

**Immersion education in the Netherlands: The relation
between individual characteristics of bilingual primary
school students and their English proficiency.**

Jason Gosling

4159063

Meertaligheid en Taalverwerving

Master's thesis

January 7th, 2019

Utrecht University

Supervisor: René Kager

Second reader: Elena Tribushinina



Faculty of Humanities

Version September 2014

PLAGIARISM RULES AWARENESS STATEMENT

Fraud and Plagiarism

Scientific integrity is the foundation of academic life. Utrecht University considers any form of scientific deception to be an extremely serious infraction. Utrecht University therefore expects every student to be aware of, and to abide by, the norms and values regarding scientific integrity.

The most important forms of deception that affect this integrity are fraud and plagiarism. Plagiarism is the copying of another person's work without proper acknowledgement, and it is a form of fraud. The following is a detailed explanation of what is considered to be fraud and plagiarism, with a few concrete examples. Please note that this is not a comprehensive list!

If fraud or plagiarism is detected, the study programme's Examination Committee may decide to impose sanctions. The most serious sanction that the committee can impose is to submit a request to the Executive Board of the University to expel the student from the study programme.

Plagiarism

Plagiarism is the copying of another person's documents, ideas or lines of thought and presenting it as one's own work. You must always accurately indicate from whom you obtained ideas and insights, and you must constantly be aware of the difference between citing, paraphrasing and plagiarising. Students and staff must be very careful in citing sources; this concerns not only printed sources, but also information obtained from the Internet.

The following issues will always be considered to be plagiarism:

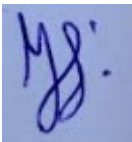
- cutting and pasting text from digital sources, such as an encyclopaedia or digital periodicals, without quotation marks and footnotes;
- cutting and pasting text from the Internet without quotation marks and footnotes;
- copying printed materials, such as books, magazines or encyclopaedias, without quotation marks or footnotes;
- including a translation of one of the sources named above without quotation marks or footnotes;
- paraphrasing (parts of) the texts listed above without proper references: paraphrasing must be marked as such, by expressly mentioning the original author in the text or in a footnote, so that you do not give the impression that it is your own idea;
- copying sound, video or test materials from others without references, and presenting it as one's own work;
- submitting work done previously by the student without reference to the original paper, and presenting it as original work done in the context of the course, without the express permission of the course lecturer;
- copying the work of another student and presenting it as one's own work. If this is done with the consent of the other student, then he or she is also complicit in the plagiarism;
- when one of the authors of a group paper commits plagiarism, then the other co-authors are also complicit in plagiarism if they could or should have known that the person was committing plagiarism;
- submitting papers acquired from a commercial institution, such as an Internet site with summaries or papers, that were written by another person, whether or not that other person received payment for the work.

The rules for plagiarism also apply to rough drafts of papers or (parts of) theses sent to a lecturer for feedback, to the extent that submitting rough drafts for feedback is mentioned in the course handbook or the thesis regulations. The Education and Examination Regulations (Article 5.15) describe the formal procedure in case of suspicion of fraud and/or plagiarism, and the sanctions that can be imposed.

Ignorance of these rules is not an excuse. Each individual is responsible for their own behaviour. Utrecht University assumes that each student or staff member knows what fraud and plagiarism entail. For its part, Utrecht University works to ensure

Submit this form to your supervisor when you begin writing your Bachelor's final paper or your Master's thesis.

Failure to submit or sign this form does not mean that no sanctions can be imposed if it appears that plagiarism has been committed in the paper.

I hereby declare that I have read and understood the above.	
Name: Jason Gosling	
Student number: 4159063	
Date and signature: 07-01-2019	

Abstract

The purpose of this study is two-fold: to examine the relation between the age of acquisition, native language and amount of classroom L2 exposure of immersion students and their English proficiency and to describe the future programmes of two bilingual primary (immersion) schools in the Netherlands. Immersion education is still at an early stage in the Netherlands and it is important to study the influence of these characteristics on English proficiency. Part one of this thesis makes use of a quantitative experiment whilst part two is achieved by interviews with the schools' programme coordinators.

For part one, 97 students from two bilingual primary schools were subject to tasks measuring their receptive vocabulary, productive vocabulary and receptive grammar. The results show that students with different ages of acquisition did not differ in their English proficiency whereas students with different native languages and amounts of classroom L2 exposure did differ. Students with English as their native language and students with a higher amount of classroom L2 exposure scored significantly higher on respectively, receptive grammar and productive vocabulary, and receptive vocabulary and grammar.

Part two of this thesis is of a descriptive character and focusses on variables in the classroom. The programmes of both schools will stay relatively similar to the previous year as large changes are deemed risky due to the early stage immersion education is in. School 1 will undergo several changes due to the large influx of Dutch students while school 2 will continue the same path as before since the students' progress is going well.

Acknowledgements

First and foremost, I would like to thank my supervisor dr. René Kager for his support, extensive feedback and most of all his ideas. Without his help, this thesis would not have seen the light of day. Second, I extend my gratitude to my second reader, Elena Tribushinina for her willingness to cooperate and her critical reading. Finally, my appreciation goes out to the two schools, their team of coordinators and teachers, and of course the young pupils. Visiting the schools was insightful to say the least and, apart from providing me with a dataset, taught me much about how a classroom looks like.

Table of Contents

ABSTRACT	IV
ACKNOWLEDGEMENTS	V
TABLE OF CONTENTS	VI
LIST OF TABLES.....	VIII
1: INTRODUCTION.....	1
1.1: POSITION OF ENGLISH IN THE NETHERLANDS	2
1.2: HISTORY OF ENGLISH TEACHING IN THE NETHERLANDS	2
1.2.1: <i>eibo</i>	2
1.2.2: <i>wto</i>	3
1.2.3: <i>tto</i>	3
1.3: DEFINITION OF IMMERSION.....	4
1.4: TPO-PILOT.....	5
1.5: RESEARCH QUESTIONS	6
2: THEORETICAL FRAMEWORK.....	8
2.1: MEASURES OF L2 PROFICIENCY.....	8
2.2: INDIVIDUAL DIFFERENCES IN IMMERSION	8
2.2.1: <i>AoA</i>	9
2.2.2: <i>Native Language</i>	10
2.2.3: <i>Amount of Classroom L2 Exposure</i>	11
2.3: HYPOTHESES	13
3: METHOD.....	14
3.1: PARTICIPANTS	14
3.2: MATERIALS	16
3.3: PROCEDURE	17
3.4: DESIGN AND ANALYSIS	18
4: RESULTS.....	19
4.1: DESCRIPTIVES.....	19
4.1.1: <i>English Proficiency Scores based on Grade</i>	19
4.1.2: <i>English Proficiency Scores based on AoA</i>	20
4.1.3: <i>English Proficiency Scores based on NL</i>	21
4.1.4: <i>English Proficiency Scores based on EXP</i>	21
4.1.5: <i>Comparison with FoTo-Study</i>	22
4.2: TESTING OF HYPOTHESES	23
4.2.1: <i>Hypothesis 1</i>	23
4.2.1.1: Receptive Vocabulary.....	24
4.2.1.2: Receptive Grammar	25
4.2.1.3: Productive Vocabulary.....	25
4.2.2: <i>Hypothesis 2</i>	26
4.2.2.1: Receptive Vocabulary.....	26
4.2.2.2: Receptive Grammar	27
4.2.2.3: Productive Vocabulary.....	27
4.2.3: <i>Hypothesis 3</i>	27
4.2.3.1: Receptive Vocabulary.....	28
4.2.3.2: Receptive Grammar	28
4.2.3.3: Productive Vocabulary.....	28

4.2.4: <i>Other variables</i>	28
5: DISCUSSION	29
5.1: RQ1A.....	29
5.2: RQ1B.....	29
5.3: CONNECTION TO THE LITERATURE.....	31
5.3.1: <i>AoA</i>	31
5.3.2: <i>NL</i>	31
5.3.3: <i>EXP</i>	32
5.4: IMPLICATIONS.....	33
5.5: LIMITATIONS	33
5.6: FURTHER AREAS OF RESEARCH.....	34
6: CONCLUSION	35
7: INTRODUCTION	36
8: THEORETICAL FRAMEWORK	37
8.1: QUALITY OF EXPOSURE	37
8.2: USE OF STUDENTS' L1s	38
8.3: TEACHER L2 PROFICIENCY	38
8.4: LANGUAGE ALLOCATION	39
9: METHOD	41
9.1: PARTICIPANTS	41
9.2: MATERIALS	41
9.3: PROCEDURE	41
9.4: DATA ANALYSIS	41
10: RESULTS	43
10.1: NEW STUDENTS AND ADMISSION POLICY	43
10.2: STUDENTS' NL	43
10.3: TEACHERS' L2 PROFICIENCY AND LANGUAGE ALLOCATION	44
10.4: EXP.....	44
11: DISCUSSION	46
11.1: SCHOOL 1	46
11.2: SCHOOL 2	47
12: CONCLUSION	49
13: GENERAL CONCLUSION	50
REFERENCES	51
APPENDIX A	55
APPENDIX B	56
APPENDIX C	67

List of Tables

Table 1.1	14
Table 1.2	14
Table 1.3	15
Table 1.4	15
Table 1.5	16
Table 1.6	17
Table 2.1	19
Table 2.2	20
Table 2.3	21
Table 2.4	22
Table 3.1	24
Table 3.2	25
Table 3.3	26

1: Introduction

In 2014, a five-year pilot on immersion education in primary schools was implemented in the Netherlands. This pilot attempts to see whether immersion, or ‘tweetalig primair onderwijs’ (tpo), can be effective in the Netherlands. It has the complementary goal of developing a high-quality bilingual (Dutch/English) programme for Dutch primary schools. Immersion is a type of bilingual programme with a large focus on language learning via offering content in the target language (TL), in this case English. As such, it is believed to be very beneficial for the pupils’ second language (L2) development. Additionally, by learning another language it is expected the pupils will learn more about the culture associated with that language. A more precise definition of immersion programmes and how they can differ from each other will be given in section 1.3. Two primary schools have also started their own immersion programmes, though separate from the pilot. For my internship, I came into contact with these schools and had the opportunity to learn about their programmes. Working with these schools helped form the goal of this thesis: to study the relation between individual characteristics of immersion students and their English proficiency. Additionally, based (in part) on the results of this first part, the future programmes of the two schools will be examined in terms of policy and curriculum. The first part (chapters 1-6) will consist of measuring the students’ English proficiency skills with an additional focus on the (individual) differences between students and how these manifest in their English development. The second part (chapters 7-12) will be accomplished by means of interviews with the programme coordinators. Thus, there will be a quantitative (part one) and qualitative component (part two). Each component will have its own theoretical framework, methodology, results, discussion and conclusion. Before delving into the research questions (RQs), the introduction discusses several phenomena that help give a better view of immersion. These include the position and history of English (teaching) in the Netherlands, the definition of immersion and information on the pilot. Defining immersion is especially important since immersion programmes can have different goals and set-ups for achieving those goals. The introduction ends with the thesis’s RQs. Following the introduction come the theoretical framework, methodology, results, discussion and conclusion of first, the quantitative component and second, the qualitative component. Afterwards, the findings of both components will be summarised in the general conclusion.

1.1: Position of English in the Netherlands

English “is very present in the media and in the linguistic landscape” and is even becoming an L2 for the Dutch (De Bot, 2014, p. 417). English has seeped into virtually every domain, be it business, education or media. Others have also noted English is switching from a foreign language (FL) to an L2 in the Netherlands (Ammon & McConnell 2002, as cited in Edwards, 2016). Defining what is and is not an L2 or an FL is an entire debate onto itself, but for the sake of clarity the following description of English in the Netherlands will be used. Since English is chronologically the second language most Dutch people learn in schools, English will be termed as the L2 in this thesis. It is believed that speaking English is highly beneficial for “students’ future work and lives within Dutch society” and is used to display “status and prestige” (Edwards, 2016, p. 66). The importance of English is reflected in its prominent role in Dutch education. Ammon and McConnell (2002) highlight the Netherlands as “one of the most advanced countries in Europe concerning the integration of instruction of English in the national education system” (as cited in Edwards, 2014, p. 38).

1.2: History of English Teaching in the Netherlands

A brief history of English teaching and its different forms is given to provide a more detailed picture of the situation of English teaching in the Netherlands. This will additionally offer perspective on how immersion education might be viewed by the Dutch public (which in part can help with the pilot’s success).

1.2.1: eibo. English is taught at every level of Dutch education: primary, secondary and higher education. It has been implemented most recently at the primary school level. English in primary schools has been compulsory in the Netherlands since 1986 (Jenniskens et al., 2017). This type of English teaching was called ‘engels in het basisonderwijs’ (eibo) or English in primary education and involved the mandatory teaching of English in grades seven and eight. The goal was to enable the students to effectively communicate in English and not to be “linguistically correct” per se (Engel, Trimbos, Drew & Groot-Wilken 2007, as cited in Thijs, Trimbos, Tuin, Bodde & de Graaff, 2011, p. 8). English was made compulsory for a variety of reasons: the rise of English as an essential international language, European policy on modern early Foreign Language Learning (FLL) and the need to prepare pupils for English in secondary school (Thijs et al., 2011).

1.2.2: vvto. The demand for learning English at an increasingly younger age has resulted in the emergence of ‘vroeg vreemdetalenonderwijs’ (vvto) or early FLL. Vvto involves giving English lessons as early as the first grade. Given vvto pupils are often only four years old, they are typically familiarised with English in the form of games or singing. Four-year-olds are assumed to be less suitable for explicit instruction in the L2 compared to older children, hence the different approach compared to pupils in grades seven or eight (DeKeyser 2003, as cited in Graham, Courtney, Marinis & Tonkyn, 2017). Apart from the earlier starting age, vvto can be also seen as the next step of eibo in terms of L2 allocation. In 2013, the Dutch ministry of education, culture and science (OCW) allowed primary schools to instruct their pupils in another language (English, German or French) for a maximum of 15% of the instruction time. This limit was set due to the fear of pupils’ first language (L1) development slowing down due to spending too much time on English (Jenniskens et al., 2017). The study by Jenniskens et al. (2017) showed that this fear was uncalled for: allocating 15% of instructional time to English does not have a negative effect on the development of the L1 (Dutch).

1.2.3: tto. English in secondary schools has been taught both as a subject and in the form of a bilingual programme called ‘tweetalig onderwijs’ (tto) or bilingual education. Tto entails bilingual programmes in secondary schools and exists at each of the three levels of secondary education: VMBO, HAVO and VWO. The goal of tto is to improve students’ chances of finding success in higher education and on the job market. Tto involves explicit L2 instruction as well as teaching *in* the L2. This means content is learned through the L2, which is called Content and Language Integrated Learning (CLIL). CLIL is at the heart of tto curricula and is defined as:

a dual-focused educational approach in which an additional language is used for the learning and teaching of both content and language. That is, in the teaching and learning process, there is a focus not only on content, and not only on language. Each is interwoven, even if the emphasis is greater on one or the other at a given time. (Coyle, Hood & Marsh, 2010, p. 1)

Courses such as geography, history and physical education are often taught in the L2. Besides lessons in and on the L2, tto also aims to provide students an international perspective by imbedding the L2 and materials in an international context (Nuffic, 2017). Research has shown that, like vvto, tto has “no detrimental effects for Dutch”, nor for subjects taught in the L2 (Admiraal et al 2006; Huibregtse 2001, as cited in Edwards, 2014, p. 43). The success of tto “led to demands by both parents and schools to introduce some form of bilingual education in primary education as well” (De Bot, 2014, p. 411-12).

