

# Second language development of newly arrived migrant kindergarteners

Exploring educational settings and pedagogical practices in the Netherlands



Frederike C. Groothoff



# **Second language development of newly arrived migrant kindergarteners**

Exploring educational settings and pedagogical  
practices in the Netherlands



***Girl with wooden tulips***

*A bouquet of wooden tulips on the front of a dissertation about second language development. Why? For me tulips are a beautiful illustration of second language development. First there is this tulip bulb, it is present, but it looks passive and you do not know what to expect. Then the bulb bursts open and something grows out of it. It is only after long patience, care, and the right circumstances that you can see what a beautiful color was hidden all the time. Furthermore, tulips are the best example, because they seem to be so typically Dutch, almost nobody remembers they were originally from Turkey...*

ISBN: 978-94-6332-595-0

NUR: 616

Copyright © 2020: Frederike Christine Groothoff. All rights reserved.

# **Second language development of newly arrived migrant kindergarteners**

Exploring educational settings and pedagogical  
practices in the Netherlands.

## **Tweedetaalverwerving van nieuwkomers in de kleuterleeftijd**

Een verkenning van de leeromgeving en pedagogische en  
didactische praktijken in Nederland.

(met een samenvatting in het Nederlands)

## **Proefschrift**

ter verkrijging van de graad van doctor  
aan de Universiteit Utrecht  
op gezag van de rector magnificus, prof.dr. H.R.B.M. Kummeling,  
ingevolge het besluit van het college voor promoties  
in het openbaar te verdedigen op

donderdag 30 januari 2020 des middags te 4.15 uur

door

**Frederike Christine Groothoff**

geboren op 18 juli 1982  
te Houten

**Promotoren:** Prof. dr. H.C.J. de Graaff  
Prof. dr. P.P.M. Leseman

**Copromotor:** Dr. E.M.M. Le Pichon-Vorstman

This research was supported by the Netherlands Organisation for Scientific Research (NWO) under project number: 023.003.159.

This research was co-financed with a grant by the TOEFL Small Grants for Doctoral Research in Second or Foreign Language Assessment program.

This thesis is dedicated to...

... all the children, teachers, and especially my colleagues. You all inspired me to dig into  
this subject...

... my almost 100-year-old grandfather. Thanks for your everlasting interest in my research  
and your example of always being/staying curious. It is a pity you never saw more than the  
title page of my dissertation. You were the personification of a life-long-learning. How I  
love to start reading such fat books as you daily did...

... Ralph, now we can finally go on a holiday without me being afraid of not working.  
Love you to the moon and back...



---

## Acknowledgements

---

Like newly arrived migrant pupils, I needed interaction with many different people in order to develop my skills as a PhD-student. I would like to start with mentioning Sanne Kuijper and Gisi Cannizzaro: You showed me back in 2007 at the University of Groningen that science is also for ‘cool chicks’. Thanks for the inspiration and the continuing friendship. Gisi, besides my inspiration you were also my editor and *paranimf*. You definitely made the world a little better by teaching me the Oxford comma!

The idea of doing a PhD could only be executed next to my job as a primary school teacher because of Jan Bak, Karlijn de Jonge, and Annemieke Van der Krift. Thank you for giving me the opportunity to apply for the *Lerarenpromotiebeurs*. And thank you Jenny Olsthoorn and Esther van Zoest for the continuing support. All former colleagues at *Taalschool Utrecht* need to be thanked for their patience with me during all these years of me being only their part-time colleague. A special thanks has to go to Maud Bringmann, Floris van der Berg, Fransje van Dongen, Esther Hoving, and Inge Vernooij. I loved being your duo!

After the permission of my principals at *Taalschool Utrecht* I needed a professor at a university to be my (co)promotor. Thank you Sharon Unsworth for introducing me to Rick de Graaff. Rick, thank you for accepting me as a *buitenpromovenda* and for all your encouragements these last six years. Thank you Emmanuelle le Pichon-Vorstman and Paul Leseman for all your critiques and support. Emmanuelle, it was a privilege to be part of your family that week in Toronto, merci beaucoup! At Utrecht University there are many more people that need to be mentioned: Huub van der Bergh, thank you for your precious time, I loved taking the seat behind your desk! Thank you Kirsten Schutter, Frank Wijnen, Maaïke Schoorlemmer, Elma Blom, Tessel Boerma, Lotte Henrichs, Marike Kempen and all the people from the Uil-OTS, especially from the Uil-OTS labs and the Language and Education group. Thank you Suzanne Bogaerds-Hazenberg and Jiska Malschaert for being my assistants.

During my PhD I got the chance to meet a lot of interesting people at conferences and meetings in the Netherlands and abroad. Thanks to the people from the NWO, ORD, NRO, LOWAN, RUTU Foundation, EDINA, SLO, ECER, YLL, DRONGO, HOLM, SIRIUS, and the European Centre for Modern Languages for organizing such great meetings. It was a privilege to meet Natalia Gagarina, Ute Bohnacker, Andrea Young, Latisha Mary, Josefin Lindgren, Johanne Paradis, Elena Nicoladis, Anna Kirova, Annina Hessel, and so many others in Oxford, Graz, Rotterdam, Edmonton, Toronto, Dublin, Berlin, Groningen, Amsterdam, Utrecht, and Bolzano.

I also want to thank my new colleagues of *Onderwijsgroep Amstelland* and *Amstelwijjs* for supporting me these last months. It is such a pleasure working with you Ellen Goossens! I was able to round up everything for my PhD because of your support.

My family (Groothoff, Patist, and Vermeulen) also needs a big thank you. Thank you for helping me complete this journey. I promise I will make up for all the lost daughter/sister/auntie time... Hanne, thank you for your time and support as *paranimf*! Papa, without you my references would never have been so complete, thank you for still being my *bibliothecaris*. Dear Ralph, even though I was not always the best girlfriend, you never stopped believing in me! You encouraged me and kept reminding me that life is more than a PhD-project. You kept me in the balance.

I would like to end these acknowledgements with saying thank you to all the teachers and multilingual children from primary schools all over the Netherlands that participated in my project. You were a great inspiration and without you there could have never been a project like this:

**Heel erg bedankt!**

**Dziękuję! Thank you!**

**Macha Danki! Благодарность! Mahadsanid! 謝謝你**

**شكرا! Благодаря! Obrigada! Obrigado! ありがとう**

**Grazie! Paldies! Merci! Terima kasih! Salamat!**

**Muļtūmīri! Ծնողնի Kőszőnőm!**

**감사 ;Gracias!**

**Asante**

---

## Table of contents

---

<b>Acknowledgements.....</b>	<b>vii</b>
<b>Clarification of Terms.....</b>	<b>xv</b>
<b>Abbreviations .....</b>	<b>xix</b>
<b>1 Introducing the context for research with newly arrived migrant kindergarteners .</b>	<b>21</b>
1.1 The inspiration for this research.....	22
1.2 Introduction.....	22
1.3 Newly arrived migrant pupils .....	24
1.4 The educational settings in the Netherlands.....	25
1.4.1 The educational setting for newly arrived migrants in the Netherlands .....	25
1.4.3 Core objectives for education .....	27
1.5 The focus, aim, and general research question of this dissertation .....	28
1.6 Outline of this dissertation .....	29
<b>2 Theoretical perspectives .....</b>	<b>31</b>
2.1 Introduction.....	32
2.2 Ecological linguistics in education .....	32
2.3 Communicative competence .....	34
2.4 Second language learning .....	37
2.4.1 Dual language learners, bilinguals, and second language learners .....	38
2.4.2 The importance of learning the second language .....	38
2.4.3 Individual differences .....	39
2.5 Receptive vocabulary development .....	41
2.5.1 Receptive vocabulary development of second language learners.....	41
2.5.2 Receptive vocabulary development of newly arrived migrant pupils in Dutch ..	42
2.6 Narrative ability development.....	43
2.6.1 Narrative ability development .....	44

2.6.2 Macrostructure and microstructure .....	45
2.6.3 Narrative ability development of second language learners .....	47
2.7 Relation school learning environment and social and cognitive outcomes .....	48
2.7.1 How to measure educational quality? .....	48
2.7.2 Focus on teacher behavior .....	50
2.7.3 Focus on focal pupils experiences .....	51
2.7.4 Pedagogical practices in the Netherlands .....	52
<b>3 Present study: extended research questions and methodology .....</b>	<b>57</b>
3.1 Introduction .....	58
3.2 Extended research questions .....	58
3.3 Methodological considerations .....	60
3.3.1 Second language assessment .....	60
3.3.2 Research design and data collection .....	62
3.3.3 Further methodological considerations .....	63
3.4 Participating schools, teachers, and pupils .....	64
3.4.1 Participating schools and teachers .....	64
3.4.2 Demographic characteristics of the participating pupils .....	66
<b>4 Development of receptive vocabulary of newly arrived migrant kindergarteners ....</b>	<b>71</b>
4.1 Introduction .....	72
4.2 Receptive vocabulary .....	73
4.3 Method .....	73
4.3.1 Participants .....	73
4.3.2 Measurement .....	76
4.3.3 Procedure .....	76
4.3.4 Analyses .....	77
4.4 Results .....	79
4.5 Summary and conclusion .....	83

