showed *higher* vocabulary scores than pupils from the mainstream schools. Having a certificate may be a proxy for providing English lessons that are of sufficient quality, but the absence of such a certificate is not an indicator of insufficient quality.

Other possible reasons for (the absence of) differences between mainstream and early-English pupils' English vocabulary scores may be found in differences in teaching practices, or in the amount of exposure to English that pupils receive outside the school. To start with the first, schools differ in the focus of the English lessons. It has been shown that the focus is mostly on English listening skills, but that many early-English schools also focus on vocabulary development (Thijs, Trimbos, Tuin, Bodde, & de Graaff, 2011). Depending on the focus of the English lessons, pupils may become better at some aspects of English than at others, which

may be reflected here in the fact that early-English pupils did not perform better than mainstream pupils on the English vocabulary task.

Yet another possible difference between teachers may be their language use. Although schools reported that they taught English for at least 60 minutes per week, that does not necessarily mean that all teachers always use the English language for that amount of time. In personal communication with the teachers, some mentioned that they sometimes switched to Dutch during the English lessons. The actual amount of time devoted to English may thus have been less than reported, and the early-English group may thus not have been as homogeneous as intended.

One more reason for mixed findings with respect to group differences between mainstream and early-English vocabulary scores, is that pupils may differ in their amount of out-of-school exposure, a factor that has been shown to be of influence on especially older pupils' level of English proficiency (Jensen, 2017; Kuppens, 2010; Sylvén & Sundqvist, 2012; Thijs et al., 2011). It may be that mainstream pupils had more exposure to English. This may in turn have compensated for the lack of exposure at school, which can be a possible explanation for the absence of significant differences between early-English and mainstream pupils' vocabulary scores. I asked parents to fill in a questionnaire with information about pupils' out-of-school exposure to English, but the response rate was very low. For the majority of pupils, it remains unknown to what extent they were exposed to English outside the school environment. On the basis of the data that are available, it seems however not plausible that mainstream pupils had more exposure to English outside school than early-English pupils. In total, 250 questionnaires were returned. Of these questionnaires, 140 came from early-English pupils. Preliminary analyses showed that early-English and mainstream pupils did not significantly differ in the amount of out-of-school exposure to English. For pupils from both types of schools, out-of-school exposure was related to pupils' English vocabulary knowledge, but only for pupils in the highest grade.

The results of my research confirm and extend those of de Graaff (2015). In that study, which did not investigate pupils' English vocabulary, it was also found that differences between early-English and mainstream pupils' level of English were small, and that there was a large amount of overlap in their scores. It is likely that differences between schools, teachers, classroom practices, and pupils have contributed to these findings in my dissertation. Future research should investigate which classroom practices are effective, how much English teachers actually use during the English lessons, which out-of-school experiences pupils have with respect to English, and how these variables relate to pupils' English vocabulary knowledge. From the results of my research, it can be concluded that English lessons from the start of primary school, as they are currently taught in the Netherlands, do not negatively influence pupils' first language, Dutch, but at the same time also do not necessarily positively influence pupils' English vocabulary.

6.2 Limitations

6.2.1 Limitations with respect to the research design

There are a number of limitations to the research carried out in this thesis. First of all, whilst the overall number of participants included in each study is

relatively high, at the group level, it sometimes remains fairly small, especially for the simultaneous bilingual children (i.e., < 20). The aim was to include at least 40 participants per age group in each of the studies, but it was not always possible to reach that number. Requirements were that schools had at least eight years of experience with early-English education (in the studies involving children in grade 8, reported in Chapters 2, 4, and 5), and that they had a certificate for their English lessons (in the studies reported in Chapters 3 and 4). These requirements limited the number of schools eligible for participation. Furthermore, I was dependent on schools' and parents' willingness to participate in the studies. Pupils had to be tested individually, and in the case of bilingual children, after school time. This made testing time consuming and this in turn limited the number of participants that could be included. It may be the case that there were consistent but small effects between the mainstream, early-English, and bilingual pupils, but I was not able to capture these with the present sample sizes. Future research with (even) larger sample sizes is needed.

A second limitation of this thesis is that all the studies were cross-sectional, comparing children from different age groups. This was done in an attempt to investigate how the relation between early-English education and pupils' development may be different in older than in younger pupils, as older pupils had had more experience with the L2 than younger pupils. Future research could include a longitudinal design, to investigate how the relation between early-English education and pupils' development within a group of pupils changes as those pupils get older, as well as what the nature of the relation between bilingual experience and pupils' development in different domains is. In my study on pupils' executive functions, for example, I predicted from a theoretical point of view that pupils' lexical balance affected their switching abilities and that this was the direction of this relation. It is however possible that this relation may hold in the opposite direction: pupils who can easily switch between rules or languages may more easily pick up the new language, and therefore show more balanced lexicons. Research has shown that foreign language learning is a complex process, and that different factors may be at play at different points in time (Lindgren & Muñoz, 2013). Future research with a longitudinal design could contribute to clarifying how different experiences may change the relation between early-English education and pupils' development over the years.

6.2.2. Limitations with respect to other variables

The work described in this thesis shows that early-English education, sometimes indirectly, has an influence on some aspects of pupils' development, but not on other aspects. I showed, for example, that early-English pupils' English and Dutch lexicons are more balanced than that of mainstream pupils, which in turn is related to their switching skills (Chapter 2). Language balance was not related to other executive functioning processes, such as working memory. I also showed that the effects of early-English education may differ per age group. With respect to pupils' vocabulary development, for example, in Chapter 2 I found a significant advantage for early-English pupils in final grade over their mainstream peers. No such differences were found for pupils at the beginning of primary school or halfway through primary school (Chapter 2). In research by others (e.g. Unsworth et

al., 2015), as well as in my own research (Chapter 3), significant differences in English vocabulary were found already in first-grade pupils. One reason for these inconsistent findings may be that the work in this thesis focused mainly on child and language characteristics, whereas environmental influences also may have played a role.

I did not investigate certain characteristics of learners and teachers, such as pupils' motivation to learn a foreign language, or pupils' and teachers' attitudes towards the foreign language. A recent report has shown that in general, Dutch pupils' attitudes towards the English language and towards learning English are positive (Jenniskens et al., 2018). Nevertheless, pupils as well as teachers may differ in their attitudes or motivation with respect to English lessons. Research has shown that young Spanish L2 learners differ in their motivation to learn the L2, and that motivation for L2 learning and L2 attainment positively correlate with each other (Muñoz, 2017). This may as well be the case for Dutch pupils. Teachers may also differ in their attitudes towards English lessons in primary school. Previous research carried out in secondary schools showed that Dutch teachers think that English programmes are effective, and that teaching in such a programme is beneficial to their own professional development, as well as their pedagogical competences (Oattes, Oostdam, de Graaff, & Wilschut, 2018). Research with primary-school teachers in immersion programmes in Spain showed similar results (Pena Diaz & Porto Requejo, 2008). At the same time, research including teachers from various countries in Europe, but not the Netherlands (Enever, 2014; Pérez-Cañado, 2012), showed that teachers also report that their own proficiency level is not high enough to use English as the language of instruction. This may be one of the reasons for them using the native language instead, and for having difficulty with handling spontaneous interactions in the L2. Teachers also report that they are insecure about their knowledge of methodologies for teaching in a foreign language (Pena Diaz & Porto Requejo, 2008). Primary-school teachers in early-English schools in the Netherlands may experience similar insecurities and challenges, which in turn may influence their teaching practices.

Preliminary findings of a questionnaire that I distributed amongst mainstream and early-English schools in the Netherlands and that was filled in by 102 teachers (n = 50 from early-English schools) suggest that teachers' English proficiency level and attitudes towards the possible negative sides of early-English education, such as the time costs, correlate negatively. In other words, teachers who are more proficient in English are more likely to be positive about early-English education. Furthermore, I also found a positive relation between teachers' English proficiency level and the belief in their own early-English teaching competence. Given that teachers have an important influence on pupils' learning processes (Hattie, 2003), and the success of early foreign-language classes is dependent on the motivation and the dedication of the teachers (Fernández & Halbach, 2011), it may be that such variables have influenced the effectiveness of the early-English programmes in the classes that participated in this study, which in turn may have influenced pupils' development in the domains that were the focus of this study. Although I made an attempt to take into account environmental and individual differences that may have affected L2 learning, it was beyond the scope of this thesis to include many of such variables. Future research could investigate how

these environmental and individual differences may contribute to the relation between early-English education and pupils' development.

Another limitation of this research is that I was not able to match the group of bilingual children to the mainstream and early-English groups with respect to socio-economic status (SES). Mainstream schools were matched to early-English schools in terms of average income in their neighbourhood. In this way, pupils from the two types were likely to come from similar SES backgrounds. In addition, in a questionnaire, parents were asked about their level of education. Since the majority of the parents did not return the questionnaire, I was not able to investigate to what extent mainstream and early-English pupils' SES background differed from that of the bilingual group. From the data that are available for mainstream and early-English pupils, it seems that the parental level of education varied much more than for the bilingual children. With very few exceptions, bilingual children's parents had a high level of education. Since SES is related to cognitive and academic achievement (Bradley & Corwyn, 2002), possible SES differences between the three groups may have directly or indirectly influenced my results. This is especially relevant to my findings on pupils' executive functions, since both SES and bilingualism are related to children's executive functions (Calvo & Bialystok, 2014). The executive functions advantage that I found in favour of the bilingual group might thus be due not only to the fact that these children were being raised bilingually, but also to their (possibly) favourable SES background. Parental education levels, however, are also related to children's level of intelligence (Bradley & Corwyn, 2002). If the mainstream, early-English, and bilingual children came from highly different SES backgrounds, that would likely have been reflected in the pupils' scores on an intelligence test. That appeared not to be the case. Although this does not rule out the possible influence of SES background, it shows that the three groups were comparable on a measure that is closely related to SES.

A final limitation is the rather large gap in bilingual experiences between the early-English and bilingual group: whereas early-English pupils had very limited experience with the use of two languages, the bilingual group used both languages on a daily basis and they were highly proficient in both languages. The results showed that a limited amount of experience had minimal influence on pupils' development. To investigate whether a larger amount of instructed L2 learning has greater effects on pupils' development, future research should include a group of pupils enrolled in bilingual primary education programmes, which provide more English exposure at school (e.g. 30-50% of the time in Dutch bilingual schools). Previous research has suggested that such programmes may influence pupils' development of executive functions (Purić et al., 2017), provided that the pupils have had enough exposure to the L2 (Carlson & Meltzoff, 2008; Purić et al., 2017). It was not possible to include such a group in the research reported here, since bilingual primary education is only in its pilot phase in the Netherlands. This pilot started just a few years ago, such that only first-grade pupils could have been included in the study, had I been able to include such a group. The first results of this pilot are however positive, showing that pupils from bilingual primary schools have better knowledge of English vocabulary and grammar than early-English pupils (Jenniskens et al., 2018).

6.3 Relevance of the findings

The results reported in this thesis are relevant to the field of L2 and bilingualism research as well as to educational practice. The only language pair studied in this thesis was Dutch-English, because English is the most frequently taught foreign language in Dutch schools. The results are nevertheless also relevant to knowledge about L2 learners and bilinguals who speak a different language combination. The effect of bilingualism on the development of executive functions is said to be language independent (Barac et al., 2014), and the results of this thesis may thus be generalized to speakers of other languages. For phonological awareness, there is an ongoing discussion whether a possible bilingual advantage is language specific or not (Bruck & Genesee, 1995; Marinova-Todd et al., 2010). For the domains of speech perception and the use of the PPVT-4 as an L2 vocabulary test, it is clear that the results of this thesis are specific to Dutch learners of English. With respect to speech perception, for example, the speech contrasts investigated in this study may be difficult to Dutch learners of English, but that does not necessarily hold for L2 learners with another mother tongue. The same is true for the results of the study on the PPVT-4, as the PPVT-4 may contain a different number of cognates to children who have a first language other than Dutch.