1.3: Definition of Immersion

Whilst vvto was occurring in the Netherlands, a new approach to bilingual education called immersion was being developed abroad. Immersion started and has been popularised in Canada because of its English- and French speaking population. Immersion programmes were developed to provide “Canada’s majority-group English-speaking students with opportunities to learn Canada’s other official language” (French) (Genesee, 1994, p. 1). Immersion has since been implemented in many countries across different continents. It has gained huge traction in the United States because of its large Spanish speaking population who wish to retain their (minority) L1. In other countries, many of which in Europe, the demand comes from the recognition of English as an important international language (Nikolov 2010, as cited in Hickey & de Mejía, 2017). The same demand applies for the Netherlands.

As stated, immersion is a type of bilingual education and can exist in many forms. In some countries one type of immersion can be more suited to its population than another type. The focus here will be on the type of immersion that is expected to be implemented in the Netherlands. Eight core features of immersion programmes are given by Johnson and Swain (1997):

1. The L2 is a medium of instruction
2. The immersion curriculum parallels the local L1 curriculum
3. Overt support exists for the L1
4. The program aims for additive bilingualism
5. Exposure to the L2 is largely confined to the classroom
6. Students enter with similar (and limited) levels of L2 proficiency

7. The teachers are bilingual
8. The classroom culture is that of the local L1 community (p. 6-8)

These features are not applicable to all types of immersion as some programmes aim to strengthen a minority or heritage language. As this is not the case with the tpo-pilot or the two schools, the features mentioned above can be applied (with the exception of features five and six). There is a wide variety of potential programmes depending on the country, population, language policy etc. An important distinction is made between early and late immersion and between partial and total immersion. Early immersion starts in the first grade while late immersion starts in grade six or seven (Johnson & Swain, 1997). Initially, as with vvto, there were worries that pupils' L1 development would suffer because of the amount of time allocated to the L2. This led to the emergence of partial immersion which entails 30-50% L2 instructional time over the course of primary school. Total immersion, on the other hand, starts with 100% L2 instructional time and finishes primary school with about 50% (Baker, 2011).

An immersion programme's success is believed to depend on, among others, individual factors such as socio-economic status (SES), motivation, L1, gender and attitudes. In this thesis, the *age of acquisition (AoA)*, *native language (NL)* and *amount of classroom L2 exposure (EXP)* and their relation to students' English proficiency will be examined. The rationale behind choosing these factors will be explained in section 2.2.

1.4: Tpo-pilot

Due to tpo being in such an early stage in the Netherlands, there is no clear framework for how the ideal immersion programme should be set up. There are many examples of immersion programmes in other countries, but one cannot copy an education programme and expect it to work in another country. This inability to copy is due to possibly having a different TL, the role the TL has in that country and the typological distance between the NL and TL. Baker (2011) highlights the complexity of (successful) immersion programmes:

The language policy and language practice in schooling are only one element amongst many that make a school more or less successful. A recipe for success is unlikely to result

from one ingredient (e.g. the language of the classroom). A great variety of factors act and interact to determine whether bilingual education is successful or not. (p. 202)

In order to develop the tpo-pilot, a report was made by van den Broek, de Graaff, Unsworth and van der Zee (2014). This report contained a study of international literature and interviews with immersion experts. Its aim was to provide information on the set-up, execution and evaluation of the pilot. The authors concluded that immersion programmes do not have a negative effect on the L1 and enhance the reading, listening, speaking and writing skills of the students' L2 (van den Broek et al., 2014). They also mentioned the schools in the pilot show some very important similarities with the schools examined in the literature (for example, 30-50% instruction in the L2) and is therefore expected to show (relatively) similar results in terms of L1/L2 proficiency. The FoTo-study by Jenniskens et al. (2018) details the progress of the pilot's schools and their students. More details on the results of the FoTo-study will be given in chapter 2. The results will be used to compare the students of the pilot schools with the participants of this study as it can be useful for the qualitative part of this thesis. Seeing how these two groups of students compare can provide insight into the appropriate future path of the programmes of the two schools. The FoTo-study will serve as a benchmark to see whether the participants possibly underperform on a certain test (e.g. productive vocabulary). If the comparison with the benchmark shows that the participants underperform compared to the pilot's students, the schools can adjust their programme to improve that particular domain.

1.5: Research Questions

To conclude the introduction, the goal of this thesis is two-part: to study the relation between individual characteristics of immersion students and their English proficiency and, based (in part) on these results, to examine the future programmes of the two schools in terms of policy and curriculum. This goal is relevant because as of yet there is not much information on this topic in the context of the Netherlands. It will add to the existing body of research that is mainly focused on the tpo-pilot since the participants are from different schools than the pilot schools. The English proficiency will be indicated by measures of *receptive vocabulary (RV)*, *receptive grammar (RG)* and *productive vocabulary (PV)*. Additional attention will be given to differences between Dutch natives, English natives and students with a different NL than Dutch or English (internationals).

This thesis's goal will be achieved by answering the following RQs. RQs 1a and 1b focus on the quantitative part and RQ 2 on the qualitative part.

1.
 - a. Is there a correlation between, on the one hand, the students' *AoA*, *NL*, *EXP* and, on the other hand, the students' English proficiency?
 - b. Do students with different *AoAs*, *NLs* and *EXPs* significantly differ in their English proficiency?

2. What are the schools' plans regarding the influx of new students the coming year(s) in terms of policy and curriculum design and how are they going to accommodate the Dutch, English and international students?

2: Theoretical Framework

2.1: Measures of L2 Proficiency

The indicators of L2 proficiency will be based on indicators used by others in the field of immersion. L2 proficiency in research settings is often divided into several subcomponents and can be measured in a variety of ways. These subcomponents are writing, speaking, listening and reading. The former two are productive skills, *i.e.* skills involving the person producing language either via speech or writing. The latter two are receptive skills, *i.e.* skills that are comprehension-based and do not involve the producing, but the understanding of language (listening and reading). The indicators of proficiency used are *receptive vocabulary*, *receptive grammar* and *the productive vocabulary*. These are indicators used before in other studies such as the FoTo-study. This makes it possible to compare the results from the participants with the students in the FoTo-study. The tests for the three areas of L2 proficiency are: the PPVT-4 for RV, the TROG-2 for RG and the EVT-2 for PV. The use of vocabulary and grammar as indicators of L2 proficiency is self-explanatory; knowledge of the meaning/function of words and knowledge on the construction of sentences are necessary for learning a language. Both receptive and productive vocabulary are taken as measures because both are important abilities and they often do not develop at the same pace. Immersion students' expressive skills (writing and speaking) in the L2 seem to lag behind native speakers (Genesee 1987; Harley, Allen, Cummins & Swain 1990, as cited in Hummel, 2014). This is not the case for receptive skills (listening and reading). Although they both deal with vocabulary, they are distinct skills and can thus paint two sides of a picture. Additionally, a non-verbal intelligence test (SON-R) will be taken into account to see whether non-verbal intelligence (IQ) correlates with the students' English proficiency. More detail on the tests are given in section 3.2.

2.2: Individual differences in Immersion

Individual factors in language learning have long been of interest for researchers. Not only can these factors provide insight into theories of language acquisition (e.g. Universal Grammar vs. Usage-Based Theory), they can additionally be used to positively influence L2 development (Sun, Steinkrauss, Tendeiro & de Bot, 2016). Especially the latter function is relevant for immersion programmes in the Netherlands because of the early stage it is in. There are many factors on the individual level (e.g. age, attitudes etc.) that should be taken into account when creating a language

learning programme. To rank these factors in order of importance for L2 development is difficult and not the goal of this thesis. As stated, several variables which are believed to be important indicators of L2 proficiency in immersion programmes will be examined to see how they relate to the students' English proficiency. These variables are *AoA*, *NL* and *EXP*. The reason for choosing these variables is that they are believed to be important for L2 development in immersion settings (Genesee 2004; Johnson & Swain 1997; Pfenninger, 2014).

AoA is the age at which the child is subject to significant or substantial exposure to the L2 (Hummel, 2014). This exposure can be obtained via the child's environment (naturalistic setting) or via instruction in school (instructional setting). Although it can be difficult to pinpoint, *AoA* is important because an earlier *AoA* could mean the child will have had more exposure to the TL. There can thus be some overlap between *AoA* and *EXP* which will be elaborated on in section 2.2.1. *AoA* is also important as it can be translated to the implementation of an early or late immersion programme. The second variable, the child's *NL*, needs no elaborate explanation as to why to it is an important indicator of proficiency in the TL. The *NL* is important to keep in mind though in case there are no or small differences between native and non-native English-speaking pupils. This experiment has a large number of internationals which makes it more interesting to include *NL*. *EXP* is also an important factor for L2 proficiency in immersion programmes (Housen 2012; Johnstone 2002; Lightbown 2000; Muñoz 2012, as cited in Buyl & Housen, 2014). It is especially important pedagogically because schools have the option to alter this variable to best fit their programmes and goals. More detailed information on these three variables will be given in sections 2.2.1, 2.2.2 and 2.2.3.

2.2.1: AoA. As stated, *AoA* is the age at which the child is subject to significant or substantial exposure to the L2 (Hummel, 2014). Quantifying the *AoA* can be achieved in many ways. In this thesis, the significant L2 exposure can either be gained by having parents who speak the L2, having lived in the L2 environment or by following an immersion programme in the L2. To clarify, a student's *AoA* can be zero if the student is born in an English-speaking country or if one of their parents is a native L2 speaker. It can also be any other age (though immersion rarely starts before age four) depending on at what age the student starts with an immersion programme. *AoA* can thus be identified by naturalistic or instructional L2 exposure.

The effect AoA can have on immersion students' L2 proficiency is not straightforward. There are studies showing both early and late immersion programmes can positively influence L2 proficiency (Pfenninger, 2014; Pfenninger & Singleton, 2017). Both types of programme can work if accompanied by “appropriate and continuous instruction” (Genesee, 2004, p. 27). A potential explanation for an ultimate attainment advantage of an early AoA is given by Muñoz (2008). She argues that an early AoA “produces long-term benefits when associated with greater time and massive exposure” (Muñoz, 2008, p. 582). This suggests that AoA on its own might not have a large effect on L2 proficiency if it does not translate into a higher amount of L2 exposure. Genesee (2004) offers additional reasons for the successes found with an early AoA. These include the “optimal fit between learning styles of young learners and effective L2 pedagogy” and the idea that an early AoA facilitates “students’ innate or natural language learning ability” (Genesee, 2004, p. 12). A late AoA might be advantageous because older students are reported to have an initial rate advantage over younger students (due to enhanced cognitive functions) (Genesee, 1987; Lindholm-Leary & Genesee 2014; Muñoz 2015; Swain 1996, as cited in Lambelet & Berthele, 2015). AoA has been shown to influence linguistic domains differently in immersion. An early AoA can be beneficial for listening comprehension and oral communication while a late AoA can be better for “reading and writing skills, and lexical and grammatical knowledge” in total immersion (Lapkin, Swain, Kamin & Hanna 1980; Muñoz 2008, as cited in Pfenninger, 2014, p. 533). Since the participants of this experiment follow early partial immersion programmes (although the participants’ average EXP lies below the average EXP for partial immersion, which is 30-50%), it would be interesting to see if similar results are found.

2.2.2: Native Language. NL is an important variable in any immersion programme for the obvious reason that students whose NL is the TL are expected to outperform those with a different NL. It is an important factor to include though for the following reason. The results might show there is no or small difference on a particular area of proficiency between Dutch natives, English natives and internationals. This could have implications on the pedagogical or policy level for a particular programme. If there is no or small difference between students with different NLs, it could mean the programme is set up correctly for majority-language students. At the same time, it could imply the English natives are progressing at a far slower rate since they had a head-start

compared to other students. One reason there might be an insignificant or small difference is the typological distance between Dutch and English. The “typological proximity can considerably affect the rate and levels of attainment of L2 learning” (Buyl & Housen, 2014, p. 192). Positive transfer might thus play a role for Dutch students in particular.

The FoTo-study by Jenniskens et al. (2018) looked at students’ home language and its relation to RV, RG and PV. The home language was the language the student hears most at home. Although home language differs from the student’s NL (the variable used in this experiment), some overlap can be expected since a child’s NL is often the parents’ NL. Students with English as a home language scored significantly higher on the three tests than students without English as their home language. Interestingly, students with a different home language than Dutch or English scored significantly higher on RG than the Dutch home language students (though not on RV or PV) (Jenniskens et al., 2018). This finding is surprising since migrant children often underperform compared to the majority language students (Steinlen, Håkansson, Housen & Schelletter, 2010). The authors of the FoTo-study did not mention how this result came into being.

2.2.3: Amount of Classroom L2 Exposure. EXP is determined by measuring the amount of L2 instruction per week compared to the total amount of instruction time per week. As stated, some overlap is expected between AoA and EXP because an earlier AoA probably means the student will have had more exposure to the TL. A correlation check between these variables will be conducted to see if there is overlap, though in this experiment it is not expected. This is because sufficient difference is assumed between the two variables for the following reasons. While students in the first grade have the same EXP and AoA, this is not the case for the third grade students. In terms of AoA, the students from one school started one year earlier than the students from the other school. The students who started earlier also have a higher EXP. The difference between the two variables thus lies in the third grade students.

The EXP variable in immersion research is not often taken as a single variable, but as one of many variables that together make up the total amount of L2 input the child receives. This makes it difficult to accurately predict EXP’s influence on L2 proficiency based on previous research. Nevertheless, EXP is generally believed to be related to the students’ L2 proficiency. Lyster and

Genesee (2013) concur and note that in general more L2 exposure leads to higher levels of immersion students' L2 proficiency. However, Genesee and Lindholm-Leary (2013) mention the relation between exposure and L2 outcomes is complex. This complexity is backed by contradicting studies. They refer to the finding that “students in total IMM programmes generally acquire higher levels of L2 proficiency than students in partial IMM programmes” (Genesee 2004, as cited in Genesee & Lindholm-Leary, 2013, p. 12). In another study though, late-immersion students sometimes perform equal to early-immersion students despite the disparity in L2 exposure (Genesee 1981, as cited in Genesee & Lindholm-Leary, 2013, p. 12). Others add to this by mentioning several other studies that “have also suggested that L2 proficiency does not necessarily increase with longer classroom contact with the L2” (Steinlen et al., 2010, p. 70).

EXP is especially important considering the claim that young learners learn language (in part) implicitly. Implicit learning “is a slow process requiring plentiful and high-quality input”, which emphasises the importance of amount and quality of L2 exposure (Graham et al., 2017, p. 924). The role quality of input has on L2 proficiency will be elaborated on in section 8.1. There is also the question whether amount of input (in or outside the classroom) influences different linguistic domains such as vocabulary or grammar in the same manner. Unsworth (2013) notes that “despite the general agreement on the role of the amount of input in second or foreign language acquisition”, it remains difficult to predict the influence of amount of input on the different domains of language (as cited in Sun et al., 2016, p. 560).

Jenniskens et al. (2018) showed that EXP positively influences the students' RV and RG. There is a discrepancy between EXP in the FoTo-study schools and the schools in this experiment though. Whereas in this experiment the EXP lies between 7% and 19% of the total instruction time per week, the average EXP in the FoTo-study is 36% for the first grade and 24% for the third grade. It will be interesting to see whether a lower EXP (compared to the schools in the pilot) will have a similar effect as EXP had on RV and RG in the FoTo-study. EXP was not included as a variable in their analysis for PV, hence the absence of a similar finding for PV. The FoTo-study findings are particularly relevant for this thesis as the students in the pilot more closely resemble the participants of this study than those in the aforementioned studies. The study contained no mention

of EXP influencing vocabulary development in a different manner than grammatical development, such as with NL.