<b>5 Development of narrative ability of newly arrived migrant kindergarteners.....</b>	<b>87</b>
5.1 Introduction.....	88
5.2 Narrative ability development.....	89
5.2.1 Microstructure and macrostructure.....	89
5.2.2 Measuring narrative ability with the MAIN .....	89
5.3 Method.....	92
5.3.1 Participants .....	92
5.3.2 Measurement .....	94
5.3.3 Procedure.....	96
5.3.4 Analyses .....	97
5.4 Results.....	105
5.4.1 Microstructure .....	106
5.4.2 Macrostructure.....	117
5.5 Summary and conclusion.....	126
<b>6 Pedagogical practices: focus on teacher behavior .....</b>	<b>129</b>
6.1 Introduction.....	130
6.2 Pedagogical practices: focus on teacher behavior.....	130
6.2.1 Previous research with the CLASS.....	130
6.2.2 Dutch studies using the CLASS .....	132
6.3 Method.....	133
6.3.1 Participants .....	133
6.3.2 Measurement .....	134
6.3.3 Procedure.....	136
6.3.4 Analyses .....	137
6.4 Results.....	137
6.4.1 Quantitative analysis of the CLASS.....	137
6.4.2 Qualitative analyses of the CLASS .....	139
6.5 Summary and conclusion.....	140
<b>7 Pedagogical practices: focus on focal pupils' experiences .....</b>	<b>145</b>

7.1 Introduction.....	146
7.2 Pedagogical practices: focus on focal pupils' experiences .....	146
7.2.1 Previous studies using snapshot.....	147
7.2.2 Dutch studies using snapshot.....	147
7.3 Method.....	148
7.3.1 Participants .....	148
7.3.2 Measurement .....	149
7.3.3 Procedure.....	150
7.3.4 Analysis .....	153
7.4 Results.....	153
7.4.1 Quantitative analysis of the snapshot data.....	153
7.4.2 Additional field notes .....	159
7.5 Summary and conclusion.....	161
<b>8 The influence of the school learning environment on the second language development of newly arrived migrant kindergarteners .....</b>	<b>167</b>
8.1 Introduction.....	168
8.2 The influence of the school learning environment on receptive vocabulary development.....	169
8.2.1 Pedagogical practices: focusing on teacher behavior .....	170
8.2.2 Pedagogical practices: focusing on focal pupil experiences .....	173
8.3 The influence of the school learning environment on narrative ability development .....	175
8.3.1 Pedagogical practices: focusing on teacher behavior .....	176
8.3.2 Pedagogical practices: focusing on focal pupils' experiences .....	184
8.4 Summary and conclusion.....	201
8.4.1 The influence of the school learning environment: focusing on teacher behavior .....	203
8.4.2 The influence of the school learning environment: focusing on focal pupils' experiences .....	203

<b>9 General discussion and conclusions .....</b>	<b>207</b>
9.1 Introduction.....	208
9.2 Overview of the results .....	208
9.2.1 Receptive vocabulary development .....	208
9.2.2 Narrative ability development. ....	208
9.2.3 The school learning environment .....	209
9.2.4 Linking characteristics of the learning environment to language development	211
9.2.5 Summary of the study.....	212
9.3 General discussion .....	213
9.3.1 Second language development .....	214
9.3.2 The school learning environment .....	215
9.3.3 The influence of the school learning environment on second language development .....	216
9.4 Implications for educational practice .....	217
9.5 Limitations and future research.....	219
9.5.1 Participants .....	219
9.5.2 Instruments and procedure.....	220
9.5.3 Language development and communicative competence.....	222
9.6 Concluding remarks .....	223
<b>References .....</b>	<b>225</b>
<b>Appendix 1. Table with parameters settings Chapter 4.....</b>	<b>251</b>
<b>Appendix 2. Tables with parameters settings Chapter 5 .....</b>	<b>253</b>
<b>Appendix 3. Description of the CLASS .....</b>	<b>259</b>
<b>Appendix 4. Boxplot CLASS .....</b>	<b>261</b>
<b>Appendix 5. Snapshot items .....</b>	<b>263</b>
<b>Appendix 6. Tables with parameters settings Chapter 8 .....</b>	<b>267</b>
<b>Nederlandse samenvatting.....</b>	<b>279</b>
<b>Curriculum Vitae .....</b>	<b>287</b>



---

## Clarification of Terms

---

### **Communicative competence:**

Communicative competence is a central concept in language teaching: this concept puts the learner at the heart of his or her learning. It concerns the learner's ability to understand the content of the conversation and to express himself/herself in authentic communication situations, and in the context of this thesis, at school. Communicative competence includes four domains: linguistic, strategic, discursive and sociocultural (e.g., Savignon, 1976; Canale, 1983).

### **Discourse competence:**

Discourse competence is part of communicative competence and implies combining grammatical forms and meanings to realize a cohesive and coherent piece of spoken or written text in different genres. For example, in the context of this thesis, learners must be able to tell a coherent story based on a picture sequence the interlocutor cannot see.

### **DL2:**

Unlike foreign language learning, second language learning refers to learning the dominant language of the environment. Dutch as second language means that the Dutch language is learned on top of the languages that the child already knows.

### **DL2-school:**

A school or class which is temporally (in general one to two years) attended by language learners with a focus on learning the second language. In the context of this thesis, Dutch is the only language of instruction.

### **Guiraud Index Score:**

A measure of lexical diversity. It is calculated by dividing the number of types (different words) by the square root of the total of tokens (total number of words) of a story.

### **Home language(s):**

The language(s) acquired, and often still spoken, in the family context.

### **Internal State Terms:**

Internal state terms provide important information about the awareness of a narrator and his/her understanding of the story characters' mental states: what they know, what they value, and what they believe e.g., see, hear, happy, think, hungry.

### **Kindergarten:**

The first two years of Dutch primary education, called group 1 and 2. Pupils in Dutch kindergarten are generally four to six years old. In the Netherlands, since 1985, education begins with these two years of kindergarten integrated into primary education. In most other countries these two years are considered pre-primary.

**Kindergartener:**

A pupil attending kindergarten.

**L1:**

Languages acquired in the family context. In this thesis L1 and home language are used interchangeably.

**Lexical diversity:**

The range of words used in oral or written language. This is measured in the context of this thesis by calculating the Number of Different Words and the Guiraud Index Score.

**Linguistic Competence:**

Linguistic competence is part of communicative competence and focusses on the skills and knowledge required to accurately express and understand the literal meaning of utterances. It includes features and grammatical rules of the language.

**Multilingual Assessment Instrument for Narratives:**

An assessment tool to measure narrative ability, developed by Gagarina et al. (2012).

**Mainstream school:**

A primary school in the Netherlands without specific classes for second language learners.

**Measure of Lexical Richness:**

This measurement takes into account the frequency bands of the words used. A story has a high richness if many low-frequent words are used.

**Narrative ability:**

The ability to tell a cohesive and coherent story. In the context of this thesis, we measured narrative ability by using the Multilingual Assessment Instrument for Narratives.

**Narrative macrostructure:**

When a narrative is analyzed on the macrostructural level the so called “story grammar” is analyzed. Story grammar includes information about the setting of a story, the goal, the attempt and the outcomes of episodes and reactions to the events. In the context of this thesis we analyzed the story structure, the story complexity, and the use of internal state terms in the stories.

**Narrative microstructure:**

Narrative microstructure concerns the use of lexical items. Narrative microstructure can be analyzed on the word or sentence level. In the context of this thesis we analyzed the lexical diversity and the lexical richness of the words used for the story.

**Number of Different Words:**

A lexical diversity measure: NDW includes all the different words (types) of a story, thus the total number of words minus the repetitions and derivations of the same root.

**Productive vocabulary:**

This is the ability to use a word by saying it or writing it down.

**Peabody Picture Vocabulary Task:**

A task to measure receptive vocabulary. Developed by Dunn and Dunn (2005)

**Primary school (Dutch):**

Primary school in the Netherlands is compulsory from the age of 5. A mainstream primary school consists of eight years of schooling. In the Netherlands, pre-primary schooling is integrated into primary school since 1985. Children generally start going to primary school at the age of 4 until 12.

**Receptive vocabulary:**

This is the ability to recognize a word, knowing its meaning. Sometimes also called “passive vocabulary.”

**Second language:****Second language learner:**

A second language learner is someone who must learn the language of the environment in addition to the languages he/ she already knows.

**Snapshot:**

A snapshot procedure can be used during observations. During a snapshot procedure pupils are observed during a certain time frame, for example for 10 seconds, after which notes are taken about the observed activities and behavior. In the present study the snapshot procedure is based on the Emerging Academic Snapshot by Ritchie, Howes, Kraft-Sayre, and Weiser (2001) and Early et al. (2010).

**Sociocultural competence:**

Sociocultural competence is part of communicative competence and refers to the awareness of rules and norms of a language in a target culture and context (for instance, rules of politeness).

**Story Grammar:**

Story grammar is the structure of a narrative, which includes components such as the setting and a (or multiple) logical event structure (e.g., initiating event, internal response, plan, action, consequence, and reaction Stein & Glenn, 1979).

**Story Structure:**

Story structure is often called story grammar. A complete structure of a story contains in our study: a setting and three episodes containing an initiating event, an internal response, a plan, an attempt, a consequence, and a reaction.

**Strategic competence:**

Strategic competence is part of communicative competence and concerns how the learner implements planned (verbal, and non-verbal) strategies to succeed in the communicative exchange.

**Structural Complexity:**

Structural Complexity is focusing on whether the structure of a story contains full episodes containing a goal, an attempt, and an outcome.



---

## Abbreviations

---

### Microstructural measures

<b>GIS</b>	Guiraud Index Score.
<b>NDW</b>	Number of Different Words.
<b>MLR</b>	Measure of Lexical Richness.

### Macrostructural measures

<b>IST</b>	Internal State Terms.
<b>SC</b>	Structural Complexity.
<b>SS</b>	Story Structure.