Despite the narrow focus on Dutch and English only, the outcomes in general are important to speakers of other language pairs. The outcomes of Chapter 2 suggest that limited exposure to an L2, regardless of which language that is, is not enough to enhance the development of executive functions. Chapter 3 implies that individual differences rather than bilingualism are related to the development of phonological awareness. It seems that being exposed to an L2 has no positive, and, importantly, no negative effect on phonological awareness. The results of Chapter 4 suggest that little exposure to an L2 is not enough for children to learn to better distinguish between two members of a phonemic contrasts in the L2 than children who do not get that exposure. It is expected that such findings will also hold for other language pairs. Chapter 5 concludes that participants' L1 will influence their scores on the PPVT-4, which may influence the results when using the PPVT-4 in cross-language comparisons – presumably in all such comparisons.

6.3.1 Relevance of the findings for early-English education

The findings of the current thesis are not only relevant to theories about (bilingual) children's development, but also to early-English education. Early-English programmes, whilst not new, are booming in the Netherlands and beyond. International research on early foreign language learning programmes such as the ones adopted in the Netherlands is scarce, and the studies that have focused specifically on the Dutch situation almost exclusively concentrated on pupils' English language development. That may be explained by the fact that policy makers, but also teachers and parents, have high expectations of early-English education. The Ministry of Education, for example, actively encouraged primary schools to lower the starting age for English lessons, in order to make pupils more proficient in English (Ministerie van Onderwijs Cultuur en Wetenschap, 2005). The Educational Council supported the claim of the Ministry (Onderwijsraad, 2008), and Nuffic, the organisation for internationalisation in education, states on its webpage that learning English at a young age is not only beneficial to pupils' English

language development, but also to pupils' language development in general (Nuffic, 2018).

The results of this thesis show that the claims made by policy makers are likely too strong, as there are often no differences in development between mainstream and early-English pupils. Where differences were found, they were small, and often unstable (in the case of phonological awareness) or inconsistent (in the case of English vocabulary development). In other words, offering English early on in children's school career can by no means be seen as a guarantee for subsequent successful English language development. It must however be acknowledged that this thesis focused on a limited part of pupils' competencies, and the fact that no differences have been found in these competencies does not mean that there may be no differences in other skills where early-English education has been claimed to have an influence, such as pupils' level of international awareness, for example (Nuffic, 2018).

The schools that participated in my studies were heterogeneous with respect to variables such as the neighbourhood they were located in, the population of pupils, and teaching practices. Since such variables may influence pupils' cognitive and academic achievements, this could in part account for the absence of differences between groups of mainstream and early-English pupils. I was not able to capture all differences between schools, and therefore, this thesis does not present the whole picture of the influences of early-English education on pupils' cognitive and linguistic development. At the same time, I took a sample that is representative for the state of early-English education in the Netherlands at the present time. It is a fact that there is no agreed curriculum on early-English education, only a maximum amount of English instruction is defined but no minimum, and whilst there are requirements for teachers' English language proficiency, these requirements are not strictly tested. Consequently, there is a lot of variation in early-English education in the Netherlands.

6.3.2 Relevance for the knowledge on the development of L2 learners with limited L2 experience

The findings of this thesis are relevant to knowledge about the language and cognitive development of children who have limited L2 experience. Previous studies that focused on differences between bilingual and monolingual children mainly compared groups of bilingual and monolingual children. In those studies, bilinguals are often defined as children who are highly proficient in both languages, and monolinguals as children who are not exposed to any other languages than their native one. In the past years, the idea has arisen that bilingualism is not a categorical variable, but that it is dynamic and composed of multiple dimensions (Luk & Bialystok, 2013). This notion has been reflected in more recent research, which included L2 learners (e.g., Poarch & van Hell, 2012). L2 learners are more experienced with respect to the use of two languages, and also have a higher proficiency in the L2 than monolingual children, but they are not as experienced and proficient as simultaneous bilinguals. Some studies went a step further, and included a continuous measure of bilingualism, for example a measure of language balance (Blom et al., 2014; Thomas-Sunesson et al., 2018). Again, in all of these previous studies, the L2 learners had a relatively large amount of exposure to the L2. Few

studies included groups of children who had had very little exposure to the L2 (Carlson & Meltzoff, 2008; Kang, 2012; Purić et al., 2017), but even in those studies, participants were exposed to the L2 for more than one hour per week. In addition, those studies included participants that all belonged to the same age group. Therefore, it remained unclear if little exposure but for a longer time influences pupils' development.

In this thesis, participants belonged to different age groups, and their exposure to the L2 was very limited (i.e., approximately 60 minutes per week at school). The results show that very limited bilingual experience may already alter children's development, although almost certainly not to the same extent as is the case in simultaneous bilinguals. More specifically, I found that bilingual experience is related to children's language balance, which in turn is related to their switching skills, for inexperienced and experienced bilingual children alike. At the same time, bilingual children outperform mainstream and early-English pupils in some developmental domains (switching, working memory, speech perception), whereas early-English and mainstream pupils do not differ in these domains. This suggests that a certain threshold of bilingual experience has to be reached before differences in development show.

6.4 Conclusion

The focus of this thesis was on the development of pupils attending early-English schools. Providing English lessons from the moment that pupils enter primary school does not hinder their cognitive or linguistic development, as compared to pupils who attend mainstream schools and who do not get English lessons before the penultimate grade. Positive effects of early-English education were found, namely on pupils' English vocabulary. These effects were however often small, and did not always hold for pupils of all age groups. There were also several skills that early-English education did not affect, namely executive functions, phonological awareness, and the perception of phonetic contrasts. When it comes to promoting pupils' cognitive and linguistic development, providing English lessons from the start of primary school works no miracles.

6.5 References

Barac, R., Bialystok, E., Castro, D. C., & Sanchez, M. (2014). The cognitive development of young dual language learners: A critical review. *Early Childhood Research Quarterly*, *29*, 699–714. http://doi.org/10.1016/j.ecresq.2014.02.003

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In M. J. Munro & O.-S. Bohn (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 13–34). Amsterdam: John Benjamins. http://doi.org/10.1075/lllt.17.07bes

Bialystok, E., Luk, G., & Kwan, E. (2005). Bilingualism, biliteracy, and learning to read: Interactions among languages and writing systems. *Scientific Studies of Reading*, *9*, 43–61. http://doi.org/10.1207/s1532799xssr0901_4

Bialystok, E., Majumder, S., & Martin, M. M. (2003). Developing phonological awareness: Is there a bilingual advantage? *Applied Psycholinguistics*, *24*, 27–

44. http://doi.org/10.1017.S014271640300002X

Blom, E., Küntay, A. C., Messer, M., Verhagen, J., & Leseman, P. (2014). The benefits of being bilingual: Working memory in bilingual Turkish-Dutch children. *Journal of Experimental Child Psychology*, *128*, 105–119. http://doi.org/10.1016/j.jecp.2014.06.007

Bosma, E., Blom, E., Hoekstra, E., & Versloot, A. (2016). A longitudinal study on the gradual cognate facilitation effect in bilingual children's Frisian receptive vocabulary. *International Journal of Bilingual Education and Bilingualism. Advance Online Publication.* http://doi.org/10.1080/13670050.2016.1254152

Bradley, R. H., & Corwyn, R. F. (2002). Socioeconomic status and child development. *Annual Review of Psychology*, *53*, 371–399. http://doi.org/10.1146/annurev.psych.53.100901.135233

Bruck, M., & Genesee, F. H. (1995). Phonological awareness in young second language learners. *Journal of Child Language*, *22*, 307–324. http://doi.org/10.1017/S0305000900009806

Calvo, A., & Bialystok, E. (2014). Independent effects of bilingualism and socioeconomic status on language ability and executive functioning. *Cognition*, *130*, 278–88. http://doi.org/10.1016/j.cognition.2013.11.015

Carlson, S. M., & Meltzoff, A. N. (2008). Bilingual experience and executive functioning in young children. *Developmental Science*, *11*, 282–298. http://doi.org/10.1111/j.1467-7687.2008.00675.x

Chen, X., Wu, X., & Shu, H. (2004). Phonological awareness of bilingual and monolingual Chinese children. *Journal of Educational Psychology*, *96*, 142–151. http://doi.org/10.1037/0022-0663.96.1.142

Chen, X., Xu, F., Nguyen, T.-K., Hong, G., & Wang, Y. (2010). Effects of cross-language transfer on first-language phonological awareness and literacy skills in Chinese children receiving English instruction. *Journal of Educational Psychology*, *102*, 712–728. http://doi.org/10.1037/a0018802

Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia*, *44*, 2037–2078. http://doi.org/10.1016/j.neuropsychologia.2006.02.006

de Graaff, R. (2015). Vroeg of laat Engels in het basisonderwijs: Wat levert het op? [Early or late English in primary education: Does it matter?]. *Levende Talen Tijdschrift*, *16*(2), 3–15.

Enever, J. (2014). Primary English teacher education in Europe. *ELT Journal*, *68*, 231–242. http://doi.org/10.1093/elt/cct079

Fernández, R., & Halbach, A. (2011). Analysing the situation of teachers in the Madrid Bilingual Project after four years of implementation. In Y. Ruiz de Zarobe, J. Manuel Sierra, & F. Gallardo del puerto (Eds.), *Content and foreign language integrated learning: Contributions to multilingualism in European contexts* (pp. 241–270). Bern: Peter Lang AG, International.

Gathercole, V. C. M., Thomas, E. M., Kennedy, I., Prys, C., Young, N., Guasch, N. V., … Jones, L. (2014). Does language dominance affect cognitive performance in bilinguals? Lifespan evidence from preschoolers through older adults on card sorting, Simon, and metalinguistic tasks. *Frontiers in*

Psychology, *5*, 11. http://doi.org/10.3389/fpsyg.2014.00011

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*, *14*, 289–302. http://doi.org/10.1177/1367006910367846

Gussenhoven, C. (1999). Illustrations of the IPA: Dutch. In *Handbook of the International Phonetic Association* (pp. 74–77). Cambridge: Cambridge University Press.

Hamann, S. R., & De Jonge, A. (2015). Eliciting the Dutch loan phoneme /g/ with the Menu Task. In *Proceedings of the 18th International Congress of Phonetic Sciences Glasgow*. Glasgow: The University of Glasgow.

Hattie, J. (2003). Teachers make a difference: What is the research evidence? Paper presented at the Australian Council for Educational Research Annual Conference on Building Teacher quality. Melbourne, Australia.

Hazan, V., & Barrett, S. (2000). The development of phonemic categorization in children aged 6–12. *Journal of Phonetics*, *28*, 377–396. http://doi.org/10.1006/JPHO.2000.0121

Janssen, C., Segers, E., McQueen, J. M., & Verhoeven, L. (2015). Lexical specificity training effects in second language learners. *Language Learning*, *65*, 358–389. http://doi.org/10.1111/lang.12102

Janssen, C., Segers, E., McQueen, J. M., & Verhoeven, L. (2017). Transfer from implicit to explicit phonological abilities in first and second language learners. *Bilingualism: Language and Cognition*, *20*, 795–812. http://doi.org/10.1017/S1366728916000523

Jenniskens, T., Leest, B., Wolbers, M., Krikhaar, E., Teunissen, C., de Graaff, R., … Koppens, K. (2018). *Evaluatie pilot tweetalig primair onderwijs. Vervolgmeting schooljaar 2016/17 [Evaluation pilot bilingual primary education. Follow-up measurement schoolyear 2016/17]*.