2.3: Hypotheses

Several predictions can be made relating to the participants' English proficiency based on findings in the previous sections. While the research does not show a clear advantage for an early or late AoA, there have been studies showing that different linguistic domains are affected differently by AoA. Specifically, a late AoA "can be better" for lexical and grammatical knowledge (Lapkin et al. 1980; Muñoz 2008, as cited in Pfenninger, 2014, p. 533). The hypothesis for AoA will be based on this finding. The hypotheses regarding NL and EXP are based on the results from the FoTo-study (Jenniskens et al., 2018). Since no information could be found on EXP influencing PV (differently than RV or RG), the same effect of EXP on all three tests was expected.

Based on these findings, the following hypotheses are proposed:

1. The students with a higher AoA are expected to score significantly higher than students with a lower AoA on all three measures.
2. a. The students with English as their NL are expected to significantly outperform the students with a different NL than English on all three measures.
b. The internationals are expected to score significantly higher on RG than the Dutch natives.
3. The students with a higher EXP are expected to score significantly higher than students with a lower EXP on all three measures.

3: Method

3.1: Participants

The data was collected from 97 immersion students (42 male- 55 female) who were between the ages of 4;0. and 7;2. The average age of the participants at the time of testing was 5;4 (SD = 1;1). Table 1.1 shows the number of participants with a particular age.

Table 1.1

Number of participants divided by age

Age	Number of students
4	27
5	34
6	17
7	19
Total	97

The participants came from two Dutch/English bilingual primary schools from the city of Utrecht with immersion programmes. The table below illustrates the number of students per grade.

Table 1.2

Number of students per grade

Grade	Number of students
1	58
2	10
3	29

There were substantially more first grade students than second and third grade students. This is mostly due to school 1's large first grade and not having a second grade. It is important to keep in mind these groups are quite small. This means it will be difficult to generalise the results to a larger population. To assess the relation between NL and English proficiency, the participants were divided into three groups based on their NL. These groups are the Dutch natives, English natives and internationals (participants with a different NL than Dutch or English). There were 64 Dutch

natives, 16 English natives and 17 internationals. Table 1.3 shows the number of participants with a particular NL.

Table 1.3

Number of participants based on NL

Number of participants	NL
64	Dutch
16	English
5	Hindi
3	Chinese
2	Turkish
1	Arabic, French, Hungarian, Italian, Moroccan, Polish, Slovakian

Bilingual participants were divided into one of three groups (Dutch, English or internationals) in the following way. If either one of their NLs was English, they were assigned to the English natives. If either one of their NLs was Dutch, they were assigned to the Dutch natives. An English/Dutch bilingual would be put with the English natives. Participants without English or Dutch as their NL were assigned to the internationals. In terms of AoA, this means that the AoA for the 16 English natives is zero since they started their acquisition from birth. Aside from the participants who started acquiring English in a naturalistic setting, there were participants who started acquiring English in school. The range of AoA for these students lay between four and six. Table 1.4 illustrates the distribution of AoA and the number of students with a particular AoA.

Table 1.4

Distribution of AoA and number of participants

AoA	Number of participants
0	16
4	66
5	9
6	6

To assess the relation between EXP and English proficiency, the participants were divided into groups based on the number of minutes of classroom L2 exposure the participants received as a

percentage of their total instruction time per week. This number depended on the participants' school and grade. In table 1.5 is given the EXP in minutes (+percentage of total instruction time) per week.

Table 1.5

Amount of classroom L2 exposure in minutes (including percentage of total instruction time) per week

School	Grade	Minutes	Percentage
1	1	225	15%
	3	105	7%
2	1	240	16%
	2	240	16%
	3	280	19%

3.2: Materials

The following tests were used to determine English proficiency: the Peabody Picture Vocabulary Test 4 (PPVT; Dunn & Dunn, 2007), the Test of Reception of Grammar 2 (TROG; Bishop, 2003), the Expressive Vocabulary Test 2 (EVT; William, 2007) and a non-verbal intelligence test (SON-R; Snijders-Oomen, 1997). These are all standardised tests which means they are administered and graded in the same manner each time. This enables the comparison of (different groups of) participants. The PPVT (228 items) is a picture selection task that measures a person's RV. RV entails the words a person understands when he/she hears or reads them. RG is measured with the TROG (80 items; also a picture selection task) and involves the knowledge of grammatical rules used for understanding the meaning of sentences. On the productive side is the EVT (190 items). The EVT assesses the participants' PV, meaning the words they can produce by, in this case, speaking. The SON-R (33 items) measures non-verbal intelligence, which means there is no spoken or written language involved. It is divided into six subtests of which one, analogies, was administered. This subtest specifically measures the participant's ability for abstract reasoning. The test items of the PPVT, TROG and EVT were shown on a laptop while the SON-R's test items were presented on paper. Each test came with a score form on paper which included the correct answers. A recorder was used to record the EVT so that the participants' answers could later be checked.

3.3: Procedure

Before starting data collection, a letter was sent to the participants' parents asking for their active consent to test their children's English proficiency. Participation was voluntary and neither parents nor children were promised or given a reward. The data was collected over a period of two weeks starting the end of March 2018. The first grade participants took the PPVT and TROG in one session which took about 20-30 minutes. The second and third grade participants were subject to two sessions, the first of which was the same as the first grade participants' session one. Session two took place at least one day after the first session. It included the EVT and SON-R, also taking about 20-30 minutes. The participants were tested during regular school days and were individually taken out of the classroom. Testing took place in a quiet room. The participants were told what tasks they were about to do and what would be required of them. Each test included several practice items to test whether they understood the instructions. If they did not seem to understand what was expected of them, the instructions were repeated. The instructions for all tests were given in Dutch, except the EVT (English). If the participant did not understand Dutch well enough, the language of instruction was switched to English. Not all participants took every test. Table 1.6 shows which participants took which tests.

Table 1.6

Distribution of participants and tests

	Grade 1	Grade 2	Grade 3
PPVT (Session 1)	x	x	x
TROG (Session 1)	x	x	x
EVT (Session 2)		x	x
SON-R (Session 2)		x	x

The reason some participants took a particular test and others did not had to do with their grade. First grade students were not subject to the EVT because research has shown that young students' productive skills are not well enough developed for them to take productive tests. Testing children with a year of instructional input is "likely to result in floor effects" (Unsworth et al., 2014, p. 532). The SON-R was not administered to the first grade students because that would have required them to do two sessions. Taking part in two sessions was believed to put too much strain on the younger children.

The procedures of the tests were as follows. At the start of session one, the participant's age was checked because the PPVT has a different starting point for participants of different ages. The participant was presented with four pictures on a computer screen and subsequently heard an English word by the experimenter that corresponded with one of the four pictures. The participant's goal was to match the word with the correct picture. The participant had the option of either pointing at the correct answer or saying the corresponding number of the answer (out of four). For the PPVT as well as the other tasks, the administrator wrote down the answer and whether the answer was correct on an answer sheet. The TROG is similar to the PPVT in terms of administering and what is expected of the participants. The difference with the PPVT is that the children are presented with an English sentence instead of an English word. Each participant had the same starting point, as was also the case with the EVT and SON-R. As stated, session two consisted of the SON-R and EVT. For the SON-R, the participant was first shown a figure which changed into a (similar looking) figure. Afterwards, another figure was presented which, the participant was told, had to change in the same manner as the first figure did. Four potential options were shown from which the participant had to choose the correct one. Regarding the EVT, the participant was shown a picture (of an object or action) and then asked a question such as 'What is this' or 'What is she doing?'. The participant had to answer the question in English. If the participant did not respond within 10 seconds (as judged by the administrator) on any of the tests (except for the SON-R, for which the participant had 20 seconds to respond), the stimulus was repeated a maximum of two times. After each test was administered, the scores were inserted into an Excel file which were then transferred into an SPSS data file.

3.4: Design and analysis

This experiment included a within subjects' design. This design entails the same participants carrying out all the tasks (which, except for the younger students on session two, holds for this experiment). The dependent variable was the students' English proficiency, as indicated by the number of correct answers on the tests for RV, RG and PV. The independent variables included the AoA, NL and EXP. Extraneous variables such as gender and age and their effect on the dependent variable were also included.

4: Results

The findings of this study are structured as follows: first, the descriptives are given for the tests including the mean score, standard-deviation, minimum score and maximum score to show how the students performed on RV, RG and PV. This section will include the comparison between the participants' mean scores and the mean scores of the students in the FoTo study (Jenniskens et al., 2018). Afterwards, RQs 1a and 1b will be answered by checking for significant differences between students with different NLs, AoAs and EXPs on their test scores. Additionally, any correlations between, on the one hand, the test scores and, on the other hand, IQ, age and gender will be held into account. Each hypothesis will include information on which analysis has been conducted to check the hypothesis.

4.1: Descriptives

4.1.1: English Proficiency Scores based on Grade. In order to compare the scores to the FoTo-study's scores, the students' performance on the three tests is given in table 2.1. The students are divided based on which grade they are in.

Table 2.1

English proficiency scores based on grade (1,2,3)

	Grade	N	Mean	Standard deviation	Minimum	Maximum
1	PPVT Score	58	41.76	24.36	9	113
	TROG Score	58	16.83	15.17	1	66
2	PPVT Score	9	88.33	24.16	53	133
	TROG Score	10	45.70	19.95	15	72
	EVT Score	10	29.60	24.12	3	74
3	PPVT Score	29	78.31	32.34	26	141
	TROG Score	29	40.17	18.90	11	74
	EVT Score	29	21.86	23.88	2	74

The students in the third grade scored higher than the first grade students on both RV (by 36.55 points) and RG (by 23.34 points), but not the second grade students. This is because the second grade, which only consists of students from school 2, obtained higher scores than the third grade students from school 1. These (comparatively low) scores for the third grade students from school

1 are the reason for the second grade students seemingly outperforming the third grade students. The standard-deviations were quite large which can have two causes. First, the number of students per grade is quite low and second, there are both native and non-native speakers of English. Whether there are correlations between these scores and AoA, NL and EXP will be shown in section 4.2.

4.1.2: English Proficiency Scores based on AoA. Table 2.2 shows the English proficiency scores divided on AoA.

Table 2.2

English proficiency scores based on AoA

	AoA	N	Mean	Standard deviation	Minimum	Maximum
0	PPVT Score	16	97.12	25.64	60	141
	TROG Score	16	54.44	16.88	21	74
	EVT Score	8	55.25	21.02	14	74
4	PPVT Score	65	46.77	29.66	9	133
	TROG Score	66	19.24	17.02	1	67
	EVT Score	16	22.44	20.77	2	67
5	PPVT Score	9	64.78	18.63	30	91
	TROG Score	9	32.11	11.78	15	46
	EVT Score	9	10.56	8.32	2	28
6	PPVT Score	6	51.83	14.61	26	64
	TROG Score	6	28.00	12.31	11	38
	EVT Score	6	5.67	2.50	2	8

Participants who started acquiring English from birth outperformed the participants who started at age four, five and six on each measure. Comparing only the participants with AoAs of four, five and six shows that no one AoA scores higher than the others on all three tests. Rather, participants with an AoA of five did better on the PPVT and TROG while participants with an AoA of four outscore the others on EVT.

4.1.3: English Proficiency Scores based on NL. To see whether students with different NLs performed differently, the scores on the tests are divided based on the students' NLs. These scores are shown in table 2.2. DU stands for the Dutch natives' scores, EN for the English natives' scores and INT for the internationals' scores.

Table 2.3

English proficiency scores based on NL

	NL	N	Mean	Standard deviation	Minimum	Maximum
DU	PPVT Score	63	45.81	24.47	9	124
	TROG Score	64	19.64	15.49	1	67
	EVT Score	25	9.04	7.82	2	31
EN	PPVT Score	16	97.12	25.64	60	141
	TROG Score	16	54.44	16.88	21	74
	EVT Score	8	55.25	21.02	14	74
INT	PPVT Score	17	61.65	37.31	12	133
	TROG Score	17	27.65	19.85	2	61
	EVT_Score	6	43.67	16.44	26	67

The English natives scored higher than both Dutch natives and internationals on all three tests. The difference is larger between English natives and Dutch natives than between English natives and internationals. The internationals also outperformed the Dutch natives on each test. Whether these differences are significantly different will be shown in section 4.2.2. The standard-deviations are very dissimilar across NLs and tests.

4.1.4: English Proficiency Scores based on EXP. Table 2.4 illustrates the English proficiency scores based on the participants' EXP.

Table 2.4: English proficiency scores based on EXP

Table 2.4

English proficiency scores based on EXP

	EXP	N	Mean	Standard deviation	Minimum	Maximum
7	PPVT Score	21	67.62	28.08	26	136
	TROG Score	21	34.76	18.22	11	74
	EVT Score	21	14.86	20.15	2	74
15	PPVT Score	48	35.21	18.81	9	95
	TROG Score	48	12.73	10.99	1	66
	EVT Score	0	0	0	0	0
16	PPVT Score	19	80.37	24.74	35	133
	TROG Score	20	41.10	18.88	11	72
	EVT Score	10	29.60	24.12	3	74
19	PPVT Score	8	106.38	26.23	66	141
	TROG Score	8	54.38	12.72	33	67
	EVT Score	8	40.25	24.21	6	74

The participants with 19% EXP scored higher than the other participants on all three tests. The group with the second highest scores were the participants with EXP of 16% followed by the participants with EXP of 7%. The lowest performers were participants with 15% EXP.

4.1.5: Comparison with FoTo-Study. The comparison between this study's participants and the FoTo-study participants is based on mean scores as only those were made publicly available. The group of first grade FoTo-study participants and group of third grade FoTo-study participants are the same participants, measured twice in the span of two years. As stated, the mean scores based on NL will be compared with the FoTo-study's mean scores based on home language. Unfortunately the mean scores based on home language for PV were not provided by the authors of the FoTo-study, not allowing that particular comparison.

Starting with RV, the first grade students outperformed the FoTo-study students by on average 7.51 points and the third grade students scored on average 9.21 points higher than the FoTo-study participants. The two groups of participants seem to score similarly on RG though with the difference being 1.47 points (in favour of the FoTo-study group) for the first grade and 0.66 points for the third grade (in favour of this experiment's participants). As for PV, the third grade participants of this experiment outscored the FoTo-study's third grade students by 7.55 points. Overall, the participants of this experiment outperformed the FoTo-study participants on the vocabulary tasks, but not on RG.

Based on NL (and home language), the participants of this experiment did better than the FoTo-study participants on RV, regardless of their NL. The differences between the two groups based on NL lay between 11.82 points (the Dutch natives in the first grade) and 37.20 points (the internationals in the third grade). These differences were in favour of this experiment's participants. The comparison for RG showed a similar picture. Each group of participants in this experiment, no matter their NL or grade, scored on average higher than their corresponding FoTo-study groups.