### Language terms

<b>4;6</b>	Age, meaning 4 years and 6 months old
<b>DL2</b>	Dutch as second language
<b>DL2-school</b>	A school or class which is temporally attended (in general one year) by second language learners.
<b>L1/L2/L3/Lx</b>	First/second/third/x <sup>th</sup> language.
<b>MAIN</b>	Multilingual Assessment Instrument for Narratives (Gagarina et al., 2012).
<b>PPVT</b>	Peabody Picture Vocabulary Task (Dunn & Dunn, 2005).



## CHAPTER 1

---

### **Introducing the context for research with newly arrived migrant kindergarteners**

---

## **1.1 The inspiration for this research**

I was born and raised in the Netherlands and earned my teaching degree in 2004. In January 2011 I started working at a school in Utrecht that was devoted solely to pupils who were newly arrived in the country and received intensive training in the Dutch language before entering a mainstream school. This school was one of a handful of such “newcomer” schools in the Netherlands that included the youngest group of school-goers in the Netherlands, namely children aged 4 and 5. Finding ways to best support this special group of pupils has been one of the most rewarding challenges of my work as a teacher.

What led me to the present research is the fact that many municipalities in the Netherlands, as of 2019, do not offer a separate school or a separate class within a school just for newcomers aged 4 and 5. Rather, they enroll these children into a mainstream class for “full immersion” into the Dutch education system without any intensive language support with specific second language pedagogies. Because I had spent so many years working with this group of younger newcomers – developing materials and methods – I had the impression that they benefit from this additional support.

But did they really? Does it really make a difference for these kindergarteners if they receive Dutch language support or if they are mainstreamed? Which other factors besides type of educational facility (separate or mainstreamed) might affect how these children close the gap in their Dutch second language skills? I decided to look into this scientifically in order to shed light on the language development of newly arrived migrant kindergarteners. The expectation is that the results of this study will be of interest to teachers, researchers, and policy-makers alike.

## **1.2 Introduction**

Migration and globalization have impacted the composition of the populations in schools. Even though migration is nothing new, over the past few years the rate of arrivals of migrants in Europe is unprecedented and there is much more variety in countries of origin compared to twenty years ago (Herzog-Punzenberger, Le Pichon-Vorstman, & Siarova, 2017).

Regardless of the migration and educational background of the pupils, all newly arrived pupils need to learn the language used in school. This is particularly important when the second language is the exclusive language used for instruction (Wong Fillmore, 1983), which is the case in the Netherlands. If the second language is not sufficiently mastered, the migrant pupils may be hampered in their academic development (e.g., August, Carlo, Dressler, & Snow, 2005; Verhallen & Schoonen, 1993; Wong Fillmore, 1983). Therefore, pupils must catch up with their peers to meet their potential. According to numerous scholars however (e.g., Paradis, 2011; Unsworth, Hulk, & Marinis, 2011; Muñoz, 2008; Verhoeven, 1991; Wong Fillmore, 1983) second language development

depends on various individual and external factors and thus not all second language learners can be expected to catch up at a similar rate or to a similar level.

One of the external factors which might influence second language development of migrant children is the type of school learning environment they enter after arrival to the new country. The educational landscape for young newly arrived pupils in the Netherlands is quite diverse, ranging from totally separate classes or semi-separated classes, to immediate full inclusion into a mainstream class. It is important to know in what way these different approaches influence the pupils' development of the second language. If specific characteristics of the learning environment that influence second language development of newly arrived migrant pupils are detected, policy makers, local communities, and schools can use this information to better accommodate the needs of this group of pupils. When the needs of migrant pupils are better accommodated, they can close the educational gap with their peers more quickly and effectively. Importantly, the over-representation of this group in lower tracks of secondary education and the high rate of early school leavers among this group, which is now the case, could change in more equal opportunities (European Union, 2013).

Currently, three forms of educational settings are in place in the Netherlands for newly arrived migrant pupils: in the first form, pupils are totally segregated for one or two years; in the second form pupils are semi-segregated for one or two years; and in the third form the pupils are mainstreamed after arrival. However, to date and to our knowledge, there is no scientific evidence to support to choose any form over the other. The aim of this dissertation is therefore to understand how the language learning environment influences the development of the second language of newly arrived migrant pupils.

To this end, I analyzed the language development of forty-two newly arrived migrant pupils aged 4 to 6 and I studied different aspects of their language development in Dutch, namely receptive vocabulary and narrative ability, over a period of two-and-a-half years. Additionally, I compared the pedagogical practices between the settings focusing on teacher behavior and on experiences of the focal pupils, considering that the circumstances in which the second language was taught may be of influence on the second language development of these pupils.

In this introduction chapter, the context of the research is framed. This first chapter starts with the introduction of the population in focus: newly arrived migrant pupils (Section 1.3). Then the Dutch educational system is outlined in general, and the specifics of the educational settings that exist for newly arrived kindergarteners in the Netherlands are discussed (Section 1.4). Subsequently the focus, aim, and main research question of this dissertation are presented (Section 1.5). Chapter 1 ends with the general outline of the dissertation (Section 1.6).

### **1.3 Newly arrived migrant pupils**

In general, and also specifically in the Netherlands, pupils are considered newcomers when they were not born in the country and have not been schooled yet in the language of schooling. This definition of newcomers is very broad and therefore the status of these migrant families can vary from asylum seekers and refugees to economic migrants or expatriates. They can be asylum seekers who are war refugees or economic refugees, expats who will work in the country for a couple of years, or other individuals who have moved for any number of other reasons. In most previous studies of language development of migrant children specific groups of migrants were investigated (for example Turkish-Dutch bilinguals, Leseman, 2000; Blom, 2010), or a difference was made between refugees and other types of migrants (in the study of for example Kaplan, Stolk, Valibhoy, Tucker, & Baker, 2016).

In contrast to previous studies that differentiated between types of migrant populations (Leseman, 2000; Blom, 2010; Kaplan et al., 2016), the population of the present study includes all newly arrived migrant pupils, based on the two criteria named earlier, to which we added an age and a duration criteria: (1) The child was not born in The Netherlands, (2) The child has never been schooled in Dutch before, (3) The child was minimal four maximal six years old, and (4) the child was in the Netherlands for less than a year at the beginning of the study. More details about the participants and their countries of origin will be presented in Chapter 3, paragraph 3.4.2.

These newly arrived migrant pupils are usually considered “second language learners.” Second language learners include all learners who are exposed to the language of the environment at a later age, after having already acquired their home language. However, a distinction has to be made between different types of second language learners. Therefore we want to emphasize that the newly arrived migrant pupils in the present study differ from other second language learners in class who entered the Dutch education system with some knowledge of Dutch since they were born in the Netherlands.

It is expected that second language learners born in the Netherlands will have had more experience with Dutch before starting school than newly arrived migrant pupils because there are special school readiness programs for pupils who speak a different home language than the language of instruction in school. According to the OECD (2015) and several researchers (e.g., Baker, 2011; Cummins, 2000), pupils who speak a language at home that is different from the language of instruction in school may be at risk for academic underachievement, assuming that this underachievement is due to a language delay in the school language. Since pupils need at least 5 to 7 years to learn a new language it is seemingly likely that they could not learn similarly to pupils who do speak the language of instruction at home (Collier & Thomas, 1989; Cummins, 1981; 1989; 2008; Thomas & Collier, 2002).

In order to remedy to these delays programs have been developed in the Netherlands to boost the language development of these pupils (Leseman & Veen, 2016). Specifically, in the Netherlands, a policy started in 2000 to stimulate families with a migration background to join early child care and education (Emmelot, van Schooten, Timman, Verhallen, & Verhallen, 2001). Such programs were usually offered to children aged 2.5 to 6. Upon entering the Netherlands, the newly arrived migrant pupils involved in the present study were already 4 to 6 years old (“kindergarteners”) and expected to start immediately with compulsory Dutch primary education.

## 1.4 The educational settings in the Netherlands

The Dutch education system is made up of primary education, secondary education, and higher education. For all pupils in the Netherlands, education is compulsory from the age of 5 to 16, most pupils (98%) start attending school right after their fourth birthday (CBS, 2003). Primary education in the Netherlands consists of eight years of schooling (Groups 1 through 8, see Table 1.1). Since 1985, early childhood education has been incorporated into primary schools, which means that primary school in the Netherlands starts at the age of 4. Even though the first two years of schooling (Group 1 and 2), when pupils are 4 to 6 years old, are seen as preschool/kindergarten (called *kleuterklassen* in Dutch), they are part of the public primary schools and there is an educational curriculum including all eight years<sup>1</sup>. Throughout this dissertation we will use the term “kindergarteners” for pupils in school aged 4 to 6.

Table 1.1: Overview of the Dutch Primary Education System

	Primary Education							
	“Kindergarteners”		6 Group 3	7 Group 4	8 Group 5	9 Group 6	10 Group 7	11 Group 8
Age	4 Group 1	5 Group 2						
Level								

### 1.4.1 The educational setting for newly arrived migrants in the Netherlands

Compulsory education laws also hold for newly arrived migrant pupils. According to the Reception Conditions Directive for the reception of asylum seekers Article 14<sup>2</sup>, EU member states must ensure access to the education system as quickly as possible and school entry shall not be postponed for more than three months after arrival. In a 2015 letter the Dutch minister of education acknowledges this principle by saying that based upon international treaties refugees until the age of 18, regardless of their status, have a right to

<sup>1</sup> For an extensive description of the Dutch education system we like to refer to the website of the Nuffic, the Dutch organisation for internationalisation in education: [www.nuffic.nl/en/](http://www.nuffic.nl/en/)

<sup>2</sup> Directive 2013/33/EU of the European Parliament and of the Council of 26 June 2013 laying down standards for the reception of applicants for international protection (recast), OJ 2013 L180 (Reception Conditions Directive)

education (Ministry for Primary and Secondary Education and Media<sup>3</sup>, 2015a). Between October 1, 2017, and October 1, 2018, the number of *newcomers* (pupils who were not born in the Netherlands, who have been fewer than four years in the Netherlands, and of whom both parents were not born in the Netherlands either; Educational Inspectorate, 2019, p. 149) increased by approximately 6,000 to a total of 61,544 pupils. This number of newcomers accounted for 4.3 percent of the total of primary school pupils (Educational Inspectorate, 2019).