Jensen, S. (2017). Gaming as an English language learning resource among young children in Denmark. *Calico Journal*, *34*(1), 1–19. http://doi.org/10.1558/cj.29519

Kang, J. Y. (2012). Do bilingual children possess better phonological awareness? Investigation of Korean monolingual and Korean-English bilingual children. *Reading and Writing*, *25*, 411–431. http://doi.org/10.1007/s11145-010-9277-4

Kuo, L.-J., & Anderson, R. C. (2010). Beyond cross-language transfer: Reconceptualizing the impact of early bilingualism on phonological awareness. *Scientific Studies of Reading*, *14*, 365–385. http://doi.org/10.1080/10888431003623470

Kuo, L.-J., Uchikoshi, Y., Kim, T.-J., & Yang, X. (2016). Bilingualism and phonological awareness: Re-examining theories of cross-language transfer and structural sensitivity. *Contemporary Educational Psychology*, *46*, 1–9. http://doi.org/10.1016/J.CEDPSYCH.2016.03.002

Kuppens, A. H. (2010). Incidental foreign language acquisition from media exposure. *Learning, Media and Technology*, *35*, 65–85. http://doi.org/10.1080/17439880903561876

Lesniak, A., Myers, L., & Dodd, B. (2014). The English phonological awareness skills of 5;0–6;0-year-old Polish–English, Portuguese–English bilingual

speakers and English monolingual children. *Speech, Language and Hearing*, *17*, 37–48. http://doi.org/10.1179/2050572813Y.0000000029

Lindgren, E., & Muñoz, C. (2013). The influence of exposure, parents, and linguistic distance on young European learners' foreign language comprehension. *International Journal of Multilingualism*, *10*, 105–129. http://doi.org/10.1080/14790718.2012.679275

Lobo, V. R. (2013). *Teaching L2 English at a very early age : A study of Dutch schools*. Radboud University. Retrieved from http://www.lotschool.nl

Loizou, M., & Stuart, M. (2003). Phonological awareness in monolingual and bilingual English and Greek five-year-olds. *Journal of Research in Reading*, *26*, 3–18. http://doi.org/10.1111/1467-9817.261002

Luk, G., & Bialystok, E. (2013). Bilingualism is not a categorical variable: Interaction between language proficiency and usage. *Journal of Cognitive Psychology*, *25*, 605–621. http://doi.org/10.1080/20445911.2013.795574

Marinova-Todd, S. H., Zhao, J., & Bernhardt, M. (2010). Phonological awareness skills in the two languages of Mandarin–English bilingual children. *Clinical Linguistics & Phonetics*, *24*, 387–400. http://doi.org/10.3109/02699200903532508

Ministerie van Onderwijs Cultuur en Wetenschap. (2005). *Bijlage: Herziene kerndoelen basisonderwijs [Attachment: revised objectives primary education]*. The Hague.

Muñoz, C. (2017). Tracing trajectories of young learners: Ten years of school English learning. *Annual Review of Applied Linguistics*, *37*, 168–184. http://doi.org/10.1017/S0267190517000095

Nuffic. (n.d.). *Standaard vroeg vreemdetalenonderwijs Engels [Standard early foreign language education English]*.

Nuffic. (2018). Wat is vroeg vreemdetalenonderwijs? [What is early foreign language education?]. Retrieved July 18, 2018, from https://www.nuffic.nl/primair-onderwijs/vroeg-vreemdetalenonderwijs-vvto/wat-is-vroeg-vreemdetalenonderwijs

Oattes, H., Oostdam, R., de Graaff, R., & Wilschut, A. (2018). The challenge of balancing content and language: Perceptions of Dutch bilingual education history teachers. *Teaching and Teacher Education*, *70*, 165–174. http://doi.org/10.1016/J.TATE.2017.11.022

Onderwijsraad. (2008). *Vreemde talen in het onderwijs [foreign languages in education]*. The Hague: OBT bv.

Pena Diaz, C., & Porto Requejo, M. D. (2008). Teacher beliefs in a CLIL education project. *Porta Linguarum*, *10*, 151–161.

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

Pérez, A. M., Peña, E. D., & Bedore, L. M. (2010). Cognates facilitate word recognition in young Spanish-English bilinguals' test performance. *Early Childhood Services*, *4*(1), 55–67. http://doi.org/10.1044/1092-4388(2013/12-0056)

Poarch, G. J., & van Hell, J. G. (2012). Executive functions and inhibitory control in multilingual children: Evidence from second-language learners, bilinguals,

and trilinguals. *Journal of Experimental Child Psychology*, *113*, 535–551. http://doi.org/10.1016/j.jecp.2012.06.013

Potapova, I., Blumenfeld, H. K., & Pruitt-Lord, S. (2016). Cognate identification methods: Impacts on the cognate advantage in adult and child Spanish-English bilinguals. *International Journal of Bilingualism*, *20*, 714–731. http://doi.org/10.1177/1367006915586586

Purić, D., Vuksanović, J., & Chondrogianni, V. (2017). Cognitive advantages of immersion education after one year: Effects of amount of exposure. *Journal of Experimental Child Psychology*, *159*, 296–309. http://doi.org/10.1016/j.jecp.2017.02.011

Reder, F., Marec-Breton, N., Gombert, J.-E., & Demont, E. (2013). Second-language learners' advantage in metalinguistic awareness: A question of languages' characteristics. *British Journal of Educational Psychology*, *83*, 686–702. http://doi.org/10.1111/bjep.12003

Rubin, H., & Turner, A. (1989). Linguistic awareness skills in grade one children in a French immersion setting. *Reading and Writing*, *1*, 73–86. http://doi.org/10.1007/BF00178839

Sylvén, L. K., & Sundqvist, P. (2012). Gaming as extramural English L2 learning and L2 proficiency among young learners. *European Association for Computer Assisted Language Learning*, *24*, 302–321. http://doi.org/10.1017/S095834401200016X

Thijs, A., Trimbos, B., Tuin, D., Bodde, M., & de Graaff, R. (2011). *Engels in het basisonderwijs [English in primary education]*. Enschede.

Thomas-Sunesson, D., Hakuta, K., & Bialystok, E. (2018). Degree of bilingualism modifies executive control in Hispanic children in the USA. *International Journal of Bilingual Education and Bilingualism*, *21*, 197–206. http://doi.org/10.1080/13670050.2016.1148114

Unsworth, S., Persson, L., Prins, T., & de Bot, K. (2015). An investigation of factors affecting early foreign language learning in the Netherlands. *Applied Linguistics*, *36*, 527–548. http://doi.org/10.1093/applin/amt052

van Bezooijen, R., & Gerritsen, M. (1994). De uitspraak van uitheemse woorden in het Standaard-Nederlands: Een verkennende studie [The pronunciation of foreign words in Standard Dutch: An explorative study]. *De Nieuwe Taalgids*, *87*, 145–161.

van der Leij, A., Bekebrede, J., & Kotterink, M. (2010). Acquiring reading and vocabulary in Dutch and English: The effect of concurrent instruction. *Reading and Writing*, *23*, 415–434. http://doi.org/10.1007/s11145-009-9207-5

Vega, C., & Fernandez, M. (2011). Errors on the WCST correlate with language proficiency scores in Spanish-English bilingual children. *Archives of Clinical Neuropsychology*, *26*, 158–164. http://doi.org/10.1093/arclin/acq097

Appendix A

Table

Overview of Studies Comparing Phonological Awareness in Monolingual and Bilingual Children

Study	**Groups compared**	**Age**	**Tasks used**	**Language of testing**	**Significant differences**	**Results***
Rubin & Turner, 1989	English children in French immersion (n = 32), and English monolinguals (n = 16)	6;6 (immersion) 6;7 (monolinguals)	Phoneme deletion Syllable deletion Word reading (irregular, regular, non-word) Spelling (real and non-word)	English	Immersion > monolingual: phoneme and syllable deletion Monolingual > immersion: reading irregular words	Mixed
Bruck & Genesee, 1995	English children in French schools (n = 91 at T1; n = 77 at T2), and English monolinguals (n = 72 at T1, n = 60 at T2)	5;9 at T1, retested one year later in grade 1 (T2) Longitudinal design	Syllable counting Phoneme counting (T2 only) Onset [phoneme] deletion *Same-different tasks*: Rhyme Syllable beginning Syllable-end Cluster onsets Singleton onsets Rime First phoneme Last phoneme	English	T1: Bilingual > monolingual: Onset deletion, cluster onset, singleton onset, rime T2: Bilingual > monolingual: Syllable counting Monolingual > bilingual: Phoneme counting	Mixed

Bialystok et al., 2003 (study I)	French-English bilinguals (n = 36), English monolinguals (n = 36)	bilinguals: 5;5 (kindergarten), 6;3 (grade 1), 7;9 (grade 2) Monolinguals: 5;8 (kindergarten), 6;2 (grade 1), 7;3 (grade 2) Cross-sectional design	Phoneme substitution -without cue -with sound cue -with picture cue	English	No overall differences Bilinguals > monolinguals: no-cue condition Monolinguals > bilinguals: sound condition	Mixed
Bialystok et al., 2003 (study II)	English children in French schools (*n* = 36) English monolinguals (*n* = 39)	Bilinguals: 5;9 (kindergarten), 6;10 (grade 1), 7;8 (grade 2) Monolinguals: 6;0 (kindergarten), 7;0 (grade 1), 7;10 (grade 2) Cross-sectional design	Phoneme substitution -without cue -with sound cue -with picture cue	English and French	No overall differences Sound cue condition: Monolinguals > bilinguals	Negative
Bialystok et al., 2003 (study III)	Spanish-English bilinguals (*n* = 25), Chinese-English bilinguals (*n* = 31) [Mandarin & Cantonese], English monolinguals (*n* = 33)	Spanish-English: 6;7 (grade 1), 7;4 (grade 2) Chinese-English: 6;6 (grade 1), 7;6 (grade 2) Monolingual: 6;7 (grade 1), 7;5 (grade 2) Cross-sectional design	Sound-meaning task [rhymes with target, means same as target] Phoneme substitution (with sound cue) Phoneme segmentation	English	Spanish-English > monolingual > Chinese-English: phoneme segmentation	Mixed

					English-Greek bilingual > English monolingual: Cluster onset oddity, initial phoneme identification, single phoneme onset oddity, phoneme elision	
Loizou & Stuart, 2003	English-Greek bilinguals (*n* = 16), Greek-English bilinguals (*n* = 18), English monolinguals (*n* = 16), Greek monolinguals (*n* = 18)	English-Greek: 5;10 Greek-English: 5;10 English monolingual: 5;9 Greek monolingual: 5;9	Rhyme oddity Syllable completion Onset oddity Initial phoneme identification Single phoneme onset oddity Phoneme elision	Greek and English	Greek bilingual = Greek monolingual Initial phoneme identification (p = .059) and Phoneme elision (p = .051) in favour of monolinguals English-Greek > Greek-English: compound score small units (initial phoneme identification, single phoneme onset oddity, phoneme elision) and compound score large units (rhyme oddity, syllable completion, onset oddity)	Mixed

			Greek tasks: English-Greek > Greek-English, Greek monolinguals: compound score small units English tasks: English-Greek > Greek-English, English monolinguals: compound score small units, compound score large units

Chen et al., 2004 (study I)	Cantonese children in Mandarin immersion (*n* = 170) Mandarin monolinguals (*n* = 105)	grade 2, grade 4 (no age defined) Cross-sectional design	Tone awareness Onset awareness Rime awareness	Mandarin	Cantonese-Mandarin > Mandarin monolingual: onset awareness, rime awareness (both grade 2 only, and only on items with onsets that exist only in Mandarin) Mandarin monolingual > Cantonese-Mandarin: onset awereness (grade 4 only, and only on items with onsets that exist in Mandarin but are similar to Cantonese)	Mixed