4.2: Testing of Hypotheses

4.2.1: Hypothesis 1. The first hypothesis is the following: students with a higher AoA are expected to score significantly higher than the lower AoA students on all measures. In order to test this hypothesis, three multiple linear regressions (one for each test) were computed including Pearson correlational analyses. The correlational analyses will be used to check for correlations and the multiple regression analyses for significant differences between students with different AoAs. Although a multiple regression analysis enables the ability to see which variables best predict another variable (for instance, the RV scores), it was chosen for the reason mentioned previously. Only performing a correlational analysis can often produce untrustworthy results because it does not take the other variables into account. Prior to these tests, tests of normality (the Kolmogorov-Smirnov test and the Shapiro-Wilk Test) were carried out. These indicated that, whether divided on age, gender, AoA, NL or EXP, all the scores on the tests (except scores for IQ and the male students for RV) were not normally distributed. This meant a large bootstrapped

model with multiple variables was required to find reliable results. Putting the variables together in a multiple linear regression model lowers the chances of finding significant effects where there are none (Type 1 error). The variables entered into the multiple regression analysis were age, gender, AoA, NL and EXP. Since ordinal and categorical variables such as NL cannot be computed into a multiple linear regression, NL (and gender) were made into dummy variables so they could be entered in the analysis. Six factors were ultimately entered into the analysis: age, AoA, EXP, female, English and Other. The English dummy variable indicated the difference between the English and Dutch natives while Other indicated the difference between internationals and Dutch natives. The female variable showed the difference between the female and male participants.

4.2.1.1: Receptive Vocabulary. A significant, although not strong (strong = $r > .7$), negative correlation was found between AoA and RV scores, $r = -.484$, $p < .001$. A negative correlation means that as AoA increases, the scores on RV decrease. However, there was another significant correlation, namely between AoA and English, $r = -.948$, $p < .001$ (which holds for all three measures). This indicates a high level of multicollinearity between the two variables which is backed by the collinearity statistics (collinearity tolerance and VIF). This multicollinearity makes it difficult to determine the importance of each of these variables separately for RV scores. As assumed, there was no large overlap between AoA and EXP, $r = -.200$, $p = .026$. This means there was no overlap between these two variables and their relation with RV scores are not affected by the other. Table 3.1 shows the coefficients for the RV model.

Table 3.1

Multiple linear regression model for receptive vocabulary (bootstrapped)

	B	Bias	SE	Sig.	95% CIs
(Constant)	-82.764	-2217	29.933	.008	-149.934, -30.717
Age	20.027	-.080	2.356	.001	15.384, 24.294
AoA	-3.134	.434	4.541	.469	-10.712, 7.073
EXP	2.669	.053	.835	.001	1.078, 4.360
Female	3.230	.186	4.351	.466	-5.408, 12.080
English	37.178	1.711	20.032	.058	2.624, 79.766
Other	2.866	-.229	8.343	.729	-13,444, 18.825

The results indicated that the model explained 62.5% of the variance and that it was a significant predictor of RV performance, $F(6,89) = 24.71$, $p < .001$. As shown above, AoA was negatively but not significantly related with RV ($B = -3.13$, $p = .469$). The B value indicates that as AoA increases with 1 the RV scores decrease by 3.13 points. This means that a higher AoA did not lead to significantly higher scores on RV, rejecting the RV part of hypothesis 1.

4.2.1.2: Receptive Grammar. AoA negatively correlated (quite) strongly with RG scores, $r = -.512$, $p < .001$. The coefficients for RG are given in table 3.2.

Table 3.2

Multiple linear regression model for receptive grammar (bootstrapped)

	B	Bias	SE	Sig.	95% CIs
(Constant)	-69.631	-3.098	20.926	.004	-115.674, -30.814
Age	11.619	.071	1.585	.001	8.524, 14.997
AoA	1.394	.432	3.421	.673	-4.588, 8.827
EXP	1.621	.056	.558	.008	.556, 2.762
Female	1.260	.183	2.654	.641	-3.560, 6.763
English	40.505	1.970	15.414	.013	11.612, 73.276
Other	1.317	.063	4.525	.760	-7.431, 10.398

The model explained 65.6 % of the variance and was a significant predictor of performance on RG, $F(6,90) = 28.66$, $p < .001$. It also showed that AoA ($B = -1.39$, $p = .673$) was not a significant predictor of RG scores. Students with a higher AoA did not score significantly higher than students with a lower AoA. This result rejects the RG part of hypothesis 1.

4.2.1.3: Productive Vocabulary. A significant negative correlation was found between PV scores and AoA, $r = -.726$, $p < .001$, suggesting that as AoA increases, the PV scores decrease. This would be in line with the hypothesis. Table 3.3 shows the coefficients for the PV model.

Table 3.3

Multiple linear regression model for productive vocabulary (bootstrapped)

	B	Bias	Std. Error	Sig.	95% CIs
(Constant)	-24.773	1.350	25.646	.361	-73.594, 28.881
Age	3.743	.053	3.631	.305	-2.954, 11.375
AoA	-1.101	-.176	3.289	.726	-8.167, 5.429
EXP	1.197	-.056	.667	.096	-.203, 2.410
Female	5.535	-.504	4.831	.278	-4.847, 14.566
English	39.424	-.418	15.169	.017	7.127, 67.616
Other	23.784	.348	8.421	.017	6.415, 40.469

The model explained 77.4% of the variance and was a significant predictor of PV performance, $F(6,32) = 18.56$, $p < .001$. A higher AoA does not correspond with significantly higher scores on PV ($B = -1.10$, $p = .726$) compared to a lower AoA, rejecting the PV part of hypothesis 1.

4.2.2: Hypothesis 2. Hypothesis 2 is two-fold: (2a) Students with English as their NL are expected to outperform students with a different NL than English on all three measures. The second part (2b): The internationals are expected to score significantly higher on RG than the Dutch natives. These hypotheses were tested using the same multiple linear regressions and Pearson correlational analyses as with hypothesis 1.

4.2.2.1: Receptive Vocabulary. The strongest correlation between NL and any variable (or test score) was between having English as a NL and RV scores, $r = .545$, $p < .001$. This suggests that those with English as their NL tend to score quite high on RV. The model (due to implementing dummy variables) unfortunately only showed the comparison between the English- and Dutch natives and the internationals and Dutch natives. The performance of the English natives and internationals thus cannot be checked for significant differences but can be implied by looking at the mean scores given in table 2.2.

Table 3.1 shows the English natives almost scored significantly higher than the Dutch natives ($B = 37.18$, $p = .058$). Since the p-value is not below .05, it cannot be said the English natives significantly outscored the Dutch natives. The difference in mean score between the English natives and internationals was smaller (35.47 points) than between the English- and Dutch natives (51.31 points). This implies the English natives also did not score significantly higher than the internationals (see table 2.2). These two findings reject the RV part of hypothesis 2a.

4.2.2.2: Receptive Grammar. As with RV, there was a fairly strong correlation between English as a NL and the test scores (RG), $r = .596$, $p < .001$. A similar pattern to RV (but with statistical significance) was expected in the model. This was confirmed by the regression analysis (table 3.2) which showed students with English as their NL scored significantly higher than the Dutch natives ($B = 40.51$, $p = .013$). The mean scores for RG showed the English natives outscored the internationals by 26.79 points and the Dutch natives by 34.80 points. Whether the difference between English natives and internationals is significant is hard to determine. However, since the former scored twice as high as the latter, a significant difference is assumed here. Thus, the RG part of hypothesis 2a cannot be rejected. As for hypothesis 2b, the internationals did not score significantly higher than the Dutch natives on RG ($B = 1.32$, $p = .760$), rejecting the hypothesis.

4.2.2.3: Productive Vocabulary. English as a NL correlates (fairly) strongly with PV scores, $r = .677$, $p < .001$. The model, as shown in table 3.3, showed that the English natives scored significantly higher than the Dutch natives ($B = 39.42$, $p = .017$). As the difference in mean score between the English natives and internationals was only 11.58 points, it is assumed there is no significant difference. These two findings partially reject the PV part of hypothesis 2a. The model also showed the internationals significantly outperformed the Dutch natives, which was not hypothesised ($B = 23.78$, $p = .017$).

4.2.3: Hypothesis 3. Hypothesis 3 is as follows: a higher EXP is expected to result in significantly higher scores on all three measures. This hypothesis was tested in the same manner as hypotheses 1 and 2.

4.2.3.1: Receptive Vocabulary. There was no strong or significant correlation between EXP and RV scores ($r = .022$, $p = .416$) or between EXP and other variables. The model (table 3.1), however, states that a higher EXP corresponds with higher RV scores ($B = 2.67$, $p = .001$), failing to reject the RV part of hypothesis 3.

4.2.3.2: Receptive Grammar. As with RV, no strong or significant correlation between EXP and the test scores (RG) was found, $r = -.023$, $p = .412$. This negative correlation (although not significant) goes against the hypothesised effect of EXP. Nevertheless, the regression analysis showed students with a higher EXP significantly outperformed the students with a lower EXP on RG ($B = 1.62$, $p = .008$). This means the RG part of hypothesis 3 cannot be rejected.

4.2.3.3: Productive Vocabulary. There was a significant though fairly weak correlation between EXP and PV scores, $r = .434$, $p = .003$. In line with the weak correlation, the model showed that a higher EXP does not lead to significantly higher PV scores ($B = 1.20$, $p = .096$), rejecting the PV part of hypothesis 3.

4.2.4: Other variables. This section will shortly elaborate on the effect of extraneous variables such as age, gender and IQ on the test scores. The same regression analyses were used as with testing the hypotheses but now including the IQ scores as an additional variable. Beginning with RV, the model showed that, out of the three variables mentioned, only age was a significant predictor ($B = 20.03$, $p = .001$). Whether the student was male or female or had a higher IQ had no significant effect on RV scores. The same finding applied for RG for which only age was a significant predictor of the test scores ($B = 11.62$, $p = .001$). As for PV, none of the three variables significantly predicted PV scores.

5: Discussion

This experiment was set up to answer RQs 1a and 1b. RQ1a was the following: is there a correlation between, on the one hand, the students' *AoA*, *NL*, *EXP* and, on the other hand, the students' English proficiency? RQ1b was: do students with different *AoAs*, *NLs* and *EXPs* significantly differ in their English proficiency? Both RQs were answered with the findings of the experiment. Further information on how the answers are supported by the results will be given. The findings will then be related to the literature discussed in the theoretical framework. Additionally, this section will discuss the implications of the findings, limitations of this experiment and suggestions for further research.

5.1: RQ1a

AoA negatively and significantly correlated with the scores on every test with fairly strong to strong correlations (between -0.484 and -0.726). As for *NL*, there were significant positive correlations between English as a *NL* and all three tests with correlation strengths similar to those of *AoA*. Having a different *NL* than English (Dutch natives or internationals) did not correlate with English proficiency scores. *EXP* correlated only significantly with *PV* though not very strong. No significant correlation was found between *EXP* and *RV* and *RG*. To shortly summarise, *AoA* and English as a *NL* significantly correlated with all three tests, but *EXP* correlated only with *PV*. Where the correlation between *AoA* and test scores was negative, the correlation between English as a *NL* and test scores was positive. *AoA*'s negative correlation means the higher a student's *AoA*, the lower their scores were on all three tests. Reversely, having English as a *NL* lead to higher scores on the tests. *EXP*'s correlation with *PV* was also positive. Thus, there were correlations between, on the one hand, the students' *AoA*, *NL*, *EXP* and, on the other hand, the students' English proficiency. The correlations differed per variable, as was to be expected. To further investigate the relation between these variables and English proficiency, RQ1b was created.

5.2: RQ1b

RQ1b was answered by testing the hypotheses, which were as follows:

1. The students with a higher *AoA* were expected to score significantly higher than students with a lower *AoA* on all three measures.

2. a. The students with English as their NL were expected to significantly outperform the students with a different NL than English on all three measures.
b. The internationals were expected to score significantly higher on RG than the Dutch natives.
3. The students with a higher EXP were expected to score significantly higher than students with a lower EXP on all three measures.

As stated, there was no overlap between AoA and EXP which means that the effect of either of these variables on English proficiency scores is not due to the other variable. The findings for AoA were contrary to what was expected. A higher AoA did not lead to significantly higher scores on RV, RG or PV. Hypothesis 1 can thus be rejected. Hypotheses 2a and 2b revolved around the student's NL. The English natives did, although it was very close ($p = .058$), not significantly outscore the Dutch natives or the internationals (based on differences in mean score) on RV. The opposite was shown for RG with a significant difference in favour of English natives compared to Dutch natives and internationals (although for the internationals it is an assumption based on mean scores). For PV, the English natives significantly outperformed the Dutch natives but not the internationals. These findings partially reject hypothesis 2a. Hypothesis 2b was also rejected as there was no significant difference between the internationals and Dutch natives on RG. Hypothesis 3's focussed on the effect of EXP. Results showed that a higher EXP corresponds with higher scores for RV and RG, but not for PV. There was a near significant difference ($p = .096$) in favour of a higher EXP for PV though. Hypothesis 3 can thus be partially rejected.

To answer RQ1b: students with different AoAs did not significantly differ in their English proficiency. The same holds for students with different NLs on RV, but not on RG and PV where there were significant differences between the three groups of NLs. Different EXPs lead to significant differences on RV and RG but not on PV. Thus, students with different AoAs do not significantly differ, whilst students with different NLs and EXPs do.

5.3: Connection to the literature

The results both support and conflict with what the literature states on the effects of these variables on English proficiency in immersion programmes. The findings will be related to the literature per variable.

5.3.1: AoA. In a naturalistic second language acquisition setting, there seems to be an ultimate attainment advantage for early starters whilst late starters seem to have an initial rate advantage (Muñoz, 2008). Whether this applies to an instructional setting is yet unclear: there is no consensus on a clear advantage in favour of either an early (age four) or late (age six-eight) start (Genesee, 2004). This experiment's results seem to be in line with this statement. When looking at specific linguistic domains though, AoA has been shown to affect certain domains differently (at least in total immersion programmes) (Pfenninger, 2014). Lexical and grammatical knowledge (which includes RV, RG and PV) seem to benefit from a later AoA, which was hypothesised (Lapkin et al. 1980; Muñoz 2008, as cited in Pfenninger, 2014, p. 533). However, the results are contrary to what these researchers have found as no significant differences were found between students with different AoAs on any test. The difference between this experiment and the experiments mentioned by Muñoz (2008) and Pfenninger (2014) is that this experiment is based on partial immersion programmes as opposed to total immersion programmes. This might explain (some of) the different findings.

5.3.2: NL. The literature on NL was based on the FoTo-study as the FoTo-study population and this study's population are relatively similar (both mostly Dutch natives living in the Netherlands) as opposed to studies on immersion programmes abroad. This experiment, for the most part, corroborated Jenniskens et al.'s (2018) finding that students with English as a home language outperformed the students with a different home language. Only on RV did this study's participants with English as their NL not significantly (although it was very close) outperform the other students. This is quite surprising as first, the FoTo-study suggested otherwise and second, it would seem logical that the English natives would score significantly better in their NL than Dutch natives in their L2. Additionally, the internationals did not score significantly higher than the Dutch natives on RG, as was the case in the FoTo-study (hypothesis 2b). As Jenniskens et al. (2018) did not provide a reason for this specific result, it is difficult to explain the difference

between the two outcomes. A possible reason might be a different distribution of native languages between the internationals of the two studies, but this information was not provided by Jenniskens et al. (2018). The internationals did significantly better than the Dutch natives on PV. This was not the case in the FoTo-study and therefore not hypothesised. A possible explanation might be found in the number of internationals that took the PV test, being only six. The internationals' performance, although not outperforming the Dutch natives as much on RV and RG, might thus be explained by reasons other than their NL. Overall, the results on NL's effect on English proficiency supports previous findings with the few exceptions mentioned. These exceptions could be the subject for further study which will be elaborated on in section 5.5.