Even though newly arrived migrants should have access to education on the same terms as children who were born in the country, some newly arrived migrants start in a separate education facility, a so-called *preparatory class*. The European legislation describes such preparatory classes as “aimed at facilitating the access of minors to the national education system, and/or specific education designed to assist their integration into that system” (European Parliament, 2009; Article 14, 2). In the Netherlands, such separate preparatory classes may differ per municipality. Each school board decides how to act when newly arrived migrants settle in their district. According to the Educational Inspectorate the total number of asylum seeker schools as well as large, relatively independent educational facilities for newcomers was 75 on October 1, 2019 (Educational Inspectorate, 2019). However, in addition to these there were also smaller facilities.

In some regions school boards have decided to mainstream (fully integrate) all newly arrived migrant pupils after arrival. In most of these regions, there might be some form of specific language support for a few hours a week, inside or outside the classroom. In other regions there are separate preparatory classes, and the organization of those may take two different forms: (1) an independent full- or part-time school for pupils with Dutch as a second language or (2) a full- or part-time class for pupils with Dutch as a second language within a mainstream school. The difference between the two types of separate classes is that the second form, a separate class within a mainstream school, shares the building with a mainstream school and therefore it is assumed that pupils attending these language support classes have more contact with pupils from the mainstream school. They might share the playground, attend sporting facilities together, or celebrate festivities together.

Most separate classes in the Netherlands are available only *after* kindergarten, around the age of 6 or 7, when pupils start “Group 3,” the third year of primary education in the Netherlands<sup>4</sup>. This implies that most newly arrived migrant kindergarteners (aged 4 and 5) usually start immediately in a mainstream class, in most regions there is no separate class for pupils their age with Dutch as a second language.

---

<sup>3</sup> Ministerie van Onderwijs, Cultuur en Wetenschap - OCW.

<sup>4</sup> The third year of primary education in the Netherlands is compatible with Grade 1 in the U.S. system and Year 2 in the U.K. system.

Schools in the Netherlands that provide Dutch as a second language education are financially supported by the government if they provide education to at least four newly arrived migrant pupils in their first year after arrival. In 2018, this financial support lasted for up to one year for every newly arrived migrant pupil (Ministry for Primary and Secondary Education and Media, 2018; Article 32). However, since December 2016,<sup>5</sup> this financial support includes a second year of funding for minors with a refugee status (Ministry for Primary and Secondary Education and Media, 2018; Article 33). On October 1, 2015, at the beginning of the data collection for this study, 6350 newly arrived migrant pupils received a first-year subsidy in primary education (Ministry for Primary and Secondary Education and Media, 2015b).

To be clear, this separate form of education for pupils with Dutch as a second language is meant to be temporary. The amount of time a pupil spends attending a separate class differs per region, but it is usually one year, with a maximum of two years, after which the pupils should enter a mainstream class.

In what follows, the separate language classes, whether at an independent school facility or within a mainstream school, will be referred to as “Dutch as a second language education.” The abbreviation “DL2-school” will likewise be used as an umbrella term for all forms of separate language classes for Dutch as a second language education. Mainstream schools without a separate language class will be referred to as “Mainstream schools.” One thing that is important to explain about the educational setting for migrants in the Netherlands is the language of schooling. Newly arrived migrant pupils mostly come to the Netherlands with no knowledge of the Dutch language, but often with a multilingual repertoire (see for instance, Herzog-Punzenberger et al., 2017). However, in the Netherlands, Dutch is the only language of instruction in schools, with some exceptions of Dutch-English primary schools and schools with Frisian-Dutch in a bilingual region of the Netherlands (Le Pichon-Vorstman, Erning, & Baauw, 2016).

### **1.4.3 Core objectives for education**

In the Netherlands, schools do not follow a national curriculum with set teaching materials. However, there are predefined end goals per grade, including for kindergarten levels (“Group 1” and “Group 2”). Each school may decide how to organize the education to reach the predefined end goals. DL2-schools for newly arrived migrant pupils can also decide how to organize their own curriculum. The end goals in primary schools have been specified for this particular group by the National Institute for Curriculum Development in the Netherlands (SLO).<sup>6</sup> The overall aim of the education for newly arrived migrant pupils

---

<sup>5</sup> Regulation Ministry of Education, Culture, and Science from November 21. Nr. PO/FenV/1092893

<sup>6</sup> At the beginning of this study these goals could be retrieved from <http://www.doelnieuwkomers.slo.nl/>, however in 2018 this website became unavailable since the goals for newly arrived migrants were under reconsideration. Up until June 2019 no new overview of end-goals for this group were available. The old website can still be retrieved via <https://web.archive.org/>

is, at the end of primary school (around 12 years old), to reach the same level of language and literacy proficiency as the minimum attainment level for mainstream pupils.

This minimum attainment level includes the understanding of 4,500 words and the capacity to use 2,000 words actively for kindergarteners (National Institute for Curriculum Development in the Netherlands, 2013), referring to a specific word list with basic words a child should know before entering the third year of primary school (“Group 3”), when they will start learning to read and write (Verhallen, 2009). However, an inspection of school guides and websites of separate facilities for newly arrived migrant pupils reveals that most schools only state “an appropriate level of Dutch” as end goals for time at a DL2-school, without an indication of what is meant by “appropriate.”

In my own experience as a teacher at a DL2-school, I have witnessed how difficult it can be to determine when a pupil is ready to enter a mainstream school. The question namely remains: when is the Dutch language of that pupil strong enough? For mainstream schools it is also difficult to receive a pupil who has little knowledge of Dutch compared to their peers. Especially for older pupils it seems that the Dutch educational system is not well-equipped to handle pupils who need more time to catch up.

### **1.5 The focus, aim, and general research question of this dissertation**

In some municipalities, school boards have provided separate DL2-schools for newly arrived migrants. Most of these DL2-schools for newly arrived migrant pupils start in Group 3, for pupils aged 6. This means that most newly arrived migrant kindergarteners are mainstreamed after arrival, a decision that lies with school boards. Thus, when newly arrived migrant pupils are younger than seven, there are two possibilities:

- (1) The pupils are schooled within a separate schooling for learning Dutch as a second language (DL2) in:
  - a. A separate language school, or
  - b. A separate language class within a mainstream school.
- (2) They are mainstreamed, that is, they participate in a mainstream class (where specific attention might be paid to second language development).

It is unclear on which grounds school boards take a decision whether to open a DL2-school or class for newly arrived kindergarteners. Maarse and Muller (2017) have recently argued that decisions are mainly made based on practical considerations, concerning logistics, infrastructure, and number of pupils (Maarse & Muller, 2017), whereas following Vermeer (2015) the question *should be* what the best option is for an optimal development of the pupils.

Newly arrived migrant kindergarteners are the population in focus in this study, as there was a specific need to explore the optimal school circumstances of young migrants in their first years in the Netherlands. The fact that in different municipalities different kinds

of educational facilities were provided for this specific group of pupils raised the following question in educational practices: should schools provide separate schooling for kindergarteners and if so, how? While set in the Dutch context, this study will contribute to the current international discussion (e.g., Dryden-Peterson, 2015; European Union, 2013; Herzog-Punzenberger, 2016; Hindman, Wasik, & Snell, 2016) on the education of newly arrived migrant pupils.

In relation to Vermeer's (2015) question: what is the best option for optimal development of the pupils, the general research question of this dissertation is formulated as follows:

*To what extent do pedagogical practices contribute to the second language development of newly arrived migrant kindergarteners in the first two-and-a-half years after arrival in the Netherlands?*

In the initial stages of this project the distinct organizational structure of the schools for newly arrived migrants (being a separate language school or a mainstream school) was considered to include distinct pedagogical practices as well, but this was adjusted by studying the actual pedagogical practices in the classrooms. In the present study we refer with "pedagogical practices" to many different circumstances in the classroom and these could be practices by the teacher as well as by the peers. To answer this general research question, I used an interdisciplinary approach combining language learning and teaching approaches. Therefore, this dissertation involves a study of the language development of newly arrived migrant pupils and a study of the pedagogical practices surrounding them in class. Both approaches are explained in Chapter 2.

## **1.6 Outline of this dissertation**

This dissertation proceeds as follows: After the general introduction to the context in this first chapter, Chapter 2 provides a closer look at the theoretical perspectives of the present study, leading to the formulation of extended research questions in Chapter 3. Chapter 3 will also introduce the population of the study and some methodological considerations.

In the next two chapters I study the second language development of the newly arrived migrant kindergarteners focusing on the development of receptive vocabulary (Chapter 4), and on the development of narrative ability (Chapter 5) and I compared the results of the students depending on the organization of the educational setting in which they were schooled, making a distinction between pupils in DL2-schools and Mainstream schools.

Subsequently, I studied the school learning environment of newly arrived migrant kindergarteners through a thorough analysis of the pedagogical practices in the classroom: in Chapter 6 the different school types (DL2-schools and Mainstream schools) are

compared based on an analysis of the pedagogical practices in the classroom interactions and activities through the observation of teachers in ten different schools. The focus in Chapter 6 is on teacher behavior. In Chapter 7, the pedagogical practices are investigated via the activities and interactions of the specific focal pupils. Chapter 7 zooms in on the focal pupils' direct experiences and interactions with their teacher and peers. It also includes observations about the languages the pupils hear and speak during the day.