Chen et al., 2004 (study II)	Cantonese children in Mandarin immersion (*n* = 136) Mandarin monolinguals (*n* = 153)	Grade 1, grade 2, grade 4 (no age defined) Cross-sectional design	Tone awareness Onset awareness Rime awareness (all assessed with both oddity task and same-different task)	Mandarin (tone, onset, rime) Words that are shared in both languages: onset, rime	Cantonese-Mandarin > Mandarin monolingual: tone awareness (grade 1, less so in grade 2, not in grade 4), onset awareness (grade 2 only, on shared words and non-words, and only on oddity task), rime awareness (grade 2 and 4, only on shared words in oddity task), rime awareness (shared words and pseudowords, in same-different task) Mandarin monolingual > Cantonese-Mandarin: onset awareness (grade 4 only, on Mandarin words) IQ as covariate in all analyses	Mixed

Bialystok et al., 2005	Cantonese-English bilinguals (*n* = 29) Hebrew-English bilinguals (*n* = 30) Spanish-English bilinguals (*n* = 33) English monolinguals (*n* = 40)	Cantonese-English bilinguals: 6;6 Hebrew-English bilinguals: 6;8 Spanish-English bilinguals: 6;10 English monolinguals: 6;9	Phoneme counting	English Hebrew Spanish Cantonese	In English: Hebrew-English, Spanish-English > Cantonese-English, monolinguals Controlling for age (because Cantonese-English bilinguals are younger)	Positive
Chen et al., 2010 (study I)	Chinese children receiving English instruction (80 min/week) (*n* = 185) Chinese monolinguals (*n* = 177)	Chinese-English: 7;3 (grade 1), 9;3 (grade 3) Chinese monolingual: 7;4 (grade 1), 9:3 (grade 3) Cross-sectional design	Onset Rime	Chinese Pinyin and English	Chinese-English > Chinese monolingual: English onset and rime (grade 1 and 3), Chinese onset and Chinese rime (grade 3 only)	Positive
Chen et al., 2010 (study II)	Chinese children receiving intensive English instruction (120 min/day) (*n* = 79) Chinese children receiving regular English instruction (80 min/week) (*n* =	Intensive: 6;4 Regular: 6;5 at T1 Longitudinal design from start grade 1 to end grade 2, tested five times	Onset awareness Rime awareness Phoneme awareness Tone awareness	Chinese Pinyin and English	Intensive > regular: Chinese (T5 only): onset, rime, phoneme awareness Intensive > regular: English: phoneme (T3, T4, T5), rime (T4, T5), onset, (T5)	Positive

Marinova-Todd et al., 2010	Mandarin-English bilinguals (*n* = 62) English monolinguals (*n* = 21) Mandarin monolinguals (*n* = 61)	Mandarin-English: 5;6, 6;4 English monolingual: 5-6 Mandarin monolingual: 5;4, 6;5	*Mandarin:* Tone discrimination Syllable deletion Onset-rime combination Initial sound identification Rhyme detection *English*: Elision (speech segements from words) Blending Sound matching (initial and final phonemes)	Mandarin and English	Mandarin-English > English monolingual: Elision, Blending Mandarin-English > Mandarin monolingual: 5- and 6-year-olds: Onset-rime, Initial sound, Rhyme, 6-year-olds: Tone	Positive
Kuo & Anderson, 2010	Mandarin-Southern-Min bilingual (*n* = 134) Mandarin monolingual (*n* = 59)	Kindergarten (6 year-olds), grade 1, grade 2 (no age defined) Cross-sectional design	Onset awareness Rime awareness Tone awareness	Mandarin (but only syllables tested)	Mandarin-Southern-Min > monolingual Mandarin: onset awareness (Kindergarten and grade 1 only), rhyme awareness on items involving novel syllables, tone awareness on items involving novel syllables	Positive

Kang, 2012	Korean children attending English kindergarten (*n* = 70) Korean [functionally] monolingual (15 min English/day) (*n* = 56)	Korean-English bilingual: 5;11 Korean monolingual: 5;9	Phoneme awareness Rime awareness *Korean only*: Syllable awareness Body-coda awareness	English and Korean	Korean-English > monolingual Korean: Korean: syllable, body-coda, phoneme, rime Korean-English > monolingual Korean: English: phoneme, rime	Positive
Reder et al., 2013	French children enrolled in German partial immersion (50%) schools (*n* = 43) French monolinguals (*n* = 52)	French-German immersion: 6;7 French monolingual: 6;8	Phonological categorization (syllable and phoneme) Phonological deletion (syllable and phoneme)	French	No differences [French-German immersion > monolingual: deletion (p = .06)]	No differences
Lesniak et al., 2014	Polish-English bilinguals (*n* = 18) Portuguese-English bilinguals (*n* = 18) English monolinguals (*n* = 18)	Polish-English bilingual: 5;11 Portuguese-English bilingual: 5;11 English monolingual: 6;3	Syllable segmentation Syllable identification Alliteration (similar onset phoneme) Rhyme detection Rhyme generation Phoneme identification (detect onset phoneme) Phoneme segmentation	English	Monolingual > Polish-English, Portuguese-English : alliteration, rhyme generation, phoneme identification, phoneme segmentation Monolingual > Portuguese-English: rhyme detection	Negative

Janssen et al., 2015	Turkish-Dutch sequential bilinguals (*n* = 29) Dutch monolinguals (*n* = 32)	Turkish-Dutch bilingual: 4;9 Dutch monolingual: 4;8	Rhyme awareness Phoneme blending	Dutch	Monolingual > Turkish-Dutch: Rhyme awareness	Negative
Kuo et al., 2016	Native English enrolled in a Japanese-English (50/50) school (*n* = 41) Native Japanese enrolled in a Japanese-English (50/50) school (*n* = 40) Monolingual English (*n* = 48)	English-Japanese: 7;9 Japanese-English: 7;2 English monolingual: 7;7	Onset-awareness -real word -non word	English	No overall difference Controlling for English vocabulary (Monolingual > Japanese-English) Japanese-English > English-Japanese > monolingual: Onset-awareness items with onset phonemes that are shared between Japanese and English	Mixed
Janssen et al., 2017	Turkish-Dutch sequential bilinguals (*n* = 64) Dutch monolinguals (*n* = 75)	Turkish-Dutch bilingual: 5;1 Dutch monolingual: 4;11	Rhyme awareness Phoneme blending	Dutch	Monolingual > Turkish-Dutch: Rhyme awareness	Negative

*Positive: results point to advantage for bilinguals (with or without null results); negative: results point to advantage for monolinguals (with or without null results); mixed: results show advantages as well as disadvantages for bilinguals; no differences: only null results

Samenvatting in het Nederlands

In heel Europa bieden steeds meer basisscholen vroeg vreemdetalenonderwijs (vvto) aan. Nederland is daarop geen uitzondering. Momenteel is ongeveer één op de vijf Nederlandse basisscholen een vvto-school. Hoewel scholen ook de mogelijkheid hebben om Frans, Duits of Spaans als vreemde taal te onderwijzen, kiest het merendeel van de scholen (>90%) voor Engels als vreemde taal. Op vvto-scholen krijgen kinderen vanaf het moment dat zij de school betreden gemiddeld één uur per week Engelse les aangeboden. Op scholen waar het traditionele Engels in het basisonderwijs (Eibo) onderwezen wordt, starten de Engelse lessen in groep 7 en wordt er gemiddeld 45 minuten Engels per week gegeven. Hoewel het aantal scholen dat vvto aanbiedt snel toeneemt is er weinig bekend over de effecten van dit type onderwijs op de ontwikkeling van basisschoolleerlingen. In dit proefschrift wordt onderzocht welke invloed vvto heeft op de ontwikkeling van de taal en cognitie van leerlingen.

Kinderen die naar een vvto-school gaan komen op veel jongere leeftijd op school in aanraking met het Engels dan kinderen die naar een Eibo-school gaan. Engels heeft echter ook een prominente rol in de Nederlandse samenleving: films worden bijvoorbeeld ondertiteld en niet nagesynchroniseerd. Beide groepen kinderen hebben daarom waarschijnlijk enige kennis van het Engels. Wel is het aannemelijk dat de kennis bij vvto-leerlingen groter is dan bij Eibo-leerlingen. Hun kennis is echter niet te vergelijken met die van kinderen die van huis uit Nederlands-Engels tweetalig opgevoed worden en die vanaf zeer jonge leeftijd dagelijks aan beide talen blootgesteld worden. Wanneer kinderen tweetalig opgevoed worden kan dat invloed hebben op hun ontwikkeling, zowel op het gebied van taal als op het gebied van cognitie (Barac, Bialystok, Castro, & Sanchez, 2014). Het is onbekend of vvto, waarbij kinderen in veel mindere mate aan twee talen worden blootgesteld, een vergelijkbare invloed heeft. In dit proefschrift wordt dit voor drie ontwikkelingsdomeinen onderzocht. De centrale vraag is of de ontwikkeling van *de executieve functies, het fonologisch bewustzijn* en *de perceptie van Engelse spraakklanken* van vvto-leerlingen verschilt van die van Eibo-leerlingen en in hoeverre hun ontwikkeling lijkt op die van tweetalige kinderen.

Hoofdstuk 2 beschrijft een onderzoek naar de ontwikkeling van executieve functies van vvto- en Eibo-leerlingen. Executieve functies zijn hersenprocessen die zorgen voor de uitvoering van doelgericht en flexibel gedrag. In deze studie worden drie sub-processen onderzocht: switching (ook wel flexibiliteit genoemd), inhibitie en werkgeheugen. Switching is het vermogen van mensen zich aan te passen aan wisselende regels of omstandigheden. Inhibitie betreft het vermogen irrelevante responsen te onderdrukken en te kiezen voor meer doelgerichte gedragingen. Werkgeheugen betreft het vermogen om informatie te onthouden en te manipuleren. Het werkgeheugen kent een verbale en een niet-verbale component (Diamond, 2013).

De twee talen van tweetalige kinderen zijn beide voortdurend actief en met elkaar in competitie. Tweetalige kinderen zijn daarom constant – maar grotendeels onbewust – bezig de lexicale competitie tussen hun twee talen te beheersen. Dat

beheersingsproces doet een beroep op componenten die gerelateerd zijn aan de executieve functies, zoals het wisselen tussen beide talen en het onderdrukken van de op dat moment niet gebruikte taal. Deze taalcontrole zou de ontwikkeling van de executieve functies stimuleren (Green, 1998; Green & Abutalebi, 2013). Uit eerder onderzoek naar tweedetaalleerders blijkt dat een voorwaarde voor dit stimuleringsproces is dat beide talen in gelijke mate ontwikkeld zijn, zodat beide talen in gelijke mate geactiveerd worden en direct met elkaar in competitie zijn (Blom, Küntay, Messer, Verhagen, & Leseman, 2014; Thomas-Sunesson, Hakuta, & Bialystok, 2018). Onderzoek naar de relatie tussen de verwerving van twee talen en de ontwikkeling van de executieve functies is echter uitgevoerd bij kinderen die in grotere mate aan de tweede taal blootgesteld waren dan bij vvto het geval is.

De eerste studie (beschreven in hoofdstuk 2) heeft twee doelen. Het eerste doel is om te achterhalen of de mate van balans in de ontwikkeling van beide talen gerelateerd is aan de ontwikkeling van de executieve functies bij kinderen die in beperkte mate ervaring hebben met een tweede taal. Het tweede doel is om te onderzoeken of vvto-leerlingen beter ontwikkelde executieve functies hebben dan Eibo-leerlingen. Aan dit onderzoek deden 204 leerlingen uit de groepen 1, 5 en 8 mee. Bij al deze kinderen werden de vaardigheid in switching, inhibitie, werkgeheugen, Engelse en Nederlandse woordenschat gemeten.