5.3.3: EXP. There is a large body of literature on the effect of quantity of L2 exposure in immersion programmes. This is not surprising as increasing EXP is to a certain extent in the hands of the programme coordinators. There is no consensus on L2 exposure's effect on L2 proficiency in an instructional setting (or on certain domains of L2 proficiency (Steinlen et al. 2010; Unsworth 2013). There seems to be a tendency for L2 exposure to positively correlate with L2 proficiency though (Lyster & Genesee, 2013). However, these studies speak of L2 exposure and not EXP. L2 exposure encompasses all L2 exposure a child receives while EXP is the L2 exposure received in the classroom. That is why the hypotheses are based on the FoTo-study; the authors included EXP as a variable for RV and RG. The same results as in the FoTo-study were found where a higher EXP lead to significantly higher scores on RV and RG. The results thus corroborate Jenniskens et al.'s (2018) findings in terms of RV and RG. Keeping in mind the discrepancy between EXP in this study and the FoTo-study, this result emphasises EXP's importance as even a smaller difference (compared to FoTo-study schools) in EXP (between 7-19%) can result in higher scores on RV and RG. The FoTo-study did not study the effect of EXP on PV meaning this experiment's results on EXP and PV cannot be linked to the FoTo-study. The fact that students with a higher EXP did not significantly (although near significance levels) outperform students with a lower EXP on PV was somewhat surprising. This null result, which was not hypothesised, is hard to explain as no information could be found on EXP influencing receptive and productive skills differently.

5.4: Implications

The results of experiment might imply that AoA on its own does not automatically lead to higher L2 proficiency. In this study, students with an earlier AoA were not subjected to a high EXP (as indicated by the low correlation). Thus, AoA seems to not have as large an effect on L2 proficiency on its own. This would support Muñoz's (2008) idea that AoA's influence on L2 proficiency depends on whether an early AoA is accompanied by more L2 exposure. This could potentially explain AoA's lack of effect on L2 proficiency. In terms of NL, there is not much that can be implied from the results. NL was taken into account to see if there were large differences between Dutch natives and internationals and small differences between English natives and other students. Neither of these differences were large enough to be of concern. If, for example, the English natives did only slightly better than the other students, there could be a problem in how the programme was set up (disproportionate amount of attention given to Dutch natives or internationals for example). The potentially most important result might be on EXP. A small difference in EXP (between 7-19%) was shown to lead to significantly higher scores on both receptive tasks. This could mean it is easier for programme coordinators to influence the students' L2 proficiency as a rather small change in EXP is needed.

5.5: Limitations

This study has several limitations. First, there is the aforementioned multicollinearity between AoA and English as a NL. This was expected though (as well as unavoidable considering the limitations in terms of detailed questionnaires) since the English natives all started acquiring English at birth. Nevertheless, enough difference between the two variables is assumed. This is because English as a NL was responsible for significant differences on two of the three measures (RG and PV) while differences in AoA did not lead to significant differences on any measure. Second is the small number of participants which, especially when divided into smaller groups based on AoA/NL/EXP, makes it difficult to generalise the results to a larger population. This is because the fewer participants there are, the larger the chance is that the results are due to chance. Third, apart from AoA, NL and EXP, there are more important variables in L2 acquisition. Examples of these are SES, language aptitude, motivation, L2 exposure outside the classroom and attitude towards language learning. Including these variables was impossible for practical reasons. If (some of) these variables were included, a larger amount of the variance in the scores could have

been accounted for. These variables are to a large extent not under the control of the programme coordinators. Variables such as language allocation and teachers' English proficiency are also important for students' L2 development. This is why these will be discussed in the qualitative part of this study. Fourth, using a multiple linear regression with dummy variables meant the model did not show whether there were significant differences between English natives and internationals. These were based on mean scores which is less reliable.

5.6: Further Areas of Research

Future research could focus on the relation between NL and RG because this study's results failed to replicate the findings in the FoTo-study (internationals outperforming the Dutch natives). Moreover, since the internationals scored significantly better on PV (which was not the case in the FoTo-study), this could also be subject to further research. Additionally, EXP's effect on receptive vs. productive skills in immersion education could be subject to further study as not much literature exists on this topic. For any of these suggestions, a larger number of participants is advised to be able to generalise the results.

6: Conclusion

The goal of part one was to study the relation between the *AoA*, *NL* and *EXP* of immersion students and their English proficiency. As immersion education is still at an early stage in the Netherlands, it is vital to know which characteristics influence English proficiency in what way. To achieve this goal, two RQs were created: (RQ1a) is there a correlation between, on the one hand, the students' *AoA*, *NL*, *EXP* and, on the other hand, the students' English proficiency? and (RQ1b) do students with different *AoAs*, *NLs* and *EXPs* significantly differ in their English proficiency? Three hypotheses were constructed in order to answer RQ1b: (1) students with a higher *AoA* are expected to score significantly higher than students with a lower *AoA* on all three measures, (2a) students with English as their *NL* are expected to significantly outperform students with a different *NL* than English on all three measures, (2b) internationals are expected to score significantly higher on *RG* than Dutch natives and (3) students with a higher *EXP* are expected to result in significantly higher scores than students with a lower *EXP* on all three measures. A quantitative experiment was conducted in which bilingual primary school students were subject to tests on their English *RV*, *RG* and *PV*.

Results showed English as a *NL* and *AoA* significantly correlated with all three tests while *EXP* correlated only with *PV*. There were no significant differences on the three tests between students with different *AoAs*, rejecting hypothesis 1. Hypothesis 2a was partially rejected as significant differences were found on *RG* and *PV*, but not on *RV* (between English natives and other students). Moreover, since the internationals and Dutch natives did not score significantly differ in their performance on *RG*, hypothesis 2b was rejected as well. Hypothesis 3, as with 2a, was partially rejected because a higher *EXP* only lead to significantly higher scores on *RV* and *RG* but not on *PV*.

This experiment had several limitations, the largest of which was the small number of participants. Speaking in terms of applicability, the results on *EXP* might be the most valuable for coordinators of immersion programmes as a rather small difference in *EXP* lead to significantly higher scores on both receptive tasks.

7: Introduction

The second part of this thesis revolves around RQ2 which is: what are the schools' plans regarding the influx of new students the coming year(s) in terms of policy and curriculum design and how are they going to accommodate the Dutch, English and international students? Answering RQ2 provides a more detailed examination of first, the schools' programmes, and second, discusses other factors that influence immersion students' L2 development. Several topics/variables concerning the classroom/teachers will additionally be discussed. These variables include teachers' English proficiency, language allocation, quality of exposure (QoE) and use of the students' L1s. So whilst the quantitative experiment focussed on individual characteristics, the qualitative component focusses on factors based on the classroom. Thus, additional attention is given to variance in the English proficiency scores that was not explained by the regression models. As stated in section 5.5, a limitation of the study was the inability to include more variables that influence students' L2 proficiency. This part of the thesis seeks to in part remedy this limitation and discuss several other variables that are deemed important in immersion education. These variables will be detailed in section 8. Part two is a follow-up on part one and therefore makes use of part one's results. The results might influence the programme coordinators in their decision to, for example, decrease or increase the EXP. Following this introduction come the theoretical framework, method, results, discussion of the interviews and conclusion. The theoretical framework will address several factors in immersion while the results and discussion will revolve around the interviews and their contribution to RQ2. This thesis will end with a general conclusion on parts one and two.

8: Theoretical Framework

The theoretical framework will discuss several variables that were not included in part one. The models used in part one explained between 62.5% and 77.4% of the variance found in the English proficiency scores. A certain amount of variance, though impossible to determine precisely, is due to other variables. Early language learning has often been studied in relation to individual characteristics and less so in the perspective of “teaching and teacher-related factors” (Graham et al., 2017, p. 925). This is not to say these factors are insignificant for students’ L2 development. Therefore they are included in this section. QoE and students’ use of their L1s will be elaborated on shortly. Since estimations of teacher L2 proficiency levels and information on language allocation have been provided by the programme coordinators, these variables will also be discussed in light of the two programmes. This qualitative study is of a descriptive character hence the absence of a hypothesis.

8.1: Quality of Exposure

QoE has been described as the “authenticity and richness of children’s L2 input” (Jia & Aaronson 2003, as cited in Sun, Steinkrauss, Wieling & de Bot, 2018, p. 408). High QoE can be achieved with a combination of the following features: “a balance between adult-initiated group work, child-chosen ‘potentially instructive’ play activities, curriculum differentiation and cognitive challenge, and sustained shared thinking” (Siraj-Blatchford & Sylva 2004; as cited in Stephen, McPake, Pollock & McLeod, 2016, p. 65). There is not much literature on the effect of QoE on students’ L2 development in an instructional setting (Graham et al., 2017). This is partly due to the difficulty of assessing QoE as it can be composed of multiple factors. These include parental L2 proficiency, teacher L2 proficiency and didactics, native speaker input and different types of L2 media input. Nevertheless, QoE is vital for children’s L2 learning in an instructional setting (Graham et al. 2017; Sun et al. 2016). This aligns with Genesee’s (2004) statement in which he names quality of instruction as the strongest predictor of L2 proficiency in a classroom setting. Thus, based on previous research, it is safe to assume that QoE plays a large role in the development of immersion students’ L2.

8.2: Use of Students' L1s

There are arguments for and against allowing students to use their L1 or the programme's L1 (Dutch in this case). An argument in favour is that since many students often know the programme's L1, students use it to “move the task along”, “focus their attention”, and “interact personally” (Swain & Lapkin 2000, as cited in Cammarata & Tedick, 2012, p. 254). The L1 would thus aid in acquiring the L2. Moreover, for minority students (English in this case), mastery of their L1 provides a solid foundation for acquiring the majority language (Cummins 1986, as cited in Swain & Lapkin, 2005). An argument against using the programme's L1 is that it will take time away from the programme's L2 which would hinder students' L2 development.

The classrooms of schools 1 and 2 contained a mix of different L1s which could potentially create some difficulties. Swain and Lapkin (2005) discuss the possibility of allowing students to use their L1 in the classroom. If a student has an L1 the other students do not have, it will become difficult to support that student's L1. Is it worth putting much time in supporting that specific L1? Particularly in the early years of immersion, the focus lies on the programme's L1 since that is most often the majority language. The majority language students as well as the others naturally need to learn the majority language. An argument for allowing the use of different L1s in the classroom comes from a sociocultural standpoint: “speaking and writing are tools that we use to learn” (Swain & Lapkin, 2005, p. 178). When L2/third language proficiency is inadequate, people might turn to their L1 to deal with the task at hand (Swain & Lapkin, 2005). Also, if a student's L1 “is not supported, then potentially, though not necessarily, as we have seen, subtractive bilingualism rather than additive bilingualism may result” (Swain & Lapkin, 2005, p. 177). Often though, the decision to support which of the students' L1s is based on the practical aspect of whether the teacher (or other students) has knowledge of that language.

8.3: Teacher L2 Proficiency

Apart from educational background, teaching experience in (bilingual) primary schools and vision of the best manner of instructional L2 learning (for example, focus on implicit vs. explicit instruction), there is the teacher's level of L2 proficiency. Do teachers have to be native speakers of the L2 or not? Both options seem to have advantages and disadvantages. Native speakers usually have an edge over non-native speakers in terms of pronunciation and knowledge

of the L2 culture. Non-native speakers, on the other hand, have experience learning the L2 themselves and can share strategies they used to learn the L2 (Ellis & Berger 2003, as cited in Sutherland, 2012). Whether native or non-native, the optimal level of teacher language proficiency is difficult to determine (Graham et al., 2017). There is as of yet no consensus on what the precise effect of the L2 mastery/knowledge of the teacher is on young students' L2 development (Corda, Phielix & Krijnen, 2012). Still, it is logical to assume a teacher who knows the language more thoroughly will have a more positive effect than a teacher who is less knowledgeable.

The report by van den Broek et al. (2014) showed that a teacher's L2 proficiency (as well as didactical skills) can have an effect on the students' L2 development. The FoTo-study even found correlations between self-assessed levels of L2 proficiency and student performance. Students of teachers who rated themselves at C2 level (according to the Common European Framework of Reference or CEFR) scored significantly higher on RV (though similarly on RG) than students of teachers who rated themselves at C1 level (Jenniskens et al., 2018). The self-assessed level of the school 1 teachers complies with B1/B2 level and B2/C1 level for the school 2 teachers (n.a., personal communication, July 12, 2018). Self-assessment of L2 proficiency is not as clear-cut as a standardised test but it still gives a picture of L2 proficiency.

8.4: Language Allocation

There are multiple ways of exposing students to the L2. Language allocation in the classroom can be provided in three ways. First, the one teacher – one language (OTOL) approach in which there are two teachers, each one communicating in either language of the programme. Second, the one situation – one language (OSOL) approach entails the teacher(s) teaching in both languages, but in separate situations. Often, the one language is used one part of the day and the other language the other part of the day. Third, the sandwich-method means the teacher(s) mix the two languages during lessons. Choosing one of these options is often based on practical and organisational grounds (Jenniskens et al., 2018). School 1 makes use of a mix between OTOL and OSOL. Teachers use either language for different activities (OSOL), but there is one native English teacher who teaches the more advanced students. Some students thus learn via OSOL and some via OSOL and OTOL. School 2 only uses the OSOL-approach. The effect these 3 methods have on students' L2 proficiency has been studied by Jenniskens et al. (2018). OTOL-students

significantly outperformed the OSOL- and sandwich-students on RV. For RG, the sandwich-students did significantly worse than the OTOL- and OSOL-students, but there was no significant difference between the OSOL- and OTOL students.

9: Method

9.1: Participants

The data for this qualitative study was collected from face-to-face interviews with the two programme coordinators of schools 1 and 2. Both signed an informed consent form agreeing to participate. Participant 1 (school 1) is female, age 46 and has three years of experience as programme coordinator. Participant 2 (school 2), also female and age 46, has seven years of experience. Information on the identity of the interviewees, as with the schools, will not be provided.

9.2: Materials

Two semi-structured interviews with open-ended questions were conducted. Both interviewees were asked eight questions (see appendix A) with the possibility of follow-up questions. There was no set time for the length of the interviews. The topics revolved around the future programmes of the schools. The interviews were recorded so the interviews could be transcribed (see appendix B and C).

9.3: Procedure

The interviewees were contacted via email in June 2018 explaining what the purpose of the interviews are, asking for their permission and to settle a date for the interview. Interview 1 with the programme coordinator of school 1 was held on 11 July 2018 and interview 2 (with the school 2 programme coordinator) on 12 July 2018. The questions were sent several days before the interview took place so the interviewee could prepare for the interview. The interviews took place in schools 1 and 2 and were conducted by the researcher. Before the start of the interview, the interviewees were asked to sign an informed consent form and told the interviews would be recorded. No reward was given for participation. Interview 1 lasted 40 minutes and interview 2 took 21 minutes.

9.4: Data analysis

An interview and qualitative analysis was deemed appropriate to answer RQ2 as interviews with the programme coordinators offer the option to ask detailed questions about the programmes. The qualitative analysis enables the easy analysis of the data using different topics. The data was

first examined to see which parts of the interview would be useful in helping describe the programmes. These parts were then transcribed. The analysis consisted of dividing these parts based on their topic.

10: Results

The findings of the interviews are divided by topic with the interviewees' statements given for each topic. The subjects of the topics were based on their importance in describing the programmes' futures.