In Chapter 8 the results from Chapters 4 to 7 come together: the findings from Chapter 6 and 7 about the characteristics of the school learning environment are taken into consideration as explanatory variables of the second language development found in Chapter 4 and 5. After chapter 8 we will answer the main research question: to what extent do pedagogical practices contribute to the second language development of newly arrived migrant kindergarteners in the first two-and-a-half years after arrival in the Netherlands?

In Chapter 9 all significant outcomes of the studies on the language development and the pedagogical practices in the school learning environment of newly arrived migrant kindergarteners are presented. Chapter 9 then continues with the discussion of all the (sub) research questions. Additionally, in the conclusion the outcomes of our language measures and classroom observations will be evaluated in the light of the theoretical perspectives and the strengths and limitations of this study will be discussed. Chapter 9 includes the implications of this thesis for educational practices: what can teachers and school boards take away from the outcomes and insights of this study?

## CHAPTER 2

---

### **Theoretical perspectives**

---

## **2.1 Introduction**

The focus of this dissertation is the influence of the school setting and pedagogical practices on the second language development of newly arrived migrant kindergarteners, aged 4 to 6. There have been previous studies of the language development of newly arrived migrant pupils, but these studies were mostly cross-sectional and often with older children, or children with a longer exposure to the new second language than newly arrived migrant pupils. In this dissertation pupils in the age range of 4 to 6 and with various language backgrounds were assessed during two-and-a-half years starting within the first six months after having arrived in the Netherlands. Before describing the second language development of these pupils in Chapters 4 to 8, this present chapter will discuss relevant theories concerning second language development of newly arrived migrant pupils and the factors that might influence this.

Most theories that will be presented in this chapter reflect an interdisciplinary approach, with some coming from a more linguistic perspective while others representing a more educational and pedagogical perspective. The literature that will be reviewed emphasizes the view that it is more informative to approach language development from multiple angles in order to obtain a complete picture. In the present study, when both child characteristics as well as characteristics of the school learning environment are taken into consideration, the language development of newly arrived migrant kindergarteners can be evaluated more thoroughly. Also, as the literature will show, language development itself should be investigated with different measures at different levels to obtain a more complete image of the language development and influencing factors.

This theoretical chapter starts with the presentation of two overarching theoretical frameworks. Section 2.2 concerns ecological linguistics and Section 2.3 communicative competence. These two frameworks set the context for this dissertation as a whole and they introduce some methodological considerations for this study, which will be elaborated upon in Chapter 3.

The chapter continues with theories about second language learning. First, in Section 2.4 the difference between different types of second language learners will be discussed. Then the stages in second language learning will be introduced briefly. Then it continues with the topics of language learning and individual differences in language learning. Section 2.5 concerns receptive vocabulary and Section 2.6 narrative ability, the two language components central to this study. Section 2.7 is about the relation between the educational setting and social and cognitive outcomes.

## **2.2 Ecological linguistics in education**

There are multiple factors contributing to second language development. There are factors directly from the child, such as age and linguistic background, but also factors from the

wider environment, such as the learning environment at school. Bronfenbrenner (1979) noticed that in developmental psychology studies dyads, or two-person systems, are common, but that in experiments the traditional focus was still on a single experimental subject, instead of the relationship between two participants, for example mother–child, or as in the present study teacher–pupil, or focal pupil–peer. Bronfenbrenner used the term “ecosystems” to explain how human behavior and development goes beyond a single person because it is nested within and across a set of interdependent structures. Bronfenbrenner initiated a shift in approach from only focusing on the child’s characteristics to an incorporation of the wider environment surrounding the child.

Van Lier (2002) continued to develop this idea of Bronfenbrenner in the field of educational linguistics. He also put great emphasis on the wider context in language learning. According to Van Lier, “the educational linguist must observe the living entity, and learn to understand critically what it does to whom, by whom, and for whom in the multiplexity of semiotic ecosystems in which it (language) operates, or rather co-operates with other meaning-making processes” (Van Lier, 2002, p. 145). Translated to the second language research in an educational setting an ecological view on second language development would mean relating pupil outcomes to the school context and to teacher behavior outcomes. Furthermore, interactions with peers should also be considered as possible factors of influence on the second language development.

With a more ecological view second language developmental research becomes socio-linguistic research. First of all, to look at multiple aspects of language in a more natural context of data elicitation the language assessment becomes more ecologically valid (Botting, 2002), this can for example be done by using narratives, as will be explained later on in Section 2.6. Second, when relevant child characteristics, like Exposure to Dutch, are related to language development the context in which the child grows up is taken into account. Finally, the inclusion of the educational setting and the interactions with teachers and peers allows us to investigate the language development within an even broader context.

The present study is an attempt to look at the second language development of newly arrived migrant pupils from an ecological viewpoint. We will consider many child factors and we will look at the influence of the school learning environment. Pupil’s language outcomes will be related to teacher behavior and language interactions in the classroom. By doing so this current research looked at the second language development of newly arrived migrant kindergarteners from an ecological and interdisciplinary perspective. Both linguistic and sociolinguistic factors were considered, as well as classroom and second language educational perspectives.

### 2.3 Communicative competence

Including both linguistic and sociolinguistic factors as well as classroom and second language educational perspectives has consequences for the way language development is assessed. Is language seen as a compilation of different aspects which can be assessed separately or is it seen in total, focusing more on communication as a whole? “If language is viewed as a social practice of meaning-making and interpretation, then it is not enough for language learners just to know grammar and vocabulary. They also need to know how the language is used to create and represent meanings and how to communicate with others and to engage with the communication of others” (Liddicoat & Scarino, 20013, p.15).

This holistic view on language can be captured in the theory of communicative competence: Language is seen as communication and being able to communicate means being competent in many different facets of communication. Following Savignon (1976), communicative competence provides an inclusive description of the knowledge required to use language because communicative competence includes, in addition to the knowledge of grammatical structure, the knowledge of how language is used to achieve specific communicative goals. Communicative competence therefore means considering the communicative situation as a whole. In these communicative situations, mirroring ecological linguistics, other aspects are important: with whom, to whom, the relationship, the context, and the intent of communication.

In the discussion among scholars about communicative competence terms have been redefined, renamed, or broken down into more specific terms. Hymes (1972) started by challenging Chomsky’s (1965) notion of “linguistic competence”, because besides knowing whether a sentence is grammatical, a child should also know whether or not it is appropriate: “There are rules of use without which the rules of grammar would be useless” (Hymes, 1972, p. 278). Savignon added that separate tests break down a skill into language elements, and as such ignore the complexity of the communicative setting (Savignon, 1976). Canale and Swain (1980) narrowed this notion of sociolinguistic interference down to strategic competence and sociocultural competence. Canale (1983) continued and separated discourse competence from sociocultural competence. Celce-Murcia, Dörnyei, and Thurrell (1995) related their new model of communicative competence with the models of Canale and Swain (1980), Canale (1983), and Bachman and Palmer (1996). Celce-Murcia et al. (1995) separated actional competence from sociocultural competence. Further, Celce-Murcia (2007) separated actional competence into formulaic competence and interactional competence. Figure 2.1 is an attempt to visualize the discussion and evolution of the theory of communicative competence. Figure 2.1 is a combination of the Figures 2 and 3 in the article by Celce-Murcia et al. (1995) with the addition of the first two models and the final model. The first two models were added based on the chronological description in the Celce-Murcia et al. (1995) article. The connecting lines between the Bachman and Palmer (1996) and the Celce-Murcia (2007) models are drawn based on the

Celce-Murcia (2007) article as well as personal communication with Zoltán Dörnyei (March 2018).

Canale and Swain (1980) argued that in a communicative approach there must be an integration of multiple competences and thus no emphasis on one form over the other. Canale and Swain also acknowledge that certain aspects of each type of competence can be studied independently. However, investigating communicative competence would ideally not be done by testing isolated linguistic elements; instead, complete speech is recommended to be used.