De resultaten van deze studie laten allereerst zien dat kinderen bij wie het Nederlands en het Engels in gelijkere mate ontwikkeld zijn, beter presteren op een taak die switching-vaardigheden meet dan kinderen bij wie de ontwikkeling in beide talen minder in balans is. De mate van balans tussen beide talen is niet gerelateerd aan de ontwikkeling van inhibitievermogens of het werkgeheugen. Ten tweede laat dit onderzoek zien dat de ontwikkeling van vvto- en Eibo-leerlingen identiek verloopt. Beide groepen leerlingen behalen gelijke scores op switching-, inhibitie- en werkgeheugen-taken. Deze onderzoeksresultaten gelden voor alle leeftijdsgroepen (groep 1, 5, en 8).

Deze uitkomsten dragen bij aan het specificeren van de condities die ten grondslag liggen aan de versnelde ontwikkeling van de executieve functies bij kinderen die blootgesteld worden aan een tweede taal. De resultaten pleiten ervoor om in toekomstig onderzoek naar de executieve functies van tweedetaalleerders de balans tussen de ontwikkeling van beide talen mee te nemen, omdat de balans de mate van competitie tussen de betrokken talen lijkt te weerspiegelen.

In de tweede studie (hoofdstuk 3) staat de ontwikkeling van het fonologisch bewustzijn centraal. Het fonologisch bewustzijn is het vermogen om woorden op te delen in klanken en om klanken te manipuleren (Anthony & Francis, 2005). De ontwikkeling van het fonologisch bewustzijn verloopt stapsgewijs, waarbij kinderen eerst leren grotere delen van woorden te herkennen en manipuleren, zoals bijvoorbeeld het geval is bij rijmen, alvorens ze leren fonemen te herkennen en manipuleren. Eerder onderzoek met kinderen die twee talen leren – in zowel een educatieve als in een natuurlijke omgeving – hebben verschillende uitkomsten laten zien: de ontwikkeling van het fonologisch bewustzijn kan vertraagd zijn, versneld zijn, of niet verschillen van die van eentalige kinderen.

De in hoofdstuk 3 beschreven studie heeft drie doelen. Het eerste doel is om na te gaan of tweetaligheid de ontwikkeling van het fonologisch bewustzijn

stimuleert of juist belemmert. Het tweede doel is om te achterhalen welke mate van blootstelling aan een tweede taal noodzakelijk is voordat verschillen in de ontwikkeling van het fonologisch bewustzijn tot uiting komen. Het derde doel is om na te gaan in hoeverre individuele verschillen – in de ontwikkeling van de Nederlandse woordenschat, de Engelse woordenschat, de balans tussen de ontwikkeling van de Nederlandse en die van de Engelse woordenschat, het werkgeheugen, het kortetermijngeheugen, en leeftijd – de relatie tussen tweetaligheid en het fonologisch bewustzijn beïnvloeden. Daartoe werden drie groepen kinderen vergeleken: Eibo-leerlingen, vvto-leerlingen, en leerlingen die van huis uit Engels-Nederlands tweetalig opgroeien (n = 294 in totaal). De kinderen zaten in groep 1, 2 of 3 van de basisschool. Vier onderdelen van het fonologisch bewustzijn werden gemeten: (eind)rijm, het samenvoegen van individuele fonemen tot één woord, het bepalen van het eerste foneem van een woord, en het weglaten van een aangegeven foneem uit een bestaand woord, zodat een nieuw woord gevormd wordt.

Deze studie laat zien dat Eibo-, vvto-, en tweetalige leerlingen een vergelijkbare ontwikkeling in het fonologisch bewustzijn vertonen. Verder blijken leeftijd, de ontwikkeling van de Nederlandse woordenschat, het kortetermijngeheugen, en in mindere mate de Engelse woordenschat en het werkgeheugen positief gerelateerd te zijn aan het fonologisch bewustzijn. De balans tussen de ontwikkeling van de Nederlandse en de Engelse woordenschat is niet gerelateerd aan het fonologisch bewustzijn. Daarmee toont deze studie aan dat het op jonge leeftijd aanbieden van een tweede taal geen positieve maar ook geen negatieve invloed heeft op de ontwikkeling van het fonologisch bewustzijn.

Het derde ontwikkelingsdomein dat in dit proefschrift centraal staat is dat van de klankperceptie. Wanneer baby's geboren worden zijn zij in staat alle verschillende klankcontrasten van elkaar te onderscheiden. Dit geldt ook voor de klankcontrasten in talen die niet hun moedertaal zijn en die voor volwassen tweedetaalleerders moeilijk zijn. Al in de eerste maanden na de geboorte verdwijnt dat vermogen en zijn kinderen niet (veel) beter dan volwassenen in het onderscheiden van moeilijke klankcontrasten in talen die niet hun moedertaal zijn (Kuhl, 2004). Hoofdstuk 4 beschrijft een studie naar het vermogen van Eibo-, vvto-, en volledig Engels-Nederlands tweetalige leerlingen om verschillende Engelse spraakklanken te onderscheiden. De vraag die daarbij centraal staat is of vvto-leerlingen beter zijn dan Eibo-leerlingen in het onderscheiden van Engelse spraakklanken en of hun vermogens om dergelijke klanken te onderscheiden vergelijkbaar zijn met die van volledig tweetalige kinderen.

De leerlingen, uit groep 1, 5 en 8 (n = 160 in totaal), werden getest op vier Engelse klankparen: /b/-/s/ (makkelijk), /k/-/g/ (gemiddeld), /f/-/θ/ (moeilijk), en /ɛ/-/æ/ (zeer moeilijk). Uit dit onderzoek blijkt dat vvto-leerlingen niet beter in staat zijn om Engelse klanken van elkaar te onderscheiden dan Eibo-leerlingen. Daarnaast blijkt dat beide groepen minder goed zijn in het onderscheiden van Engelse klanken dan Engels-Nederlands tweetalige kinderen. Oudere kinderen blijken de Engelse klanken beter van elkaar te onderscheiden dan jongere kinderen. Voor de Engels-Nederlands tweetalige kinderen geldt dat voor alle hier geteste klankparen (/b/-/s/, /k/-/g/, /f/-/θ/, /ɛ/-/æ/). Voor Eibo- en vvto-leerlingen geldt dat leeftijd geen rol

speelt in het onderscheiden van /ɛ/ en /æ/. Deze kinderen zijn namelijk ongeacht hun leeftijd niet in staat /ɛ/ van /æ/ te onderscheiden. De verklaring die hiervoor gegeven wordt, is dat beide klanken van dit klankpaar in het Nederlands niet voorkomen, in tegenstelling tot de andere klankparen waarvan beide (/b/-/s/) of één van de twee klanken (/k/ en /f/) wel in het Nederlands bestaan. Moedertaalsprekers van het Nederlands zullen zowel /ɛ/ als /æ/ verstaan als een Nederlandse /ɛ/, die anders klinkt dan de Engelse /ɛ/ of /æ/. Waar zij bij de andere klankparen na een aantal jaren blootstelling aan het Engels leren dat de Engelse klank anders klinkt dan de klank die wel in het Nederlands voorkomt, blijft het voor hen ondoenlijk om de /ɛ/ en /æ/ van elkaar te onderscheiden.

De vraag rijst waarom vvto-leerlingen niet beter in staat zijn om Engelse klanken van elkaar te onderscheiden dan Eibo-leerlingen. Een mogelijke oorzaak kan liggen in de manier waarop de Engelse lessen vorm krijgen. Er lijkt in die lessen weinig aandacht te zijn voor het verstaan en produceren van verschillende Engelse klanken. De aanbeveling naar aanleiding van deze studie is dan ook om daar in het basisonderwijs tijdens de lessen Engels meer aandacht aan te besteden.

In alle voorgaande hoofdstukken is steeds gebruik gemaakt van de Peabody Picture Vocabulary Test (PPVT; Dunn & Dunn, 2007) om de passieve Engelse woordenschat van kinderen te meten. De PPVT is een veelgebruikte woordenschattoets, ook in studies met participanten voor wie het Engelse niet de eerste taal is. De PPVT is echter in eerste instantie ontwikkeld voor moedertaalsprekers van het Engels. In hoofdstuk 5 wordt verslag gedaan van een onderzoek waarin antwoord gegeven wordt op de vraag of de 4^{e} editie van de PPVT (PPVT-4) een betrouwbaar instrument is om de Engelse woordenschat van Nederlandssprekende basisschoolleerlingen (groep 1, 5 en 8; n = 204) en middelbare scholieren (jaar 1 en 3 van het vwo; n = 99) te meten. Ook wordt onderzocht of de kans dat leerlingen de betekenis van de Engelse woorden in de PPVT-4 kennen, samenhangt met de frequentie waarmee dat Engelse woord in het algemeen voorkomt, de frequentie waarmee het Nederlandse equivalent van dat woord in het algemeen in het Nederlands voorkomt en de mate waarin het Engelse woord en de Nederlandse vertaling van dat woord hetzelfde klinken. Daarnaast wordt onderzocht of er een verschil is tussen de Engelse woordenschatontwikkeling van Eibo- en vvto-leerlingen en tussen die van middelbare scholieren die wel of niet de tweetalige onderwijsstroom volgen.

Dit onderzoek laat zien dat de PPVT-4 geen betrouwbaar instrument is om de Engelse woordenschat te meten bij kinderen die zeer beperkt vaardig zijn in het Engels. Het instrument wordt echter betrouwbaarder naarmate kinderen een betere Engelse taalvaardigheid hebben. Daarnaast blijkt dat er meer kans is dat jonge kinderen (die ook minder vaardig zijn in het Engels) de betekenis van een Engels woord kennen wanneer dat woord in het Nederlands vaker voorkomt. Komt het woord in het Nederlands minder vaak voor, dan is de kans kleiner dat jonge kinderen de betekenis van het Engelse equivalent zullen kennen. Voor oudere kinderen is de frequentie van de Nederlandse vertaling niet van belang. Voor hen geldt dat er een grotere kans is dat zij de betekenis van een woord uit de PPVT weten wanneer dat woord in het Engels vaker voorkomt. Voor zowel jongere als oudere kinderen geldt dat er een grotere kans is dat zij de betekenis van een woord

kennen wanneer de uitspraak van dat woord meer lijkt op de uitspraak van de Nederlandse vertaling.

Daarnaast blijkt dat de er nauwelijks verschillen zijn tussen vvto-leerlingen en Eibo-leerlingen in de ontwikkeling van hun Engelse woordenschat. Alleen in groep 8 is er een verschil zichtbaar in het voordeel van vvto-leerlingen. In groep 1 en groep 5 zijn de prestaties van beide groepen leerlingen gelijk. Voor middelbare scholieren die de tweetalige stroom volgen geldt wel dat hun Engelse woordenschat duidelijk beter ontwikkeld is dan die van hun leeftijdsgenoten die het reguliere, Nederlandstalige, programma volgen.

In dit hoofdstuk wordt de aanbeveling gedaan om de PPVT niet op de wijze af te nemen die in de handleiding beschreven staat, dus om niet met de leeftijds-adequate woorden te beginnen. In plaats daarvan zou de PPVT afgenomen moeten worden door te beginnen bij een item dat bedoeld is voor kinderen die jonger zijn dan het kind bij wie de toets daadwerkelijk afgenomen wordt. Bovendien wordt opgemerkt dat voorzichtigheid geboden is bij het interpreteren van resultaten uit onderzoek waarbij de PPVT gebruikt wordt om tweedetaalleerders met verschillende moedertalen te vergelijken. Talen die verwant zijn aan het Engels kennen immers veel woorden die hetzelfde uitgesproken worden als hun Engelse equivalent. Voor talen die niet verwant zijn aan het Engels is dat in veel mindere mate het geval. Kinderen die een moedertaal hebben die in de eerste categorie valt, ervaren daarvan meer voordeel op de PPVT dan kinderen die een moedertaal hebben uit de laatste categorie.