10.1: New Students and Admission Policy

This topic revolves around new students starting the coming year, the admission policy and size of the classrooms. In terms of school size, school 1 will welcome 100 new students on top of the existing 150 with 50 more joining during the year. School 2 had 40 students and will have 14 new students starting next year. 20 more will join during the year. The class size lies around 28 for school 1 and 18 for school 2. Both schools adhere to a "postcode" policy. A "postcode" policy entails accepting only applicants within a certain area. Interviewee 1 says this is the reason for the large number of new Dutch students compared to non-Dutch natives. Also, both schools only allowed students with a certain level of Dutch proficiency. Interviewee 1 claims that the internationals whose Dutch is underdeveloped can potentially struggle with their English development. This is due to them being immersed in a Dutch environment and thus spending more time on their Dutch development. Additionally, they are exposed to more Dutch in school than English. Interviewee 2 added that if the student cannot communicate in either English or Dutch, the chances of them being accepted into the school become very slim as communication will be difficult. Also, if a first grade student's English is good enough, their Dutch development will go smoothly as he/she can focus more on their Dutch.

10.2: Students' NL

Only school 1 had information on the distribution of the new students' nationality (on which NL is based). Eighty-nine of the 100 new students are Dutch, five from India, two Filipino and one from Poland, Japan, Indonesia and one yet unknown (n.a., personal communication, July 13, 2018). Both interviewees believed there is benefit to having both Dutch and English natives in the same classroom because they can learn from each other. Nevertheless, there will be a difference between the two schools in terms of policy regarding NL distribution in the classroom. Interviewee 1 stated the distribution was not held into account, but that, in hindsight, it would have been a good idea. Interviewee 2 said the NL distribution was held into account for the group of third and fourth

graders (shared group). This was done in order to find a good balance between Dutch and English natives. Too much emphasis on Dutch, as a consequence of many internationals, can result in a delay in the students' English development, according to interviewee 2. Moreover, according to interviewee 1, (too) many English natives/internationals in the classroom can lead to a switch from teaching Dutch as an L1 to teaching Dutch to an L2. Since both schools aim to be a bilingual school primarily for Dutch children learning English as an L2, this is not desirable.

Also, interviewee 1 stated that a large number of English natives in the classroom makes it more difficult to teach all the students simultaneously because of the different levels of English proficiency. This problem will most likely occur in school 1's grades five, six and seven because these will be comprised of 50% English natives. Additionally, since the new Dutch students in these grades will not have received immersion, their English proficiency will not be very high. Because the gap in English proficiency between Dutch and English natives was deemed too large, a change in curriculum has been implemented for school 1. In order for lessons to be beneficial for both groups, Dutch natives will first receive 45 minutes of pre-teaching preparing them for the shared lesson with English natives. Interviewee 2 did not find the gap between school 2's Dutch and English natives large enough to be of concern.

10.3: Teachers' L2 proficiency and Language Allocation

In terms of teacher L2 proficiency and language allocation, not much will change for both schools. School 1 will have more teachers to accommodate the new students, but the average L2 proficiency will stay the same. These are all estimated to be above vvto-level (B1), but not yet on the preferred level of a tpo teacher (C1), indicating no change in relation to the previous year. The same language allocation will also apply next year involving a combination between the OTOL- and OSOL-approach. It is still mostly OSOL since there is only one native English teacher. For school 2, the same levels of teacher L2 proficiency (B2/C1) and language allocation (OSOL) will apply the coming year. Moreover, no additional teachers needed to be hired.

10.4: EXP

Interviewee 1 stated that because of the large influx of Dutch students, EXP is expected to thin out. This is why school 1 is aiming to increase EXP to 25-30% the coming years where last

year EXP was 15% for grade one and 7% for grade three. This aim does not apply to the third grade as more attention will be given to Dutch to teach the students to read and write. Interviewee 1 expects reaching 25-30% EXP will take several years. This is because most new teachers need to be familiarised with the bilingual classroom. Also, doubling EXP might lead to a decrease in the amount of Dutch in the classroom. Interviewee 2 stated that the EXP in school 2 will not change as the students' English development is thought to be going well and the current EXP is believed to contribute to that. Also, it is vital that the children learn Dutch and increasing the EXP could potentially delay the students' Dutch development.

11: Discussion

The purpose of this section is to answer RQ2: what are the schools' plans regarding the influx of new students the coming year(s) in terms of policy and curriculum design and how are they going to accommodate the Dutch, English and international students? The future programmes will be described as clearly as possible with the information collected from the interviews. The two programmes will be discussed and related to the literature in the theoretical framework separately.

11.1: School 1

The programme of school 1 will undergo several changes in relation to the previous year. This is largely due to the large increase of students which will double the number of students at the end of next year. The number of groups will increase from 8 to 14. Since school 1 aims to offer a Dutch/English immersion programme primarily for Dutch children, it seems logical that almost 90% of the new students is a Dutch native. The "postcode" policy is one of the reasons for this high percentage. Additional reasons include English becoming more and more important in the Netherlands and the fact that bilingual primary schools are a new occurrence. The large number of Dutch natives has multiple consequences for the programme/classroom. First, a positive consequence is that a larger percent of the classroom is on the same level of Dutch and English making it easier for the teacher in terms of differentiation. Having more English natives and internationals means there might be larger differences between the students' levels of Dutch and English. A negative consequence is that there will not be many English natives for the Dutch natives to converse with and learn from. The NL distribution was not held into account, but this does not have to be an issue.

The gap in English proficiency between Dutch and English natives was, in some instances, found to be too large to simultaneously teach both groups effectively. To address this issue, there is going to be a change in the way the curriculum is set up for grades four, five, six and seven. Before the actual IPC (school 1's method) lesson in which both groups are taught, the groups will do a separate activity. The English natives will do a task on their level while the Dutch natives are prepared for the IPC lesson. In this pre-lesson the Dutch natives will be familiarised with the

vocabulary used in the IPC lesson. This way both the Dutch and English natives will benefit from the lesson as they are now both able to understand it.

Aside from the fact that the teachers will most likely be new to immersion, the situation regarding teachers will stay the same. This means the estimated levels of teacher L2 proficiency and language allocation are similar compared to previous year. The OTOL-approach resulted in significantly higher scores than the OSOL-approach on RV in the FoTo-study (Jenniskens et al., 2018). However, since school 1 mostly uses the OSOL-approach (because there is only one native English teacher), the students will probably not gain the full benefit of the OTOL-approach for RV.

In terms of EXP, school 1 is currently closer to the EXP levels of vvto schools than of tpo schools (in the FoTo-study). The plan is to increase EXP to 25-30% the coming years which would come closer to the average EXP of the tpo schools in the pilot (24% for third grade) (Jenniskens et al., 2018). A consequence of this plan is that it will take away instructional time in Dutch. Still, this will probably not be a problem because of the large number of Dutch natives. If there were more English natives/internationals, decreasing the amount of Dutch in the classroom might be detrimental to a larger number of students' Dutch development.

11.2: School 2

School 2's programme will stay mostly similar compared to last year. Interviewee 2 states this is because the growth of the school and the students' progress are going as planned. The "postcode" policy allowed the school to create a Dutch/English immersion programme primarily for Dutch children. Non-Dutch natives could still apply but also not be accepted into the school. Too many internationals would lead to school 2 turning into an international school for which there are insufficient means. The distribution of students with different NLs was thus held into account. Interviewee 2 thinks the desired balance between students with different NLs has been found. School 2 will continue with the same teachers as last year. Grade one/two (which had two teachers) will split into two groups and grade three/four will have more students this year. This means there will not be any changes in teacher L2 proficiency or language allocation.

Moreover, the same EXP (around 16-19%) will be applied the coming year for all grades. As with school 1, this level EXP is below the average EXP of the two schools in the pilot (Jenniskens et al., 2018). There is a difference between first/second graders and third/fourth graders in terms of how they receive their English instructional time. Grades one/two will have 4 activities in English each day, roughly culminating in an hour of English per day. Grades three/four will receive their EXP during the afternoon which is approximately one hour and 15 minutes. School 2 has of yet not had a fourth grade meaning preparations for teaching these students to read and write in English is underway. There will be no difference in EXP between students with different NLs.

12: Conclusion

School 1's programme will be adjusted to accommodate the large number of new students starting next year. The "postcode" policy lead to mostly Dutch natives enrolling for next year, but since that was the school's goal it is not thought to be an issue. Next year will see more teachers to accommodate the larger number of students. Regardless, teacher L2 proficiency and language allocation will stay the same. The following changes to the programme were made: pre-lessons for Dutch natives and an increase in EXP the coming years. Pre-lessons will solve school 1's problem regarding the large difference in English proficiency between Dutch and English natives. The goal in terms of EXP is to increase it to 25/30% which is more commonly for immersion programmes instead of the previous 7-19%. The EXP increase is implemented to further develop the students' English proficiency.

School 2's programme and its students' progress is going well and therefore there will not be any large changes. The number of new students do not call for adjustments in terms of staff, NL distribution or language allocation.

13: General Conclusion

This final section concludes the findings of parts one and two. The purpose was two-fold: to study the relation between the *AoA*, *NL* and *EXP* of immersion students and their English proficiency (RQ1) and to describe the future programmes of the participants' two schools (RQ2). Each part will contribute to providing a more complete picture of the schools' programmes and immersion education in the Netherlands in general. The RQs were answered with a quantitative experiment (RQ1) and interviews with the programme coordinators (RQ2).

As for RQ1, results showed that different AoAs did not lead to significant differences in English proficiency. AoA alone, not accompanied by more L2 exposure, might not be a significant predictor of L2 proficiency. English as a NL and EXP were significant predictors of respectively, RG and PV, and RV and RG. Particularly the result on EXP could be important for the schools as even a relatively small difference in EXP (between 7-19%) resulted in significant differences on both receptive tasks.

From the interviews, for which the questions were partly based on the results of part one, information was collected on the programmes' set-up. The focus lied mostly on classroom variables as these could not be included in part one. The goal of both schools stays the same the coming years: providing a Dutch/English immersion programme primarily for Dutch children. The "postcode" policy made finding the balance between students with different NLs possible, although school 1 did not focus on this balance. While school 1 will see a large increase in Dutch students (which has advantages and disadvantages), school 2 will not expand as much. In terms of EXP, school 1 is planning to increase EXP whereas for school 2 EXP will stay the same. School 2 is not planning to change much as its students are progressing well.

Since immersion is still new for the schools, both programme coordinators are careful with making significant changes to the programme. This is because it is difficult to predict what the consequences are as there is not much to compare their programmes with in the Netherlands. To conclude, both schools are well underway with their English programmes and will continue with their own vision of immersion.

References

- Baker, C. (2011). *Foundations of bilingual education and bilingualism* (Vol. 79). Tonawanda, NY: Multilingual matters.
- Bishop, D. V. M. (2003). *TROG-2 Test for Reception of Grammar-2*. London: The Psychological Corporation.
- Broek, E. van den, de Graaff, R., Unsworth, S., & van der Zee, V. (2014). *Voorstudie pilot primair tweetalig onderwijs*. Groningen: GION. Retrieved from <https://www.epnuffic.nl/publicaties/vind-een-publicatie/voorstudie-pilot-primair-tweetalig-onderwijs.pdf>
- 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(2), 178-196. doi:10.1080/13670050.2013.866627
- Cammarata, L., & Tedick, D. J. (2012). Balancing content and language in instruction: The experience of immersion teachers. *The Modern Language Journal*, 96(2), 251-269. doi:10.1111/j.1540-4781.2012.01330.x
- Corde, A., Phielix, C., & Krijnen, E. (2012). *Wat weten we over vroeg Engels op de basisschool? Effecten van vroeg vreemdetalenonderwijs op de moedertaal en de leerprestaties in de vreemde taal: review van de onderzoeksliteratuur*. Retrieved from: <https://www.epnuffic.nl/publicaties/vind-een-publicatie/wat-weten-we-over-vroeg-engels-op-de-basisschool.pdf>
- Coyle, D., Hood, P., & Marsh, D. (2010). *CLIL; Content and language integrated learning*. Cambridge: Cambridge University Press.
- de Bot, K. (2014). The effectiveness of early foreign language learning in the Netherlands. *Studies in Second Language Learning and Teaching*, 4(3), 409-18. doi:10.14746/ssllt.2014.4.3.2
- Dunn, L. M., & Dunn, D. M. (2007). *PPVT-4: Peabody picture vocabulary test*. London: Pearson Assessments.
- Edwards, A. (2014). *English in the Netherlands: Functions, forms and attitudes* (Unpublished doctoral dissertation). University of Cambridge, UK.
- Edwards, A. (2016). *English in the Netherlands: Functions, forms and attitudes*. Amsterdam: John Benjamins Publishing Company.

- Graham, S., Courtney, L., Marinis, T., & Tonkyn, A. (2017). Early language learning: The impact of teaching and teacher factors. *Language Learning*, 67(4), 922-958. doi: 10.1111/lang.12251
- Genesee, F. (1994). Integrating Language and Content: Lessons from Immersion. *UC Berkeley: Center for Research on Education, Diversity and Excellence*. Retrieved from <https://escholarship.org/uc/item/61c8k7kh>
- Genesee, F. (2004). What do we know about bilingual education for majority language students? In T.K. Bhatia, & W. Ritchie (Eds.). *Handbook of bilingualism and multiculturalism*, (pp. 547-76). Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.464.3481&rep=rep1&type=pdf>
- Genesee, F., & Lindholm-Leary, K. (2013). Two case studies of content-based language education. *Journal of Immersion and Content-Based Language Education*, 1(1), 3-33. doi:10.1075/jicb.1.1.02gen
- Hickey, T., & de Mejiá, A. M. (Eds.). (2017). *Immersion Education in the Early Years*. New York, NY: Routledge.
- Hummel, K. M. (2014). *Introducing second language acquisition: Perspectives and practices*. Hoboken, NJ: John Wiley & Sons, Inc.
- Jenniskens, T., Leest, B., Wolbers, M., Bruggink, M., Dood, C., & Krikhaar, E. (2017, December). *Zicht op vroeg vreemdetalenonderwijs*. Retrieved from the Rijksoverheid website: <https://www.rijksoverheid.nl/documenten/rapporten/2017/12/01/zicht-op-vroeg-vreemdetalenonderwijs>
- Jenniskens, T., Leest, B., Wolbers, M., Krikhaar, E., Teunissen, C., de Graaff, R., Unsworth, S., & Coppens, K. (2018, April). *Evaluatie pilot Tweektalig Primair Onderwijs: Vervolgmeting schooljaar 2016/17*. Retrieved from: <https://www.rijksoverheid.nl/documenten/rapporten/2018/04/01/evaluatie-pilot-tweektalig-primair-onderwijs>
- Johnson, R. K., & Swain, M. (Eds.). (1997). *Immersion education: International perspectives*. Cambridge: Cambridge University Press.
- Lambelet, A., & Berthele, R. (2015). *Age and foreign language learning in school*. New York, NY: Springer.