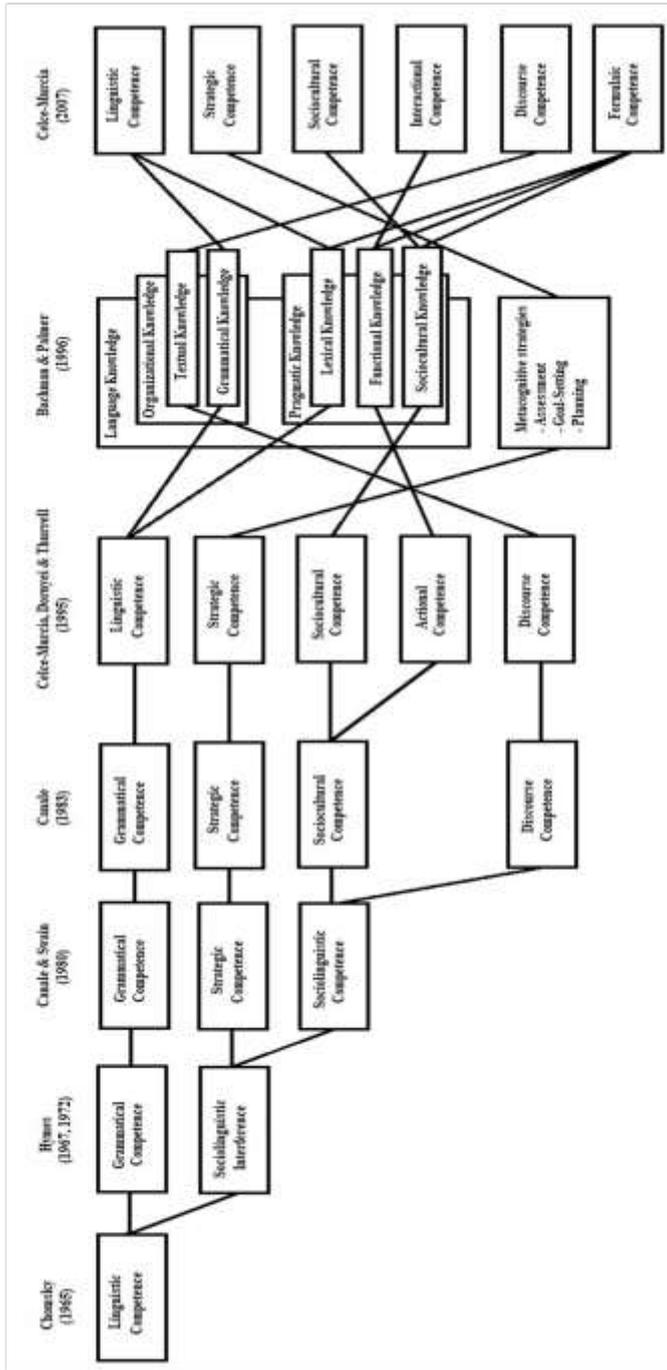


Figure 2.1: Chronological Evolution of the Theory of Communicative Competence.

In this dissertation the concept of communicative competence is broken down into four segments based on Canale (1983) because these are the core components of the theory: grammatical competence (although for present purposes the term linguistic competence used by Celce-Murcia et al., 1995, is preferred), sociocultural competence, discourse competence, and strategic competence. *Linguistic competence* focusses on the skills and knowledge required to accurately express and understand the literal meaning of utterances. It therefore includes features and rules of the language, like for example vocabulary, the formation of words or sentences, pronunciation and spelling. *Sociocultural competence* refers to the appropriate use of words and sentences. Whether or not something is appropriate depends on many different contextual factors, for example the status of the participants, the purpose of the interaction, and many different conventions of interaction. Whether or not something is appropriate also depends on the cultural context and the different languages the interlocutors speak or know. *Discourse competence* implies combining grammatical forms and meanings to realize a cohesive and coherent piece of spoken or written text in different genres. Is the order of the sentences logical and can the reader or listener follow the story? Finally, *strategic competence* concerns how to use verbal and non-verbal communication strategies in an effective way. Strategies can be used to compensate for a breakdown in communication (e.g. overcoming the fact that you do not know a specific word by describing it or acting out) or to enhance the effectiveness of the communication (e.g. slowing down your speech when you want to emphasize something).

Accordingly, second language research should approach second language development more holistically. The goal is to assess multiple aspects of language to capture all segments of communicative competence. With a narrative instrument all aspects of communicative competence can be assessed as the narrator needs to use vocabulary and grammar (linguistic competence) to tell a story, but he or she also needs to tell a coherent story (discourse competence). Furthermore, a narrator needs to adjust to the interlocutor and thus show sociocultural competence; finally, strategic competence can help make the communication more effective.

## **2.4 Second language learning**

The language used in the schools of the participants in this study was new to them. These pupils are therefore called second language learners of Dutch. However, it has to be said that Dutch was not always the second, but maybe the third or fourth language they have encountered in their young lives. In this section, bilingualism and second language learning are first briefly introduced. Second, more detail is given about how receptive vocabulary and narrative ability develop in a second language.

### **2.4.1 Dual language learners, bilinguals, and second language learners**

Learning two or more languages can take place under different circumstances. Although the group of second language learners is, by their own nature, heterogeneous and there is considerable debate about how to classify them (Genesee, Paradis, & Crago, 2004), the existing body of literature identifies two kinds of bilinguals or dual language learners: (1) simultaneous bilinguals and (2) sequential/successive bilinguals. Simultaneous bilinguals acquire two languages early on, mostly from birth: “two languages are present from the onset of speech” (Gass & Selinker, 2008, p. 28). On the other hand, successive or sequential bilinguals are bilinguals who learn the second language after establishing the first: the “second language is added at some stage after the first had begun to develop” (ibidem.). Paradis, Genesee, and Crago (2011) call simultaneous bilinguals, bilinguals, while they refer to sequential bilinguals as second language learners. Nevertheless, they emphasize that there is no real cut-off point between being bilingual or a second language learner. However, they refer to the age of 3 because at that point the first language can be well established and the learning of the first language can be visible in the learning of the second.

The critical milestones of bilinguals are comparable with those of monolinguals. For example, babbling (Oller, Eilers, Urbano, & Cobo-Lewis, 1997; Maneva & Genesee, 2001) or the production of the first words (Conboy & Thal, 2006; Marchman, Martínez-Sussmann, & Dale, 2004; Nicoladis, 2001). Paradis et al. (2011) however do emphasize the fact that bilinguals do not parallel monolinguals in all respects. Especially the amount of exposure to each of the languages influences the rate of development in bilinguals.

All participants in the current study are members of a minority group in the Netherlands and their education is in their “second” language, Dutch, which is the majority language in the Netherlands. However, the population of this study includes several children who were already bilingual when they started to learn Dutch. See Chapter 3, Section 3.4.2 on the demographic characteristics of the participants.

### **2.4.2 The importance of learning the second language**

A second language is acquired to varying degrees of proficiency depending on the context in which the acquirer needs to use the second language (Collier, 1987). For newly arrived migrant pupils arriving in the Netherlands, it is extra important to become highly proficient in the majority language of the country, Dutch, because it is also the language of instruction in school. It is generally acknowledged that a weak command of the second language, when it is the language(s) of schooling, may delay the overall cognitive development of the pupil in school (e.g., August, Carlo, Dressler, & Snow, 2005; Verhallen & Schoonen, 1993; Wong Fillmore, 1983). A limited vocabulary may limit the comprehension of a text (August et al., 2005). As a result, these pupils are at risk of being diagnosed as learning disabled (August et al., 2005) or they end up in a lower track of education (Golberg,

Paradis, & Crago, 2008). Therefore, pupils need to catch up with their peers to reach their potential.

To illustrate this dilemma better we follow Cummins (2008) when he refers to the difference between BICS, *basis interpersonal communicative skills* and CALP, *cognitive academic language proficiency*. Newly arrived migrant children need to learn Dutch to communicate with their peers and teachers, but also to follow the instruction in the school subjects. Pupils who learn the second language in the school context need proficiency in that language in all language domains and in all language skills to use it in all different content areas. In general, second language learners reach conversational proficiency at peer-appropriate level within two years, but they need at least five years to reach grade norms in the academic aspects of a second language (Cummins, 1981). Researchers need to be aware of the difference between these two types of language skills and should focus on how to support second language learners when they are in this catch-up phase to reach peer-appropriate levels on the academic level as well.

### 2.4.3 Individual differences

Although it is important for all pupils to catch up with their peers as soon as possible, according to numerous scholars (e.g., Paradis, 2011; Unsworth, Hulk, & Marinis, 2011; Muñoz, 2008; Verhoeven, 1991; Wong Fillmore, 1983) second language development depends on various individual and external factors and thus not all second language learners can be expected to catch up at a similar rate or to a similar level.

The many factors that can explain individual differences in language development can be categorized as child-internal and child-external factors (Paradis, 2011; Paradis et al., 2011). Child-internal factors are from the child itself, for example: motivation, personality, language learning aptitude, age, gender, and the structure of their first language. Child-external factors exist in the environment outside of the child, which can be the immediate environment of the family or the broader environment of the school or the society. Examples of child-external factors are: quantity, quality, and variation of the input in the first language (L1) and the second language (L2) at home or at school, the quality of interactions between child and parents, the nature of the language use in school, and the richness of the L2 the child is exposed to outside of the home and school.

In the Sections 2.5 and 2.6 more detail is given about two aspects of language development, namely receptive vocabulary and narrative ability. For both skills, a brief overview is provided about how these two language skills develop in second language learners. We will also discuss how the school learning environment might influence this development since this will be one of the foci in this present study and therefore, this particular child-external factor will be investigated more thoroughly in a separate section, Section 2.7.

Learning thousands of words as a first language learner seems to be effortless, while acquiring a large vocabulary as a second language learner seems to be more difficult. There are multiple factors that make second language learning harder. First, second language learners are likely to be exposed to a reduced sample of the language they are learning. Furthermore, the context in which they are learning is more complex than that of first language learners. While first language learners learn their first thousand words in the immediate context surrounding them when they were a baby, second language learners are exposed to new words in their second language that may be more difficult since they might not refer to physical objects to point at. The first thousand words of a second language learner are likely to also include words that refer to meanings outside of the immediate context, concerning for example abstract mathematical concepts (Lightbown & Spada, 2013).

Another aspect which might cause differences between learners is exposure. Exposure to the new language can be investigated in two ways: the quantity of exposure and the quality of exposure. Additionally, both these aspects can be measured in multiple ways. Montrul (2008) poses the problem of how exposure to input is operationalized. Is it just measuring the exposure quantitatively by number of hours, days, months and years, or do you measure the quality of exposure by asking what is actually said during the hours of exposure?

The amount of exposure can be used in investigations concerning different kinds of target language properties (e.g., Gathercole & Thomas, 2009; Hoff, Core, Place, Rumiche, Señor, & Parra, 2012; Paradis, 2011; Jia & Fuse, 2007). The acquisition of some aspects of language requires less exposure than others. Furthermore, for some language aspects, once that aspect is acquired a ceiling effect is reached, while for other aspects more exposure continues to improve scores on that aspect.