In het slothoofdstuk (hoofdstuk 6) worden de voorgaande studies geïntegreerd besproken. Daarbij wordt gereflecteerd op de Nederlandse en de Engelse woordenschatontwikkeling van leerlingen, die in elk van de drie onderzoeksrondes gemeten werd. De onderzoeken samen leveren een eenduidig resultaat op voor de Nederlandse woordenschat: die is bij Eibo- en vvto-leerlingen in gelijke mate ontwikkeld. Voor de Engelse woordenschat is het beeld minder duidelijk. Zo presteerden vvto-leerlingen in groep 5 niet beter op de Engelse woordenschattoets dan hun leeftijdsgenootjes van Eibo-scholen, terwijl vvto-leerlingen uit groep 2 en 3 hier wel beter op presteerden dan Eibo-leerlingen uit diezelfde groepen. Voor leerlingen uit groep 1 en groep 8 geldt dat er soms wel een verschil tussen beide groepen zichtbaar is en soms niet. Wanneer er een verschil zichtbaar is, is dat in het voordeel van vvto-leerlingen.

Deze gemengde uitkomsten kunnen wellicht deels verklaard worden door verschillen tussen scholen. Sommige scholen die meededen aan het onderzoek hadden een certificaat dat aantoont dat de kwaliteit van hun Engelse lessen op orde is terwijl andere dat niet hadden. Desalniettemin zijn het niet altijd de leerlingen van deze scholen die beter presteren dan leerlingen van Eibo-scholen. Het niet hebben van een certificaat betekent dan ook niet dat de kwaliteit van de Engelse lessen niet voldoende is. Een andere reden voor de gemengde uitkomsten kan zijn dat sommige scholen meer nadruk leggen op de ontwikkeling van de woordenschat tijdens de lessen Engels en dat deze nadruk weerspiegeld wordt in de scores die de kinderen behalen op een Engelse woordenschattoets. Ten slotte kan het ook zo zijn dat sommige leerlingen buiten schooltijd meer blootgesteld worden aan het Engels dan andere leerlingen. Uit de aanwezige data uit oudervragenlijsten (n = 250) blijkt dat

vvto- en Eibo-leerlingen in gelijke mate met Engelse media in aanraking komen. Deze verklaring lijkt dus niet plausibel.

In het slothoofdstuk worden ook de in hoofdstuk 2 gerapporteerde resultaten over de ontwikkeling van de executieve functies bij vvto- en Eibo-leerlingen aangevuld met de eerste resultaten van onderzoek naar de ontwikkeling van de executieve functies bij kinderen die Engels-Nederlands tweetalig opgevoed worden (n = 74). Bij deze kinderen werd de ontwikkeling van vaardigheden, werkgeheugen, en Nederlandse en Engelse woordenschat gemeten. De resultaten laten zien dat, net als bij vvto- en Eibo-leerlingen, de switching-vaardigheden van tweetalige kinderen beter ontwikkeld zijn naarmate hun Nederlandse en Engelse woordenschat meer met elkaar in balans zijn. Daarnaast blijkt dat tweetalige kinderen een beter verbaal werkgeheugen hebben dan Eibo- en vvto-leerlingen. Ook hebben 11- en 12-jarige tweetalige kinderen betere switching-vaardigheden dan hun Eibo- en vvto-leeftijdsgenoten.

Samenvattend laten de resultaten van dit proefschrift zien dat minimale blootstelling aan een tweede taal al invloed kan hebben op de cognitieve ontwikkeling en de taalontwikkeling van kinderen. Deze invloeden zijn echter beperkt en niet vergelijkbaar met de invloed van tweetalig opgroeien. Zo is bij zowel tweedetaalleerders als bij volledig tweetalige kinderen de balans in de ontwikkeling van de woordenschat in beide talen positief van invloed op het switching-vermogen. Tweetaligen laten echter betere vaardigheden zien in switching, werkgeheugen en het verstaan van Engelse klanken dan vvto- en Eibo-leerlingen, terwijl tussen die laatste twee groepen geen verschil te zien is. Dit suggereert dat kinderen een bepaalde mate van tweetaligheid moeten hebben voordat zich verschillen in de ontwikkeling voordoen. Dit onderzoek laat zien dat vvto niet voldoende is om die grens te halen en dus niet voldoende is om de ontwikkeling van de executieve functies, het fonologisch bewustzijn, of het verstaan van Engelse klanken positief te beïnvloeden. Tegelijkertijd zijn er geen negatieve effecten van het op jonge leeftijd aanbieden van een tweede taal.

De overheid, leerkrachten en ouders hebben vaak positieve en soms hoge verwachtingen van vroeg vreemdetalenonderwijs. Dergelijke verwachtingen moeten op basis van dit onderzoek bijgesteld worden: vroeg vreemdetalenonderwijs is geen garantie voor een succesvolle ontwikkeling van de Engelse taalvaardigheid noch lijkt het een positieve invloed te hebben op de cognitieve ontwikkeling. Met andere woorden: vroeg-vreemdetalenonderwijs is geen wondermiddel voor de taalontwikkeling en de cognitieve ontwikkeling van basisschoolleerlingen.

Referenties

Anthony, J. L., & Francis, D. J. (2005). Development of phonological awareness. *Current Directions in Psychological Science*, *14*(5), 255–259. http://doi.org/10.1111/j.0963-7214.2005.00376.x

Barac, R., Bialystok, E., Castro, D. C., & Sanchez, M. (2014). The cognitive development of young dual language learners: A critical review. *Early Childhood Research Quarterly*, *29*, 699–714. http://doi.org/10.1016/j.ecresq.2014.02.003

Blom, E., Küntay, A. C., Messer, M., Verhagen, J., & Leseman, P. (2014). The

benefits of being bilingual: Working memory in bilingual Turkish-Dutch children. *Journal of Experimental Child Psychology*, *128*, 105–119. http://doi.org/10.1016/j.jecp.2014.06.007

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*, 135–168. http://doi.org/10.1146/annurev-psych-113011-143750

Dunn, L. M., & Dunn, D. M. (2007). *PPVT-4 Manual*. Bloomington: NCS Pearson.

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, *1*(2), 67–81. http://doi.org/10.1017/S1366728998000133

Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology*, *25*, 515–530. http://doi.org/10.1080/20445911.2013.796377

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience*, *5*, 831–843. http://doi.org/10.1038/nrn1533

Thomas-Sunesson, D., Hakuta, K., & Bialystok, E. (2018). Degree of bilingualism modifies executive control in Hispanic children in the USA. *International Journal of Bilingual Education and Bilingualism*, *21*, 197–206. http://doi.org/10.1080/13670050.2016.1148114

Summary in English

All across Europe primary schools have started to lower the starting age for learning a foreign language. The Netherlands are no exception to this trend. Currently, one in five Dutch primary schools offers foreign language instruction from early on. Even though schools have the possibility to offer French, German, or Spanish as the foreign language, most of these schools (>90%) opt for English. Contrary to mainstream schools, in which English lessons are given from the penultimate grade (around age ten) for about 45 minutes per week, pupils in these early-English language schools receive around 60 minutes of English lessons from the moment they enter school at the age of four. Despite the fact that the number of early-English schools is ever growing, little is known about the effects of this type of education on the development of primary-school pupils. This thesis investigates to what extent early-English education influences pupils' cognitive and linguistic development.

Pupils who attend an early-English school are exposed to English in an instructed setting at a much younger age than pupils who attend a mainstream school. English also plays a prominent role in Dutch society: television programmes are for example subtitled instead of dubbed. As a result, both groups of pupils are likely to have some knowledge of English, but early-English pupils' knowledge is likely to be greater than that of mainstream pupils. Their knowledge of English will however unlikely be comparable to that of children who are being raised bilingually with both Dutch and English as their native languages, and who are exposed to both languages on a daily basis. This latter kind of bilingual experience may influence children's linguistic and cognitive development (Barac, Bialystok, Castro, & Sanchez, 2014), but it is unknown whether early-English education has the same effects. This thesis investigated this issue for three developmental domains. The overarching question is whether the development of *executive functions*, *phonological awareness*, and *the perception of English speech sounds* of early-English pupils is different from that of their mainstream peers, and whether it resembles that of (simultaneous) bilingual children.

Chapter 2 reports on a study on the development of executive functions of early-English and mainstream pupils. Executive functions are brain processes that are involved in the execution of goal-directed and flexible behaviour (Diamond, 2013). In this study, they are divided into three sub-processes: switching (flexibility), inhibition, and working memory. Switching is the ability to adjust to changing rules or circumstances. Inhibition concerns the ability to suppress irrelevant responses and to choose more deliberate behaviours. Finally, working memory is the ability to hold information in mind and to manipulate this information. Working memory can again be sub-divided in a verbal and a nonverbal component.

A bilingual child's two languages are always both active and in competition with each other. It is assumed that bilingual children are therefore constantly – but mostly unconsciously – monitoring the lexical competition between their two languages. This monitoring process calls on components that are related to the

executive functions, like switching between languages and inhibiting the language that is not used at a particular moment. The language monitoring process would therefore stimulate the development of executive functions (Green, 1998; Green & Abutalebi, 2013). Previous studies with second-language learners have shown that a prerequisite for this effect is that both languages are developed to an equal extent, and are therefore activated to comparable levels resulting in competition (Blom, Küntay, Messer, Verhagen, & Leseman, 2014; Thomas-Sunesson, Hakuta, & Bialystok, 2018). The available previous research has been conducted with children with considerably more exposure to the second language than is the case for early-English pupils.

The goal of this study was twofold: first to investigate whether the balance between the development of both languages is positively related to the development of executive functions in children who have little experience with a second language, and second to investigate whether early-English pupils show better developed executive functions than mainstream pupils. To answer this question, 204 pupils from grade 1 (4-5 years old), grade 5 (8-9 years old), and grade 8 (11-12 years old) were tested. All (of these) children performed tasks that measured their abilities in switching, inhibition, working memory, and their English and Dutch vocabulary.

The results showed that first, children whose Dutch and English vocabularies were more balanced, performed better on a switching task than children who showed less balanced Dutch-English vocabularies. The extent to which the development of both languages is balanced was not related to the development of inhibition or working memory. Second, this study revealed that the development of early-English pupils is not different from that of mainstream pupils. Both groups of pupils obtained comparable scores on switching, inhibition and working memory tasks. Both results hold for children in all three age groups.

These results contribute to specifying the conditions required for the accelerated development of executive functions in children that are exposed to two languages. More specifically, they indicate that future research should take into account the balance between the development of two languages, because this balance seems to reflect the competition between the two languages.

The focus of Chapter 3 is the development of phonological awareness. Phonological awareness is the ability to divide words into sounds and to manipulate those sounds (Anthony & Francis, 2005). Phonological awareness develops gradually: children first learn to recognize parts of words and to manipulate these, as is for example the case with rhyming, before they learn to recognize and manipulate individual phonemes . Previous studies with children who are learning two languages – whether in instructed or natural settings – have shown different outcomes: their phonological awareness may develop more slowly, more quickly, or at a comparable rate to that of mainstream pupils.