- Lindholm-Leary, K., & Genesee, F. (2014). Student outcomes in one-way, two-way, and indigenous language immersion education. *Journal of Immersion and Content-Based Language Education*, 2(2), 165-180. doi:10.1075/jicb.2.2.01lin
- Lyster, R., & Genesee, F. (2013). Immersion education. *The encyclopedia of applied linguistics*. doi:10.1002/9781405198431.wbeal0525
- Muñoz, C. (2008a). Symmetries and asymmetries of age effects in naturalistic and instructed L2 learning. *Applied Linguistics*, 29(4), 578–96. doi:10.1093/applin/amm056
- Nuffic. (2017). *Tweetalig onderwijs (tto)*. Retrieved from the Nuffic website: <https://www.nuffic.nl/voortgezet-onderwijs/tweetalig-onderwijs>
- Pfenninger, S. E. (2014). The misunderstood variable: Age effects as a function of type of instruction. *Studies in Second Language Learning and Teaching*, 4(3), 529-556. doi:10.14746/ssllt.2014.4.3.8
- Pfenninger, S. E., & Singleton, D. (2017). *Beyond age effects in instructional L2 learning: Revisiting the age factor*. Bristol: Multilingual Matters.
- Snijders, J. T. & Snijders-Oomen, N. (1976). *Snijders-Oomen Non-Verbal Intelligence Scale SON 2.5-7*. Groningen: Wolters-Noordhoff.
- Steinlen, A. K., Håkansson, G., Housen, A., & Schelletter, C. (2010). Receptive L2 Grammar Knowledge Development in Bilingual Preschools. In *Bilingual Preschools. Volume 1: Learning and Development* (Ch.3). Retrieved from: <https://files.eric.ed.gov/fulltext/ED572867.pdf>
- Stephen, C., McPake, J., Pollock, I., & McLeod, W. (2016). Early years immersion: Learning from children's playroom experiences. *Journal of Immersion and Content-Based Language Education*, 4(1), 59-85. doi:10.1075/jicb.4.1.03ste
- Sun, H., Steinkrauss, R., Tendeiro, J., & de Bot, K. (2016). Individual differences in very young children's English acquisition in China: Internal and external factors. *Bilingualism: Language and Cognition*, 19(3), 550-566. doi:10.1017/S1366728915000243
- Sun, H., Steinkrauss, R., Wieling, M., & De Bot, K. (2018). Individual differences in very young Chinese children's English vocabulary breadth and semantic depth: internal and external factors. *International Journal of Bilingual Education and Bilingualism*, 21(4), 405-425. doi:10.1080/13670050.2016.1178706

- Sutherland, S. (2012). Native and non-native English teachers in the classroom: A re-examination. *Arab World English Journal*, 3(4), 58-71. Retrieved from <http://westminsterresearch.wmin.ac.uk>
- Swain, M., & Lapkin, S. (2005). The evolving sociopolitical context of immersion education in Canada: Some implications for program development. *International Journal of Applied Linguistics*, 15(2), 169-186. doi:10.1111/j.1473-4192.2005.00086.x
- Thijs, A., Trimbos, B., Tuin, D., Bodde, M., & de Graaff, R. (2011, August). *Engels in het basisonderwijs*. Retrieved from the Stichting Leerplan Ontwikkeling website: <http://www.slo.nl/downloads/2011/engels-in-het-basisonderwijs-vakdossier.pdf/>
- Unsworth, S. (2013). Current issues in multilingual first language acquisition. *Annual Review of Applied Linguistics*, 33, 21–50. doi:10.1017/S0267190513000044
- Unsworth, S., Persson, L., Prins, T., & de Bot, K. (2014). An investigation of factors affecting early foreign language learning in the Netherlands. *Applied Linguistics*, 36(5), 527-548. doi:10.1093/applin/amt052
- Williams, K. T. (2007). *EVT-2: Expressive vocabulary test*. London: Pearson Assessments.

Appendix A

Questions for semi-structured interview

1. Hoeveel nieuwe leerlingen worden er het komende jaar verwacht?
2. Hoeveel van deze zijn Nederlandse, Engelse, and anderstalige kinderen?
3. a. Gaat deze influx de verdeling tussen deze groepen leerlingen veranderen?
b. Zo ja, hoe worden deze groepen geacommodeerd? Worden er verschillende groepen gemaakt of wordt er bijvoorbeeld meer tijd aan/in Engels besteed om de Nederlandse- en anderstaligen dichterbij het niveau van de Engelstaligen te brengen?
4. Komen er nieuwe docenten om de nieuwe leerlingen les te geven of worden de groepen groter?
5. Wat is het niveau van het Engels van de docenten?
6. Wordt er vastgehouden aan de huidige hoeveelheid Engelse lestijd of wordt die veranderd?
7. Komen er veranderingen qua methode of didactiek?
8. Er werd dit jaar eens per week apart Engelse lesgegeven aan de gevorderde leerlingen. Gezien het feit dat de Nederlands- en anderstaligen slechter scoren dan de Engelstaligen, wordt eraan gedacht deze leerlingen ook extra of apart Engels les te geven? (Alleen voor school 1)

Appendix B
Transcript interview 1

Datum: 11-07-2018

Lengte: [40:00]

Interviewer: Jason Gosling

Transcriber: Jason Gosling

[IN] = Initialen geïnterviewde

[JG] = Initialen interviewer

[00:10]

[IN] Dus kleuters, drie dagdelen en dan per groep 45 minuten. Het probleem is: er komt een groep 7 bij, of een 7^e groep bij in Februari.

[JG] En groep 4, 5, 6 zijn er al, of waren er al dit jaar?

[IN] We hadden- we zijn begonnen met 3 groepen 1-2. Hebben er nu 5.

[JG] Oh ik dacht 4.

[IN] We beginnen met 6.

[JG] Oh ok. Toen ik er was waren er nog 4, leek het.

[IN] Ja, dat kan, maar het is zo'n instroom en er belde net iemand over of ze komend jaar de kinderen kan inschrijven want ze wil hier naartoe verhuizen. En we zitten gewoon vol.

[JG] Dus er is al een stop aan de instromingen?

[IN] Ja.

[JG] En het is echt een enorm gebouw.

[IN] Maar halverwege dit jaar hebben we ook al moeten- We zijn al aan het kijken naar een volgend gebouw. Gebouw erbij, dus. Niet normaal. Er komen waarschijnlijk 3, 4 groepen per jaar.

[JG] Oh, echt enorm.

[IN] Hopelijk ben ik dan met pensioen. Ok, dus het groeit als een tierelier, waardoor ook het Engels een beetje verdunt, denk ik. We hadden een heleboel expats aan het begin, maar nu hebben we postcode. Want aan het begin hadden die geen huizen hier, nu hebben we postcodebeleid, dus niet iedereen mag zomaar naar deze school komen.

[JG] Anders kon je ze ook niet weigeren?

[IN] Ja, volgens mij kan je dan niet weigeren.

[01:40]

[03:30]

[IN] Maar dus we groeien als een tierelier. We openen nu met, we hebben nu 1-2 a, b, c, d, e

[JG] E is ook al begonnen? Want toen ik hier aan het testen was-

[IN] Ja het is echt- en 1-2 f, beginnen nu wel met minder kinderen en er komen instromingen door het jaar heen. Dus er zitten geen volle groepen, maar die komen vol aan het eind van het jaar. Omdat kinderen 4 zijn en dan mogen ze naar school komen.

[JG] Wat zou een volle groep, hoeveel-

[IN] 28. Dan hebben we 2 groepen 3; we hadden nu 1 groep 3, 2 groepen 4. We hadden nu ook een groep 4. Een groep 5, en 6 en 7 gaat apart. Dus we openen met 13 groepen. Maar niet allemaal vol, bijv. groep 7 is klein, en 6 en 7 ga ik alsnog samen Engels doen want anders is er niet genoeg tijd. En groep 4 bestaat ook uit 14-15-16 kinderen. Instroom het hele jaar door.

[JG] Dus, en daar zijn ook al leerkrachten voor gevonden?

[IN] Ja, daar is Ralf heel goed in.

[JG] Oh ok.

[IN] Uhm, op de een of andere manier. In Amsterdam hebben ze 300 leraren tekort voor 8000 kinderen.

[JG] Ja dat las ik ook. Misschien heeft een groter gedeelte van de leraren interesse in een tweetalige basisschool.

[IN] Denk het ook. En een nieuwe school, dat is ook wel waar je meer kan betekenen, waar je meer kan groeien.

[JG] Ja.

[IN] Dat is ook interessant.

[JG] En wat voor niveau moesten ze hebben?

[05:30]

[07:00]

[JG] Omdat in het rapport werd er ook een ruwe schatting van de leraren zelf gegeven. Als dat nu ook ongeveer zo is, kan ik dat zeggen. Dat het ongeveer hetzelfde of hoger is.

[IN]

Ja er was een niveau dat je voor tpo moet hebben, ik denk niet dat we dat helemaal hebben, maar daar willen we ze een op een gaan toetsen.

[07:14]

[10:20]

[IN] Kerncompetenties voor leerkrachten in het basisonderwijs. Vvto leerkracht. Dit is vvto leerkracht, B1 niveau. Dus het moet hoger zijn dan b1. Ik denk dat we wel dit zeker hebben, echt iedereen.

[JG] Ok.

[10:56]

[11:23]

[JG] Als ik dan even naar vraag 2 kijk. Dus de hoeveelheid 'expats' wordt kleiner?

[IN] Ik denk dat het minder wordt ja.

[JG] Uh dat geldt dus ook voor Engels? Willen jullie daarop minderen, zodat het een Nederlandse school-

[IN] We willen niet minderen, maar ze moeten wel een bepaald taalniveau hebben om-
Nederlands taalniveau-

[JG] Qua Nederlands ja-

[IN] Dus moeten ze naar de taalschool en dan kunnen ze pas hier naartoe. Dus het is niet zozeer buitenlandse, dat niet Nederlandstalige kinderen niet mogen komen. Ze moeten eerst een bepaald niveau hebben en dan kunnen ze hier naartoe.

[JG] Anders wordt het heel veel om rekening mee te houden, wel fijn als ze allemaal hetzelfde Nederlands niveau hebben-

[IN] Ja, anders ben je Nederlands aan het geven meer dan Engels eigenlijk.

[JG] Ja, dan-

[IN] Je bent ook Nederlands aan het geven, maar je bent Nederlands als tweede taal aan het geven. Terwijl het niet Nederlands als 2^e taal de bedoeling is, de bedoeling is engels als 2^e taal geven.

[JG] Ok. Er komen dus er 3 tot 4 groepen bij, wordt daar de verdeling van Nederlandse, Engelstalige en anderstalige kinderen-

[IN] Ik denk niet dat ze daar rekening mee hebben gehouden.

[12:45]

[13:40]

[IN] Er zullen zo'n 300 leerlingen in 14 groepen zijn. Dat is misschien interessant om te weten. We groeien komende jaar van de huidige 150 leerlingen in 8 groepen naar zo'n 300 leerlingen in 14 groepen.

[13:50]

[14:40]

[IN] Ik denk niet dat er gekeken is naar Engels, bewust.

[JG] Ok, dus de groepen worden gewoon willekeurig ingedeeld?

[IN] Dat was misschien wel beter geweest.

[14:52]

[16:00]

[JG] Is er een plan met hoeveelheid engels? Wordt dat veranderd?

[IN] In groep 3 gaat het even een beetje omlaag en dan in groep 4 is het de bedoeling in de middag Engels, dus bovenbouw de hele middag Engels is. Dat is dan rond de anderhalf uur, uur en drie kwartier in Engels en dan de zaakvakken in IPC.

[JG] Zaakvakken anderhalf uur tot een uur en drie kwartier in Engels per dag-

[IN] Maar, daar tussen ook dus raindrops of energizers. En we beginnen met 'good morning' en in de 'snack time' ook: 'go get your bags', 'go get your bags', 'sit down', 'what are you eating?'.
.

[JG] Dus die 'raindrops' komen er nog bij?

[IN] Ja, maar we willen- Je mag tussen de 30 en 50% zitten en wij gaan voor de 25/30 in de toekomst.

[17:02]

[19:45]

[IN] 25-30% is streeftijd voor bovenbouw, anderhalf uur tot 1 en 45 per dag.

[JG] Per dag.

[IN] Ja. En in 1 en 2, drie dagdelen van 45 minuten. Momenteel he, het streven is daar ook gewoon 30%. En in 1-2 moet het Engels geïntegreerd zijn en in groep 3 ietsjes minder Engels, zit je waarschijnlijk rond de 20%, in verband met lezen.

[JG] Ok.

[IN] Dat is de bedoeling, maar dat is niet wat we nu gaan doen. Daar gaan we naartoe in een paar jaar tijd. En nu wordt nog even vastgehouden gewoon aan de huidige hoeveelheid. Omdat er andere dingen zoveel veranderen, nieuwe leerlingen-

[JG] En omdat ook niet iedereen het gelijk kan, alles in het Engels-

[IN] En nieuwe docenten ja, en nieuwe onzekerheden. En nog moeten wennen aan heel veel dingen.

[20:45]

[21:55]

[IN] We gaan in de klas samen 'co-teachen', samen les geven. Dan kun je beter differentiëren. Dan ga ik voor de meer 'advanced' leerlingen.

[JG] Ok.

[IN] Terwijl, en dit is dan 2 lessen per week. Dat is 2x 45 min, anderhalf uur per week IPC in het Engels.

[JG] En dan is het ook doeltaal is voertaal.

[IN] Ja. Maar in 4, 5, 6, 7, moeten ze eerst 'pre teachen' in het Nederlands. Dus voor thema vakantie moeten ze het woordje zee, strandbal enzovoorts- Ze moeten eerst het Nederlands een beetje aanleren. Dan kunnen de 'advanced' wat anders doen.

[JG] Ok.

[IN] En dan gaan we samen 1 les geven. En dan nog een les in die week is ook Engels en IPC, maar ik ben er dan niet bij. Dan moet het verwerkt worden, waar je les in geeft. Dus een verwerking les. Een instructie les en een verwerking les. En de 'advanced' gaan dan wat hoger niveau werk doen en die kunnen dan ook hun leerkracht vragen om hulp.

[JG] En de 'advanced' en minder 'advanced' leerlingen krijgen nog steeds wel even veel engels?

[IN] Ja.

[JG] Maar dan van verschillende docenten?

[IN] Ja en nee. Maar dus, ik geef 4a en 4b les met de leerkracht en zij moeten in die week nog een keer een les. Zeg dat dit aardrijkskunde is, dan moeten ze een verwerking les doen in het Engels. Maar wij maken samen materiaal voor die tweede les. Uiteindelijk is het 'pre teachen' en dat moet minstens drie kwartier. En dat is voor de beginners, dan kunnen de gevorderde helpen natuurlijk. Dan heb je een 'co teaching' les.

[24:50]

[32:10]

[IN] Weetje, in volgend jaar groep 5, dat is mijn huidige groep 4- en ook volgend jaar in groep 6 en volgend jaar groep 7 zijn de meeste 'expats'- de helft van mijn groep van 16 kinderen zijn Engelstalig.

[JG] Ok.

[IN] Dus dat niveau is redelijk hoog. Het niveauverschil is zo groot tussen die 'advanced' en die beginners in die hogere groepen dat het gewoon eigenlijk bijna niet te doen is om die beginners op het niveau te krijgen dat-

[JG] Dat ze samen les krijgen.

[IN] Ja.

[32:55]

[32:20]

[IN] Maar vanwege het postcodebeleid krijgen we waarschijnlijk andere bevolking op school, dus Nederlandstalig en niet zoveel expats meer.

[JG] Dat maakt het ook makkelijker voor jullie eigenlijk-

[IN] Maar-

[JG] Qua differentiatie tussen moedertalen enzovoorts.

[IN] Ik geloof persoonlijk dat des te meer Engelstaligen er op school zijn, des te makkelijker het is om het niveau omhoog te krijgen.

[JG] Maar inderdaad wat je net zei over je groep 5: bijv. als het verschil te groot is, is het gewoon heel moeilijk om ze samen les te geven.

[IN] Klopt.

[33:54]

[37:15]

[IN] Weet je wat ik denk: het onderzoek heeft bevestigd wat wij al vermoeden en wat ook eigenlijk logisch is- Dat er een verschil is doordat er ‘native speakers’ zijn. Maar kinderen die het Nederlands nog niet beheersen, gaan ook niet eerder Engels leren dan Nederlands. Want ze krijgen veel meer Nederlands.