Multilingual children have by definition less exposure to one particular language than monolingual children since they must split daytime for language exposure into time for exposure to two or more languages. However, between multilingual children there is also great variance because some only receive input to the L2 in school while others also have L2 input in the home via a parent, siblings, other relatives, television, or other social contacts outside the house. Linguistic researcher should therefore take exposure into account as an influencing variable in language development, but also need to explain how exposure was defined in their study.

The field of multilingualism is complex since “multicompetence is not the sum of monolingual competences” (Cook, 1992; Grosjean, 1992; Cenoz & Genesee, 1998; Herdina & Jessner, 2000). Individual differences should therefore be carefully be considered in second language research.

## **2.5 Receptive vocabulary development**

The development of receptive vocabulary in the first language of young children has been widely studied (for a review see Law and Roy, 2008). The acquisition of vocabulary is often perceived as effortless by most parents and even by school teachers. First language vocabulary acquisition starts from birth and continues developing through at least the age of 12. In the preschool years, at the age of 3 and 4, children learn several words a day (Lightbown & Spada, 2013). After the age of 12 the development of vocabulary continues throughout adulthood (de Villiers & de Villiers, 1979).

### **2.5.1 Receptive vocabulary development of second language learners**

According to Golberg et al. (2008) the development of vocabulary is “a cornerstone” (Golberg et al., 2008, p. 41) of acquiring the new language for second language learners since it is important for their educational success overall. That is, the development of age-appropriate oral proficiency and literacy skills depends heavily on a vocabulary of sufficient size.

The order of acquisition of words (first nouns, then verbs and social expressions) seems to be similar for second language learners compared to first language learners as was established in, for example, a study on Turkish-Dutch (Özcan, Altinkamiş, & Gillis, 2016) and Moroccan-Dutch children (Boerma, 2005). The learning of a second language takes time, nevertheless, Golberg et al. (2008) followed 19 second language learners who were on average about 5 years old when starting to learn English and found that after 34 months of exposure to English their results on a vocabulary test met native-speaker expectations.

Van Druten-Frietman, Denessen, Gijssels, and Verhoeven (2015) showed that pupils with a non-Dutch background (two-and-a-half to three-and-a-half years old) had lower vocabulary scores than their Dutch peers. However, their growth rate was steeper but not steep enough to have overcome the gap with their peers in one year. Comparable were the conclusions in a study of Verhagen et al. (2016) with participants aged two- to six-year-olds. Participants with a non-Western ethnicity or a different home language had, generally speaking, a lower vocabulary level. However, also these children had a stronger growth rate and therefore the gap between the two groups decreased. Nevertheless, at the age of five the gap was still significant.

Compared to monolingual children, bilingual children know fewer words in each of their languages (e.g., Leseman, 2000; Bialystok, Luk, Peets, & Yang, 2010; Oller, Pearson, & Cobo-Lewis, 2007). This is caused by the fact that they must divide their language learning time over all their languages. Furthermore, it is likely that one of their languages was learned or is used in one specific context and not in another. On the other hand, when all vocabularies of the languages of a multilingual child are combined it seems to be comparably large, or even larger than the vocabulary size of a monolingual child (Bialystok et al., 2010).

All in all, it is expected that initially second language learners score differently on receptive vocabulary measures, but they are all expected to show progress. Therefore, in second language research it is advisable to measure pupils longitudinally and furthermore, compare them to other second language learners instead of monolingual pupils. Within the group of newly arrived migrants, multiple variables can be compared, since there are also other reasons for individual variation of second language learning and specifically for receptive vocabulary. The following subsections will discuss some of these factors in more detail.

Pearson, Fernández, Lewedeg, and Oller (1997) studied how vocabulary is affected by the amount of exposure. Their population consisted of bilingual children between eight and thirty months who were followed for about five years. Parents estimated the amount of time their children spent with speakers of each of their languages. They showed that the number of words learned in each of their languages was, to a large extent, proportional to the amount of time the children spent with speakers of each language.

In a study with 3- and 4-year-old Dutch and Turkish-Dutch children, Leseman (2000) found that Dutch vocabulary correlated moderately to strongly with the amount of Dutch exposure at home. It was also correlated with the frequency of high-level language interactions in the home (as measured by a parental questionnaire).

Even though we could not find comparable studies measuring newly arrived migrant kindergarteners, based on the other available studies we expect that the more exposure to the Dutch language these migrant pupils have, the higher their scores will be on our receptive vocabulary measure.

### **2.5.2 Receptive vocabulary development of newly arrived migrant pupils in Dutch**

In many Dutch studies into second language development a comparison is made between second language learners and monolingual children (Appel & Lalleman, 1989; Verhallen, 1994; Strating-Keurentjes, 2000; Verhagen et al., 2016). In these studies it was shown that second language learners of Dutch have a lower Dutch vocabulary than first language learners and that this difference in Dutch vocabulary persists until at least the end of primary education. That being said, the specific group of *newly arrived* second language learners has not been included in many studies. In the current study we are primarily concerned with how the second language of newly arrived migrant pupils develops and we do not focus on how these pupils differ from other second language learners or monolinguals.

One of the Dutch studies that included newly arrived migrant pupils (Jacobs, 2016) included 37 pupils from separate DL2-schools. The pupils, with a mean age of 10 years, attended the DL2-school between one and sixteen months at the time of the study. Jacobs assessed her participants once with the Peabody Picture Vocabulary Test (PPVT; Dunn & Dunn, 2005), an often-used instrument to measure receptive vocabulary. Because

no longitudinal data was obtained, conclusions could only be drawn from one observation with the PPVT: 59.8% of the participants scored below the mean compared to the norm group. All pupils but one scored below their age appropriate score. For 40.5% of the participants, no percentile score could be obtained because their raw score was too low compared to the norm tables. Of the participants, 48.7% had a percentile score below 10%, meaning that 90% of the norm group scored higher. Only four pupils scored higher than 10%, with scores of 14, 19, 47, and 58 percent. This study shows, as expected, that newly arrived migrant pupils have a smaller receptive vocabulary compared to monolingual peers. It also shows that the norms used in standardized assessments are not easily applicable to second language learners.

To summarize, one may say that receptive vocabulary size is a significant predictor of academic achievement and that the learning of vocabulary in the second language should be the focus of schooling during several years. There are however many influencing factors enhancing or limiting the development of vocabulary. To our knowledge, no research has been carried out with the newly arrived migrant group in the Netherlands who start with no knowledge of the target language. The question therefore is: How does the receptive vocabulary of newly arrived migrant kindergarteners develop during their first years in a Dutch learning environment?

## **2.6 Narrative ability development**

Receptive vocabulary knowledge is only one part of a person's language ability. The theory of communicative competence encourages researchers to investigate as many aspects of language development as possible to get a more complete picture of someone's ability. When measuring narrative ability, data can be obtained not only about linguistic competence (which words does a child use?) but also about discourse competence (is the child able to tell a coherent story?).

The ability to tell a story is essential for academic learning, since for example, the ability to tell a story is one of the fundamentals of a child's literacy skills is. (e.g., Bishop & Edmundson, 1987; Bliss, McCabe, & Miranda, 1998; Gutiérrez-Clellen, 2002; Hayward & Schneider, 2000; McCabe & Rosenthal Rollins, 1994; Swanson, Fey, Mills, & Hood, 2005; Torrance & Olson, 1984; Wallach, 2008). Furthermore, narratives are implicated in positive social exchanges (Liles, 1993). That is, a speaker needs to be able to tell about a series of events in a structured way to make clear to the listener what has happened to whom and why. Given the importance of narratives for academics as well as for social relationships, narratives are increasingly used as an assessment tool for children, including at-risk young children (Spencer & Slocum, 2010) and bilinguals (see also Maviş, Tunçer, & Gagarina, 2016).

Narrative assessments provide rich linguistic data and are ecologically valid (Botting, 2002), providing a relatively natural context of data elicitation. Furthermore,

stories have their own rules and structures, which include organizational patterns representing temporal and causal information that form a narrative schema (Berman & Slobin, 1994; Labov, 1972; Labov & Waletzky, 1967; Stein & Glenn, 1979; Westby, 2012). Even though narration is a universal human activity, it must be adapted to the cultural context, the story content, and to the listener (Lindgren, 2018). Thus, in addition to the data collected about their receptive vocabulary in this study, data is also collected about the pupils' narrative ability.

### **2.6.1 Narrative ability development**

When children start talking they start with producing a string of unrelated sentences. When they are between 2 and 3;6 years old these sentences form a series focusing on a main character or main central theme, however there is not a cause-effect relationship between the individual sentences yet. Around the age of 4 cause-effect relations emerge between the sentences and by the age of 5, most children are able to produce a story that includes both a central theme and cause-effect links between the sentences of the story (Westby, Van Dongen, & Maggart, 1989). As the child grows, especially from ages five to ten, the stories that are told become more sophisticated in their features. Children create unified plot structures, motivate the events by using internal states terms (about for example feelings of the characters), and they also include extra information to appreciate the listener's needs for information. The stories also become more complete: they include a setting of the scene and more problem resolutions sequences. Furthermore, as narrators, children give more frequent and more complex comments on the actions as they grow older (Kemper, 1984). The work of Berman and Slobin (1994) shows that the versions of stories by older children include more explicit references to cause and effect than stories by younger children. They also include more compound time referencing and more complex theory of other minds. Table 2.1 provides an overview of how narratives generally develop from preschool into adulthood (Westby, 2012).