The study in Chapter 3 has three goals. The first goal is to investigate whether bilingualism helps or hinders the development of phonological awareness. The second goal is to examine what amount of exposure to the second language is needed for such differences in phonological awareness to emerge. The third goal is to investigate to what extent individual differences in the development of Dutch

vocabulary, English vocabulary, the balance between these two languages, working memory, short-term memory, and age influence the relation between bilingualism and phonological awareness. To investigate this, three groups of children were compared: mainstream pupils, early-English pupils, and pupils who are raised as English-Dutch bilinguals (n = 294 in total). All children are in grade 1, 2 or 3 of primary school (and are thus 4-5, 5-6, or 6-7 years old). Four phonological awareness process are measured: rhyme, phoneme blending, identification of the onset phoneme in a word, and removing a phoneme from a word such that another word is formed.

This study found that mainstream, early-English and bilingual children show comparable development of phonological awareness skills. In addition, age, development of Dutch vocabulary, short-term memory, and to a lesser extent the development of English vocabulary and working memory are positively related to phonological awareness. The balance between Dutch and English vocabulary knowledge is not related to phonological awareness. These outcomes show that exposing children to a second language at a young age does not positively influence the development of phonological awareness. Importantly, there is no negative influence either.

The third developmental domain that is investigated is the domain of speech perception. When babies are born they are capable of distinguishing the speech sounds of all languages, even those not present in their mother tongue and which are hard for adult second-language learners. Within the first months of life that ability fades, and children become attuned to the speech sounds of their mother tongue (Kuhl, 2004). Chapter 4 reports a study in which mainstream, early-English and Dutch-English bilingual pupils' abilities to distinguish different English speech sounds are investigated. The central question is whether early-English pupils' abilities are better than mainstream pupils', and whether the abilities of early-English pupils are comparable to that of bilingual children.

The pupils, from grade 1 (4-5 years-old), grade 5 (8-9 years-old), and grade 8 (11-12 years-old), were tested on four different pairs of English speech sounds that differ in difficulty level for Dutch native speakers: /b/-/s/ (easy), /k/-/g/ (average), /f/-/θ/ (hard), and /ɛ/-/æ/ (very hard). This study shows that early-English pupils are not better than mainstream pupils in distinguishing English speech sounds. In addition, both groups' abilities are by no means comparable to that of bilingual children. Older children are better able to distinguishing English speech sounds than younger children. For English-Dutch bilingual children this holds for all pairs that were tested here (/b/-/s/, /k/-/g/, /f/-/θ/, /ɛ/-/æ/). For mainstream and early-English, this does not hold for the /ɛ/-/æ/ contrast: both older and younger pupils perform at chance level. The explanation put forward for this finding is that both sounds of this pair do not exist in Dutch, contrary to the other sound pairs of which both (/b/-/s/) or one of the two speech sounds (/k/ en /f/) also occurs in Dutch. Native speakers of Dutch will perceive both /ɛ/ and /æ/ as a Dutch /ɛ/, which sounds different from the English /ɛ/ or /æ/. For the former English speech sounds pupils will learn, after several years of exposure to English, that the speech sound that does not exist in Dutch is not the same as the one that does. This does not hold for the /ɛ/-/æ/ pair.

They will continue to be unable to perceive the two members of this pair as being different.

Why exactly early-English pupils are not better able to distinguish English speech sounds than mainstream pupils remains unclear. A possible reason could be the way in which English lessons are shaped. It seems that perceiving and producing English speech sounds is not given much attention in English education. It is therefore recommended to pay more attention to these skills.

All of the three aforementioned studies use the Peabody Picture Vocabulary Test (PPVT; Dunn & Dunn, 2007) to measure pupils' passive vocabulary in English. The PPVT is a frequently used vocabulary task, also in studies with participants whose first language is not English. The PPVT was however originally developed for native speakers of English. Chapter 5 presents a study which answers the question of whether the 4th edition of the PPVT (PPVT-4) is a reliable measure of English vocabulary in Dutch pupils who are either in primary school (grade 1, 5, or 8; n = 204) or in secondary school (year 1 or year 3 of the pre-university curriculum; n = 99). Futhermore, it explores whether the likelihood that pupils know the meaning of the English words in the PPVT is related to the frequency with which that word occurs in general in English, the frequency with which the Dutch translation of that word occurs in general in Dutch, and the extent to which the English word and its Dutch translation sound the same. In addition, the test was used to investigate whether there are any differences in the development of English vocabulary between mainstream and early-English primary-school pupils, or between secondary-school pupils who follow the bilingual curriculum and those who follow the mainstream Dutch curriculum.

This study showed that the PPVT-4 is not a reliable instrument to measure English vocabulary in pupils who have a very limited proficiency in English. The instrument becomes more reliable as children's proficiency levels in English increase. Furthermore, the chance that younger children (who also have lower proficiency levels in English) will know the meaning of a word that appears in the PPVT is higher when that word appears more frequently in Dutch. If it is a less-frequently used word, the chance is lower that young children will know the meaning of the English equivalent. For older children the frequency of use in Dutch is not relevant. For these children, it is the frequency of use of the English word that matters: if the word occurs more frequently in English, older pupils are more likely to know what that word means. Finally, both younger and older children are more likely to answer correctly to a word in the PPVT when the pronunciation of that English word resembles the pronunciation of the Dutch translation equivalent.

The study also showed that there are hardly any differences between early-English and mainstream pupils with respect to their English vocabulary. It was only for the oldest primary-school pupils (i.e., those in grade 8; 11-12 year-olds) that a difference in favour of the early-English pupils was visible. Pupils in grade 1 (4-5 year-olds) and grade 5 (8-9 year-olds) showed comparable performance on the PPVT. Secondary-school pupils who are enrolled in a bilingual curriculum do show a clear advantage in English vocabulary compared to their peers in the mainstream curriculum.

Chapter 5 recommends not following the manual when using the PPVT with second-language learners. Specifically, it is advised to start with an earlier set of words than the age-appropriate one. Moreover, it is noted that researchers should be cautious in interpreting results from research in which the PPVT is used to compare second-language learners with different first languages. The fact is that languages that are related to English have many more words that sound the same as their English translation equivalent. This is much less the case for languages that are not related to English. Consequently, children with a first language that belongs to the first category profit more from this when performing the PPVT than children who have a mother tongue that belongs to the second category.

The final chapter of this thesis (Chapter 6) brings together the findings from the four different studies. It includes a reflection on pupils' development of both Dutch and English vocabulary, which was measured in all previous studies. All studies showed a clear picture when it comes to the development of Dutch: both mainstream and early-English pupils show comparable vocabulary levels in Dutch. Such a clear picture is not visible for the development of English vocabulary. Whereas early-English pupils in grade 5 (8-9 years old) do not perform better on the English vocabulary test than their peers from mainstream schools, pupils in grade 2 (5-6 years old) and grade 3 (6-7 years old) do show an advantage compared to mainstream pupils. For grade 1 and grade 8 pupils (4-5 and 11-12 years old, respectively), results are mixed: sometimes there is a difference between the two groups of pupils, and sometimes there is not.

Possible reasons for these mixed results may be found in the fact that some schools have a certificate that shows that the quality of the English lessons is sufficient. Nevertheless, it is not always the pupils from these schools that outperform the mainstream pupils. Indeed, not obtaining a certificate does not mean that the quality of the English lessons is not sufficient. Another possible reason may be that some schools pay more attention to vocabulary during the English lessons, and that this focus is reflected in pupils' scores on the English vocabulary test. Finally, it may also be the case that some pupils receive more exposure to English out of school than others, which in turn results in higher vocabulary scores. The available parental questionnaire data (n = 250) however show that early-English and mainstream pupils are exposed to English media to the same extent. Therefore it seems unlikely that out-of-school exposure is a plausible explanation for this finding.

In addition to the results on the development of executive functions in mainstream and early-English pupils presented in Chapter 2, this final chapter also presents the preliminary results of research on English-Dutch bilingual children's executive functions. Tasks measuring switching, working memory, English vocabulary, and Dutch vocabulary were completed by 74 children who were being raised bilingually. The results show that, just as for mainstream and early-English pupils, bilingual children with more balanced Dutch and English vocabularies show better switching abilities. Furthermore, bilingual children show more developed verbal working memory abilities than mainstream and early-English pupils. In addition, 11- and 12-year-old bilingual children show better switching abilities than their mainstream and early-English peers.

In summary, the results of this thesis show that minimal exposure to a second language can already influence pupils' cognitive and linguistic development. At the same time, these influences are very limited and not comparable to the influence of growing up with two languages at the same time. For example, both in second-language learners and bilingual children language balance is positively related to switching abilities. However, bilinguals show better developed switching skills, working memory abilities, and perception of English speech sounds than early-English and mainstream pupils, while no such differences exist between the latter two groups. This suggests that children should reach a certain level of experience with two languages before such developmental differences can emerge. The research reported in this thesis shows that early-English education is not enough to reach that level, and that it does not influence pupils' executive functions, their phonological awareness, or their abilities to distinguish between English speech sounds. At the same time, exposing children to a second language at a young age does not *negatively* influence pupils' development either.

Policy makers, teachers, and parents often have positive and sometimes high expectations of early-English education. This thesis shows that such expectations should be tempered: early-English education by no means guarantees that pupils' English language skills will develop successfully, nor does it seem to have a positive influence on their cognitive development. It can be concluded that early-English education works no miracles for pupils' cognitive and linguistic development.

References

Anthony, J. L., & Francis, D. J. (2005). Development of phonological awareness. *Current Directions in Psychological Science*, *14*(5), 255–259. https://doi.org/10.1111/j.0963-7214.2005.00376.x

Barac, R., Bialystok, E., Castro, D. C., & Sanchez, M. (2014). The cognitive development of young dual language learners: A critical review. *Early Childhood Research Quarterly*, *29*, 699–714. https://doi.org/10.1016/j.ecresq.2014.02.003

Blom, E., Küntay, A. C., Messer, M., Verhagen, J., & Leseman, P. (2014). The benefits of being bilingual: Working memory in bilingual Turkish-Dutch children. *Journal of Experimental Child Psychology*, *128*, 105–119. https://doi.org/10.1016/j.jecp.2014.06.007

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*, 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

Dunn, L. M., & Dunn, D. M. (2007). *PPVT-4 Manual*. Bloomington: NCS Pearson.

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, *1*(2), 67–81. https://doi.org/10.1017/S1366728998000133

Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology*, *25*, 515–530. https://doi.org/10.1080/20445911.2013.796377

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience*, *5*, 831–843. https://doi.org/10.1038/nrn1533

Thomas-Sunesson, D., Hakuta, K., & Bialystok, E. (2018). Degree of bilingualism

modifies executive control in Hispanic children in the USA. *International Journal of Bilingual Education and Bilingualism*, *21*, 197–206. https://doi.org/10.1080/13670050.2016.1148114

Dankwoord

Er zijn veel mensen die mij de afgelopen jaren geholpen of gesteund hebben en die daarmee direct of indirect hun bijdrage geleverd hebben aan het tot stand komen van dit proefschrift. Aan hen ben ik mijn dank verschuldigd.