[JG] Sorry, kinderen die nog geen Nederlands hebben geleerd-

[IN] Ja dus de internationals, dus niet de ‘native speakers’-

[JG] Oh, omdat die eerst bezig zijn met NL leren-

[IN] Want die krijgen ook veel meer Nederlands. Die worden ondergedompeld in Nederland, dus zij krijgen minder Engels dan Nederlands.

[37:54]

[38:56]

[IN] Het blijkt dus, als jij tweetalig opgroeit, je heel makkelijk een 3^e taal erbij kan leren. Omdat je de structuren van talen kent. Kinderen die Nederlands nog niet beheersen, gaan ook niet eerder Engels leren.

[JG] Als je tweetalig opgevoed bent ja, maar niet elke van die internationale zijn toch tweetalig opgevoed? Sommigen hebben maar een moedertaal en die leren er nu twee talen bij eigenlijk toch? Nederlands en Engels?

[IN] Ja maar Nederlands is meer, - Nederlands is hun tweede taal en Engels is een beetje erbij. Dus eigenlijk komen ze naar Nederland en dan worden ze bijv. Grieks en Nederlands opgevoed en dan krijgen ze een beetje Engels op school erbij. Want het is niet 50/50 en hier omheen is nog veel meer Nederlands. Dus dat kleine beetje Engels is niks vergeleken met gewoon in Nederland aanwezig zijn. Dus dat Grieks en Nederlands is al tweetalig opgevoed worden. Dus die internationals worden tweetalig opgevoed.

[40:00]

Appendix C
Transcript interview 2

Datum: 12-07-2018

Lengte: [21:28]

Interviewer: Jason Gosling

Transcriber: Jason Gosling

[IN] = Initialen geïnterviewde

[JG] = Initialen interviewer

[01:18]

[JG] Hoeveel nieuwe leerlingen worden er verwacht?

[IN] We hebben nu ongeveer 40 leerlingen en het schooljaar start met ongeveer 54 leerlingen. Dus die komen er meteen na de vakantie bij, maar daarna zit er weer groei in. Dan groeien we nog met 20 leerlingen ofzo.

[JG] En het wordt dus volgend jaar groep 1 tot en met 4?

[IN] Ja.

[JG] En komen er grotere groepen, of hoe groot zijn de groepen nu?

[IN] Ja, we hebben nu 1 groep 1-2 maar die is heel groot. Dus daar hadden we 2 leerkrachten op staan. En een groep 3 die heel klein was met maar 7 leerlingen. Volgend schooljaar zijn er 2 groepen 1-2, die starten ongeveer met 17-18 kinderen allebei. En groep 3-4 wordt 20 leerlingen.

[JG] Dus dat zijn nog redelijk kleine klassen voor-

[IN] 20 leerlingen? Ja. Groep 1-2 groeit daarna wel weer verder. Dat wordt wel weer aan-

[JG] Dat komt gewoon tijdens het jaar? Of op een bepaald moment, dat ze in kunnen stromen?

[IN] Ja, dat is na hun vierde verjaardag, dat ze kunnen instromen. Dus dat is echt het hele jaar door.

[JG] Is het gebaseerd op moedertaal? Is er een limiet gezet op de hoeveelheid Engelstalige kinderen die hier kunnen komen?

[IN] Bij groep 3-4 wel en daar ben ik ook heel secuur in om die (kinderen) aan te nemen of niet. Groep 1-2- vooral de groep 1'ers niet. Dat maakt niet zo heel veel uit.

[JG] En bij groep 4 dus wel. Is dat om een bepaalde balans te hebben tussen- En welke balans zoeken jullie dan?- Will jullie eigenlijk meer een school voor Nederlandstalige kinderen die dus Engels als tweede taal willen leren of willen jullie gewoon zeggen: "het maakt niet uit welke moedertaal en-"

[IN] Toen we deze school startte waren er met name Nederlandse ouders die voor tweetalig onderwijs kozen. Maar toen merkten we dat er een hele grote vraag was ook voor anderstaligen die naar deze school wilden. En daarna kwamen ook de Engelstalige erbij. Maar dat heb ik niet voorzien toen ik deze school startte. Ik wist wel dat het ging aantrekken, maar ik wist niet hoe en dan is dat op deze school nog op een hele moeilijke positie met de buurschool hiernaast in hetzelfde gebouw-

[JG] Is er dan ook een limiet- er is dus een limiet aan ruimte waarschijnlijk dus?

[IN] Nee, want die andere school wordt kleiner en wij groeien erin, dus dat gaat op z'n natuurlijke beloop. Maar dat heeft wel gezorgd dat in het begin van de school ik ook niet veel PR kon hebben, met name omdat het heel gevoelig lag dat ik hier een nieuwe school begon. Dus nu komen er van overal vandaan mensen die hier op school willen, maar ik heb gezegd: "We

concentreren ons alleen maar op mensen die in Utrecht-Oost wonen, dus die worden nu ook alleen maar aangenomen.”

[JG] Soort postcode-beleid?

[IN] Ja. Want anders dan ben ik een soort halve internationale school, terwijl ik daar helemaal niet de middelen voor heb. Dus als jij zegt van: “Waar selecteer je op als er kinderen zijnstromen bijvoorbeeld in groep 4 of in groep 3.” Dan kijk ik heel erg naar de behoefte van de groep, van wat die nodig hebben, en wat het kind heeft en als ik denk van: “Nou, deze leerkracht kan dat aan”, dan is dat welkom en anders niet. Als ik nu een Chinees kind krijg wat alleen maar Chinees praat en nog geen Engels en geen Nederlands, dan is het niet bij ons op z’n plek.

[JG] De leerlingen moet of Nederlands of Engels spreken, of zijn er andere- ?

[IN] Nee dat is- dat is bij groep 1-2 zo, vooral bij de groep 1’ers hebben we gemerkt dat die de Nederlandse taal- als hun Engelse taalvaardigheid al heel hoog is, gaat dat in het Nederlands ook zo. Maar voor groep 3 en 4 wordt het heel erg gecompliceerd als je alleen maar anderstaligen aanneemt die je nog Nederlands moet leren, terwijl eigenlijk het doel van de school ook is dat je Engels leert.

[JG] Dus die zijn dan meer bezig met Nederlands?

[IN] Ja.

[JG] Eigenlijk omdat dat natuurlijk de taal hier van het land is.

[IN] Ja, dus dat je gewoon een Nederlandse school bent met je Nederlandse curriculum wat je moet draaien.

[JG] Er is dus geen bepaald percentage van groepen moedertalen? Van zoveel procent moet Nederlands zijn. Is er een bepaalde balans die je zoekt of ligt dat dus echt aan het kind zelf? Of die erbij past?

[IN] Ja, want achter elke- elk kind zit wel een heel verhaal, want sommigen die zijn al wel Nederlandstalig opgevoed maar zijn naar een Engelse school geweest, dus dan kunnen ze het wel verstaan maar kunnen ze het alleen nog niet schrijven. Die pikken het sneller op dan iemand die nog helemaal geen Engels of Nederlands kan. Maar daar zijn in Utrecht ook afspraken over he, dat je dan eerst naar de taalschool gaat.

[JG] Ja dat-

[IN] Dat vertelde Ralf denk ik ook, of niet? Heb je met Ralf niet gesproken?

[JG] Nee, ik heb Nicole gesproken. Die vertelde zoiets inderdaad.

[IN] Maar juist de balans van- van Nederlandstalige kinderen als de eerste taal en Engels als eerste taal, als je die balans zeg maar goed houdt, dan hebben allebei de groepen daar ontzettend voordeel van.

[JG] Dus jullie proberen ze wel bij elkaar te zetten dat ze van elkaar kunnen leren- ik dacht als het niveauverschil te groot is, zal maar zeggen het Nederlands van de Engelsen heel laag ligt, dat ze dan niet echt iets- dat ze dan eerst catch-up moeten spelen. Maar dat gaat dus beter met andere kinderen erbij, de kinderen van de andere groepen?

[IN] Nou je hebt wel je kleine- je kleine kringen, zeg maar, waarin je focust of op het Nederlands, wat die kinderen nodig hebben, of op het Engels, wat andere kinderen nodig hebben. Maar in een grote groep leren ze zoveel van elkaar dat het alleen maar juist elkaar versterkt.

[JG] Het klinkt allemaal alsof het dus al goed gaat met de verschillende groepen en de verschillende talen-

[IN] Ja, maar je zit ook met aannamebeleid, dus je kan ook niet zomaar mensen afwijzen van: Ja, jij- jouw Engelse taal is beter dus dan kan je-. Snap je, daar zit wel- er zit ook een- vandaar dat ik elk geval individueel kijk van: wat is het beste voor het kind? Wat zijn de onderwijsbehoeften?

[JG] Zijn er plannen voor de hoeveelheid Engelse lestijd? Omdat dat een groot ding is binnen tweetalig onderwijs, van hoeveel Engels wordt er gegeven per week bijvoorbeeld?

[IN] Dat is bij de kleuters- daar zijn het ongeveer vier activiteiten op een dag die in het Engels zijn. Dus dan zou je kunnen denken aan ongeveer een uur. En bij groep 3-4 hebben we echt gezegd: "In de ochtend doen we alleen maar Nederlands en dan is het 's middags in het Engels."

[JG] Een middag, hoeveel uur is dat?

[IN] Een middag is, even nadenken-

[JG] Half 1 tot half 3?

[IN] Nee. De middag is een uur en een kwartier. Dan kun je nog wel dat ze zeg maar 's ochtends de- tijdens het eten en drinken Engels hebben, maar dat is geen lestijd.

[JG] Zijn daar veranderingen in gemaakt ten opzichte van dit jaar en volgend jaar? Bijvoorbeeld op basis van het KSU rapport, of houden jullie het gewoon hetzelfde?

[IN] We houden hetzelfde.

[JG] Want het blijkt gewoon dus dat dat goed gaat, er geen probleem mee is?

[IN] Na ja, ik vind dat je in de basis gewoon ook dat je het Nederlands in orde moet hebben. Daar zijn we dan het meeste mee bezig.

[JG] En in groep 4, dan wordt ook lezen en schrijven toch geleerd, dus dan wordt waarschijnlijk nog meer tijd aan Nederlands besteed, toch?

[IN] Nee. We gaan juist in groep 4 al beginnen met lezen en schrijven van Engels.

[JG] Hebben ze dat in groep 3- begint lezen en schrijven begint dan al in Nederlands?

[IN] Ja.

[JG] Oké, ik dacht 4.

[IN] Nee, 3. Dat is de groep waar je leert lezen en schrijven in het Nederlands.

[JG] Oké, dan herinner ik me dat verkeerd denk ik. Ik denk dat het voor mij ook wel hetzelfde was. Maar de groepen- de groepen blijven ongeveer hetzelfde, maar 1-2 wordt wel groter. En er is een 3 en 4. Is dat dan de leerkracht- ?

[IN] De leerkracht van groep 3 die gaat nu groep 3-4 doen.

[JG] Oh, het is 3-4 bij elkaar?

[IN] Ja.

[JG] Oh ja, want dat ging 8 naar 17-18-

[IN] 20.

[10:28]

[14:06]

[JG] Dat is vooral op deze school denk ik, want Onder de Bogen is er iets meer- of iets minder variatie in thuistalen, want daar is een groter gedeelte Nederlands en hier zijn het echt veel Engelse internationals. Zeker in groep 3.

[GG] Ja klopt in groep 3. Als we volgend jaar groep 3 hebben, dan zit daar juist weer een grote groep met Nederlands-Nederlands in, dus ik zou juist ook voor volgend schooljaar die groep 3, zeg maar, ook willen testen. Dat zou voor mij dan- en die zijn ook bij ons gestart op school, dus dan kun je echt zeggen: “Nou, hoe hun Engels is op dat moment,” dan kan ik echt zeggen: “Oké, dit is—hebben wij goed gedaan, of dit hebben we minder goed gedaan.”

[JG] Dus daar zijn jullie al plannen mee aan het maken om ook de vooruitgang van de kinderen die dit jaar getest zijn, om die bijvoorbeeld over 2 jaar weer te testen?

[GG] Nou, ik zou het liefst de groep 3 volgend schooljaar willen testen.

[JG] Oké, zodat die ook een startmeting hebben, zal ik maar zeggen?

[GG] Ja, want dat zijn echt de kinderen waarvan wij kunnen zeggen: “Die hebben wij Engelstalig onderwijs geboden en dat hebben we wel of niet goed gedaan.”

[JG] Maar die hebben ook nog 0 Engels gehad.

[GG] Nee, die zijn bij ons bij groep 1 binnengekomen en die zijn dus nu wel getest, maar net zoals in het foto onderzoek, zeg maar-

[JG] Ik dacht even dat we het hadden over kinderen die volgend jaar in groep 3 gaan zitten, die van buiten kwamen.

[GG] Nee-

[JG] Nee, dit zijn degenen die nu in 1-2- of nu in 2 zitten?

[GG] Ja.

[JG] Oké. Dus dan is dat geen startmeting maar dan is het een follow-up.

[GG] Ja, precies.

[JG] Daar zat ik even verkeerd mee; vandaar dat ik zei startmeting, wat niet klopt.

Nu zijn we al redelijk aan het einde; zijn er ook curriculumveranderingen gedaan? Bijvoorbeeld op Onder de Bogen hebben- werken ze met IPC, werken jullie daar ook mee trouwens?

[GG] Ja.

[JG] En dat- houdt dat ook in dat- dat zijn de meer zakelijkere vakken zoals geschiedenis en aardrijkskunde. Die worden hier dan ook in het Engels gegeven?

[GG] Ja, de hele middag is dan IPC en in het Engels.

[JG] Wordt er gemerkt dat de Nederlandstalige kinderen gewoon die nieuwe stof makkelijk opnemen in het Engels of wordt eerst die stof een beetje voorbereid in het Nederlands?

[GG] Nee, die kinderen kunnen al dat allemaal aan. Maar dat komt ook omdat vooral de kinderen van groep 3, zeg maar- groep 2-3 was dan nu steeds met IPC samen en die kinderen kunnen dat allemaal volgen. Die hebben bij ons de start al meegemaakt met Engels, dus die zitten nu al op een heel hoog niveau.

[JG] Ah oké, dat is makkelijk. Dus dat blijft gewoon hetzelfde, gebruik van IPC en-

[GG] Ja, we gaan wel bezig met Early Bird om de leerlijnen van groep 4 tot en met 8 verder uit te werken.

[JG] Want dat is natuurlijk een nieuw gebied?

[GG] Ja.

[JG] Zijn er ook telkens dezelfde dagdelen qua groepen die worden- Nederlands, nee- groep 1-2 die had dus 4 kwartiertjes eigenlijk.

[GG] Ja, 4 keer 20 minuten zeg maar.

[JG] Maar werken die ook met IPC, is dat ook voor groep 1?

[GG] Ja, groep 1 en 2 ook. Alleen het IPC is dan anders ingedeeld, omdat je bij de kleuters in de activiteiten werkt, die zitten in de kring, je gaat spelen-werken. En spelen-werken is dan ingericht in IPC, dus dan heb je overal werkplaatsen-

[JG] Die telkens in de hoekjes?

[GG] Ja, dat noemen ze-

[JG] Kinderen hebben dan toch vrije keus bij van wat ze-

[GG] Soms wel, soms niet, want groep 2'ers moeten soms ook dingen doen die moeten.

[17:57]