Mijn eerste en voornaamste woord van dank gaat uit naar mijn (co-)promotoren. James, als jij me destijds niet gewezen had op de vacature voor de promotieplekken binnen het CLS was dit proefschrift er niet geweest. Bedankt dat je het vertrouwen in mij had dat ik de geschikte kandidaat zou zijn en me aanmoedigde om te solliciteren. Ik ben ook erg blij dat je mijn promotor was, dat je je zo betrokken toonde en dat je aandacht had voor zowel de grote lijnen als de kleinste details van mijn onderzoek. Je was bovendien degene die mij met Mirjam, mijn dagelijks begeleidster, in contact bracht. Mirjam, jou wil ik bedanken dat je vanaf het allereerste begin uitstraalde dat je het vertrouwen in me had dat ik dit project tot een goed einde zou brengen. Je had bovendien altijd tijd om met me van gedachten te wisselen en je wist overal een positieve noot bij te zetten. Dat was precies wat ik nodig had. Roeland, mijn andere promotor, ik heb veel van jou geleerd, maar bovenal dat er voor elk statistisch probleem een oplossing is. Het was fijn dat je deur altijd open stond en dat ik met al mijn kleine en grote vragen bij je terecht kon. Sharon, jij maakte mijn rijtje begeleiders compleet. Je kritische opmerkingen bij mijn redeneringen daagden me uit om verder te denken. Je hielp me mijn ideeën naar een hoger niveau te tillen en daar werd mijn onderzoek alleen maar beter van. Bedankt daarvoor en voor het feit dat er bij jou altijd ruimte was voor een serieus gesprek met een vleugje humor. Vier begeleiders zijn er bovengemiddeld veel. Ik prijs me echter gelukkig dat elk van jullie mijn begeleider wilde zijn en dat ik daarmee een team van begeleiders had met een brede expertise. Het enige probleem was om een tijdstip te vinden waarop we allemaal bij elkaar konden komen, maar het is elke eindeloze datumprikker waard geweest.

Dit onderzoek had nooit kunnen worden wat het is geworden zonder alle kinderen, ouders en leerkrachten die deelgenomen hebben. Ik ben alle participanten dan ook zeer erkentelijk dat zij aan mijn onderzoek mee wilden doen, en dat ik welkom was binnen alle scholen en gezinnen.

Er hebben aan dit onderzoeksproject honderden participanten deelgenomen. Dat hadden er nooit zo veel kunnen zijn zonder de aanvullende financiële donatie van Nuffic, de organisatie voor internationalisering in het Nederlandse onderwijs. Ik ben dankbaar dat Nuffic de waarde van mijn onderzoek inzag en bereid was financiële steun te verlenen.

Het testen van alle participanten en het verwerken van de data was een logistieke puzzel die ik niet had kunnen voltooien zonder de vele stage- en scriptiestudenten die mij geholpen hebben: Maartje, Jennifer, Nina, Nadine, Anne, Eline, Laura, Monique, Marthe, Loes, Laurie, Elisa, Leonie, Minouck, Marjolein, Doortje, Daphne, Nikki, Jason, Vera, Lisette, Yvet, Eva, Marieke, Iris, Anna, Alicia, Maaike, Denise, bedankt voor jullie inzet en hulp.

Margret en Bob, dank jullie wel voor jullie hulp met het verzorgen van het testmateriaal. Door jullie had ik de benodigde laptops op het juiste moment, en met werkende testen bovendien.

Ik was als promovenda verbonden aan de afdeling Taalwetenschap. Al snel nadat ik daar kwam te werken bleek dat dat een warme vriendelijke afdeling was waar ik me thuis voelde. Mijn collega's wil ik dan ook bedanken voor hun collegialiteit en gezelligheid. Een aantal mensen wil ik daarbij in het bijzonder bedanken.

Ferdy, jij en ik deelden drieënhalf jaar lang een kantoor. Onze gezamenlijke kopjes koffie volgens een vaste volgorde van handelingen vormden altijd een goede start van de dag. Je had altijd een luisterend oor voor me en als het nodig was hielp je me met de nodige praktische hulp uit de brand. Je kunt beter een goede kantoorgenoot hebben dan een verre vriend, en jij behoort zeker tot de eerste categorie. Thijs, officieel was jij geen kantoorgenoot van Ferdy en mij. Je kwam standaard twee minuten te laat om ook met ons koffie te gaan halen. Toch hoorde je er helemaal bij. Bedankt dat je je niet liet afschrikken door mijn eeuwige opmerkingen over ruitjesbloezen en dat ik je altijd om taaladvies mocht vragen. Dank jullie wel dat jullie mijn paranimfen willen zijn.

Jan Willem, mijn andere kantoorgenoot. Het was fijn dat jouw promotieonderzoek aan het mijne raakt. Dat maakte dat ik met jou mijn onderzoeksinteresses kon delen en kon discussiëren over (tweetalig) onderwijs.

Een speciaal woord van dank gaat uit naar mijn mede-promovendi. Aurora, thank you for your enthusiasm, making hairdresser appointments more fun, and for endless text messages about useful and useless things. Chantal en Elly, jullie begrijpen als geen ander wat het is om heel het land door te crossen voor één participant. Het is fijn dat we onze ervaringen – de leuke momenten en de frustraties – met elkaar konden delen. Chantal, bedankt ook dat je me af en toe aan het sporten kreeg. Theresa, jou wil ik bedanken voor je oprechte aandacht, goede gesprekken en gezellige momenten, op de universiteit en daarbuiten. Emily, Katherine, Chen, Lotte, Marjoke (ook al ben je geen promovenda), Wessel (oké, ook geen promovendus), Martijn, Mario, Chara, Polina, Sara, Remy, thank you for coffee breaks, (fancy) dinners, and all kinds of non-work related activities. Dirk, dank je wel voor het organiseren van ons fantastische PhD (blue coat) weekend in Leuven. It all made PhD life a wonderful experience.

Dank ook aan mijn collega's Frans en Henk. Frans, jou wil ik bedanken voor je geduldige uitleg van Tiaplus. Henk, bedankt dat we onze verjaardagen samen konden vieren. Vlaai bestellen bij de Refter was altijd een uitdaging, en die ging ik liever aan mét dan zonder jou. Ook bedankt dat ik je vragen mocht stellen over van alles en nog wat en dat je daar bijna altijd het antwoord op wist.

Sommige mensen hadden een meer indirecte invloed op mij of mijn proefschrift, maar zijn toch van waarde geweest. Aan de bijeenkomsten van de PI-groep Cognitive and Developmental Aspects of Multilingualism heb ik veel gehad. Het was fijn om te kunnen praten en na te denken over de vele aspecten van tweetaligheid. De verschillende visies vormden waardevolle input voor mijn onderzoek. Ton, jou wil ik speciaal bedanken voor je oprechte aandacht en je

waardevolle commentaar op mijn presentaties. The Sound Learning Meetings were also very useful to me. I presented my work multiple times and got valuable feedback. I learned a lot.

Vanessa, jij gaf me een inkijkje in de wereld van het voortgezet tweetalig onderwijs door me te betrekken bij het onderzoek op de school waar je werkte. Het was interessant om ook na te denken over het voortgezet onderwijs en die ervaringen te vergelijken met die in het basisonderwijs.

Mienke, jij was één van mijn lijntjes met Pedagogische Wetenschappen & Onderwijskunde. Het was fijn om met jou samen studenten te begeleiden en me nog steeds verbonden te voelen met de onderwijskunde.

Joep, mijn allereerste onderzoeksproject(je) voerde ik jaren geleden uit met jou als mijn begeleider. Het wakkerde mijn interesse in het doen van onderzoek echt aan. Ik ben blij dat we de afgelopen jaren contact gehouden hebben. Bedankt voor alle gesprekken over onderwijs en over goede boeken, het gaf me altijd stof tot nadenken.

Annegien, wie had in ons eerste jaar van de bachelor PWO durven denken dat onze onderzoeksinteresses uiteindelijk zo dicht bij elkaar zouden komen te liggen? Met jou kan ik altijd praten over tweetaligheid en onderwijs, en over zoveel andere dingen. Dat Groningen en Nijmegen zo ver uit elkaar liggen hoeft daarbij helemaal geen probleem te zijn, want dan ga je gewoon samen naar een conferentie in Barcelona.

Dan waren er nog de mensen die weliswaar niet direct iets bijgedragen hebben aan mijn onderzoek, maar die toch van onschatbare waarde zijn geweest. De familie Nies, Anne en Dirk, Melissa en Giel, Pim en Marinde, Martijn, Lieke, Janne, Marloes, Sigrid, bedankt dat jullie met mij meejuichten bij elke mijlpaal. En Wietske, ik denk dat jij altijd nog harder juichte dan ikzelf.

Oma, van jou heb ik geleerd dat je met wilskracht en doorzettingsvermogen een heleboel kunt bereiken, én dat je nooit te oud bent om te leren. Ik heb er in de afgelopen jaren veel aan gehad.

Pépé, malgré la distance entre nous je n'ai jamais l'impression que t'es loin. Merci pour les coups de téléphone et les journées à Sceaux. C'est toujours un plaisir de se parler et de se voir.

Mijn ouders, ik ben jullie dankbaar dat jullie me stimuleren mijn eigen pad te volgen en dat jullie me altijd proberen te helpen. Mam, bedankt dat je met me meedacht over het werven van participanten en dat je daarvoor al je contacten inzette. Papa, je suis contente que ma recherche nous a fait réaliser qu'il y a différentes manières d'être bilingue, et que maintenant on s'écrit (la plupart du temps) en français mais on se parle (aussi la plupart du temps) en hollandais.

Chris, mijn broer, jij snapt als geen ander hoe leuk het is om twee talen te spreken. Met jou kan ik daar de beste grappen over maken (én klagen over alle gekke vragen die we krijgen). Ik sluit niet uit dat een deel van mijn passie voor tweetaligheidsonderzoek daar begonnen is.

En ten slotte Tim, je bent altijd de eerste om te geloven in al mijn dromen, ideeën en plannen en helpt me ze te realiseren. Ik ben blij dat ik jou aan mijn zijde heb en het betekent meer voor me dan ik hier verwoorden kan. Bedankt, voor alles.

Curriculum Vitae

Claire Goriot (Boxmeer, 1990) started a bachelor in Pedagogical and Educational Sciences in 2009 at Radboud University. During her studies, she also participated in the honours programme of Pedagogical and Educational Sciences, and spent a semester at Roehampton University in Londen. In 2012, she completed her bachelors with a specialization in Educational Sciences (cum laude), and started the Research Master Behaviour Science, also at Radboud University. During her masters she worked as a student assistant. She graduated from this master's programme in 2014, and went on to do a PhD at the Centre for Language Studies (CLS), which was made possible by an open competition grant from the CLS. The results are reported in this dissertation. Claire is currently employed at department of Educational Sciences at the Vrije Universiteit in Amsterdam, where she works as a lecturer and researcher.

List of publications

Goriot, C., McQueen, J. M., Unsworth, S., van Hout, R., & Broersma, M. (2018). Perception of English phonetic contrasts by Dutch children: How bilingual are early-English pupils? Manuscript submitted for publication.

Goriot, C., Unsworth, S., van Hout, R., Broersma, M., McQueen, J. M. (2018). Differences in phonological awareness development: Are there positive or negative effects of bilingual experience? Manuscript submitted for publication

Goriot, C., van Hout, R., Broersma, M., Lobo, V., McQueen, J. M., & Unsworth, S. (2018). Using the Peabody Picture Vocabulary Test in L2 children and adolescents: Effects of L1. *International Journal of Bilingual Education and Bilingualism, Advance online publication*. doi: 10.1080/13670050.2018.1494131

Goriot, C., Broersma, M., McQueen, J. M., Unsworth, S., & van Hout, R. (2018). Language balance and switching ability in children acquiring English as a second language. *Journal of Experimental Child Psychology, 173,* 168-186. doi: 10.1016/j.jecp.2018.03.019.

Goriot, C., Bakker, J., Denessen, E., & Droop, M. (2016). Zijn de voordelen van tweetaligheid voor alle kinderen even groot? Een exploratief onderzoek naar de leerkrachtwaardering van de thuistaal van leerlingen en de invloed daarvan op de ontwikkeling van hun executieve functies. *Pedagogiek, 36*(2), 135-153, doi: 10.5117/PED2016.2.GORI

Goriot, C., Denessen, E., Bakker, J., & Droop, M. (2016). Benefits of being bilingual? The relationship between pupils' perceptions of teachers' appreciation of their home language and executive functioning. *International Journal of Bilingualism, 20*(6), 700-713, doi: 10.1177/1367006915586470