The ability to tell a story is related to cognitive maturity, and thus depends on age. A good story depends on several abilities of the speaker: the ability to understand an underlying schema, to correctly interpret pictures, to link narrative content together, and to verbalize story content (Lindgren, 2018). Particularly among children aged 4 to 5 years, major developments occur in the use of story grammar elements such as goals, actions, and outcomes (Muñoz, Gillam, Peña, & Gulley-Faehnle, 2003; Price, Roberts, & Jackson, 2006; Fivush, Haden, & Adam, 1995). Around the age of 7, the development of narrative ability reduces (Schneider, Hayward, & Vis Dubé, 2006).

Narratives of children below five years of age consist most often of events with relatively simple goal-attempt-outcome sequences (Blankenstijn & Scheper, 2003; Trabasso & Rodkin, 1994). Stories become more abbreviated and complete at the age of 5 to 6, but still become more elaborate as children grow older. Furthermore, Nakamura

(2009) reports that with the increase of age, narrators more overtly referred to emotions and used a wider variety of emotion terms. In a study by John, Lui, and Tannock (2003) with 7-, 9-, and 11-year olds only limited age-related differences in story structure were found. John et al. therefore concluded that a sense of story structure is well-developed by around age 7.

Table 2.1: First Language Narrative Development. Abbreviated Version of Westby's Table 2<sup>7</sup>.

Preschool	<p><i>Description:</i> unconnected sentences; order not important</p> <p><i>Action sequence:</i> series of actions, generally with a temporal sequence; centering may be present- story may have a central character or a central theme (actions that each character takes)</p> <p><i>Reactive sequence:</i> Cause-effect sequence of events; chaining of actions</p>
Early Elementary	<p><i>Abbreviated episode:</i> Centering and chaining present; stories have at least initiating event (problem) response (character's reaction to problem), and consequence</p> <p><i>Complete episode:</i> Centering and chaining present; story has an initiating event, internal response, plan, attempt (carrying out plan), and consequence</p>
Later Elementary	<p><i>Complex episode:</i> Like complete episode, but with obstacle(s) to goal and multiple attempts to reach goal</p> <p><i>Multiple sequential episodes:</i> More than one chapters are arranged in chronological order; at least one episode should be at least complete</p>
Adolescent/Adult	<p><i>Interactive episodes:</i> Two or more characters with interactive goals</p> <p><i>Embedded episodes:</i> One narrative structure embedded within another (An interactive episode may be embedded)</p>

## 2.6.2 Macrostructure and microstructure

Both linguistic competence as well as discourse competence from the theory of communicative competence can be analyzed by examining pupils' narratives, since narratives can be analyzed on two distinct, but interconnected levels (Liles, Duffy, Merritt, & Purcell, 1995): the microstructural level and on the macrostructural level.

Narrative *microstructure* concerns the use of all kinds of lexical items, and is therefore related to linguistic competence. The aspects of microstructure in stories by second language learners will show large development, because these aspects must be newly acquired by second language learners. For every new language they learn they need to learn new vocabulary otherwise they are not able to tell a good story in that language.

<sup>7</sup> Westby, 2012, p. 202–204.

For example a German-Dutch bilingual child might be able to tell a long story in German with many different words, but on the other hand will tell the same story in Dutch much shorter and with less variation in the words since this is the language the child is just starting to learn.

Narrative *macrostructure* refers to the global organization of a story beyond the word, sentence, or utterance level (Blom & Boerma, 2016) and is therefore related to discourse competence. Macrostructure needs not to be newly acquired because this global organization of stories seems to be universal and once acquired in one language can be transferred to another language. Macrostructural aspects can be analyzed using so called a “story grammar,” which includes components such as the setting and a logical event structure (e.g., initiating event, internal response, plan, action, consequence, and reaction; Stein & Glenn, 1979). If that German-Dutch bilingual child has learned that a story contains a setting a goal and a result in German, he or she will try to include these parts of the story also in Dutch, although they might not be able to elaborate on the matter in both languages. Macrostructural performance is therefore seen as less language dependent than microstructure performance, because it is suggested that macrostructure is partly dependent on cognitive schemas available for all the languages of a speaker (Iluz-Cohen & Walters, 2012; Pearson, 2002).

Thus, using narratives as an assessment provides the opportunity to distinguish between knowledge that seems to develop within the context of learning a new language, and knowledge about how to tell a story that is seen as more universal (Cummins, 1984). An advantage of using narrative tasks, according to different researchers (e.g. Cleave, Girolametto, Chen, & Johnson, 2010; Boerma, Leseman, Timmermeister, Wijnen, & Blom, 2016; Paradis et al., 2011) is that they are less biased for multilingual participants because narratives do not only tap into language specific knowledge; narratives also require general skills like cognitive, social, and pragmatic skills (Liles, 1993). The standardized part of the narratives – the story grammar or macrostructure – is expected to be equally well developed for monolingual pupils and bilingual pupils, and thus bilinguals are expected to show age appropriate behavior. The microstructure however is expected to really show second language development and on this part we will see differences in proficiency in the second language.

Pearson (2002) stresses the usefulness of narratives as a linguistic measure because of the presence of both linguistic and discourse competence in one instrument: “By separating the scoring of the stories into independent components and even subcomponents, we can examine the separate contribution of each element to more global measures of the children’s growth” (Pearson, 2002, p. 137). Furthermore, narratives are especially useful for examining the language skills of bilinguals, since there seems to be greater dissociations between their component language skills (Pearson, Oller, Umbel, & Fernández, 1996).

### 2.6.3 Narrative ability development of second language learners

The narrative development of bilinguals has been the subject of many studies. Squires, Lugo-Neris, Peña, Bedore, Bohman and Gillam (2014) report on a study on bilingual children with a retell task<sup>8</sup> with two wordless picture books by Mayer *Frog On His Own* (Mayer, 1973) or *One Frog Too Many* (Mayer, 1975). They found that bilingual children follow the general pattern for narrative development that Berman (1988, 2001) describes for monolingual children: bilingual children, like monolingual children show an increase of their use of macrostructural elements as they grow older.

Many studies have shown narrative skills, once acquired in one language is transferred to another language (Pearson, 2002; Gutiérrez-Clellen, 2002; Squires et al., 2014; Iluz-Cohen & Walters, 2012; Fiestas & Peña, 2004) – but this only concerns macrostructure. Microstructure seems to be more reliant on language ability, and thus more dependent on the level of proficiency in the language being tested.

The expectation is that the amount of exposure to the L2 might influence the microstructure since this measure is considered dependent upon L2 language proficiency. The macrostructure of a narrative, on the other hand, is expected to be similar in both languages of a child regardless of the amount of exposure to L2 since there is assumed to be a transfer of the macrostructure from one language to the other (Iluz-Cohen & Walters, 2012; Pearson, 2002). That being said, a child does require a minimum level of specific vocabulary knowledge in L2 in order to be able to narrate about macrostructure: “Narrative structure [...] belongs to a general conceptual level which is available even in the processing of the L2 once a certain linguistic threshold has been passed” (Viberg, 2001, p. 124). Before that linguistic threshold has been met the focus of the narrator is on concrete events (Viberg, 2001).

We know of only one Dutch study specifically involving the narrative ability of newly arrived migrant pupils (Jelsma, 2015; Le Pichon & Baauw, 2015). The participants of this study were between five and ten years old and they were asked to tell a story elicited by a text free picture story about a boy going to the doctor for a shot. All participants told the story twice with four to four-and-a-half months in between. The focus of this study was on two aspects of microstructure (namely, narrative complex language and lexical diversity) and one part of macrostructure (namely, the use of Internal State Term, words about the mental states of the story characters). The whole group of participants had a positive development on the use of complex language and the use of IST, but not on lexical diversity (measured by calculating the TTR) there seemed to be no difference between the two sessions on that aspect of narrative ability. We can compare the data of our participants with two of the measure of Jelsema (2015): the data on lexical diversity and on the use of IST. As will be explained in Chapter 5, we however used a different lexical diversity

---

<sup>8</sup> In a retell task the participants listens to a story by the researcher and then have to retell the story.

measure, so we might find differences on this aspect. For IST development we could expect a similar result like Jelsema, thus a higher use of ISTs.

To summarize, narrative ability is, like receptive vocabulary, an important linguistic skill that influences academic development. However, there are many influencing factors enhancing or limiting this development. Despite many recent studies involving narrative development of bilingual pupils, it is unclear how these outcomes may relate to young newly arrived migrant pupils. Therefore, the present study also addresses the question: what is the influence of child characteristics on the development of narrative ability of newly arrived migrant kindergarteners? Besides internal child characteristics also external child factors are taken into account in the present study and that is what we will deliberate on in the next section.

## **2.7 Relation school learning environment and social and cognitive outcomes**

External child factors that influence second language development concern the linguistic quality and the quantity of the input of the L2. The main source of input in the L2 for most newly arrived migrants comes from the interactions they have in their educational setting.

In studies on early education a positive relationship has been found between children's participation in early education settings and cognitive – and to some extent social and emotional – outcomes (see for a review Yoshikama et al., 2013). Many governments, including that of the Netherlands, have therefore increased access to early education for young children (especially those aged between 3 and 6 years). There is furthermore a growing recognition that it is not simply the quantity of exposure that matters, but also the quality of the experience in early education (e.g., Hatfield, Burchinal, Pianta, & Sideris, 2016; Mashburn et al., 2008; Melhuish et al., 2015; Vandell, Belsky, Burchinal, Steinberg, & Vandergrift, 2010; Wasik & Hindman, 2011; Yoshikawa, Weiland, & Brooks-Gunn, 2016). The quality of the care and education is reflected by the components of the environment that are related to positive outcomes in the academic as well as social domains (Love, Meckstroth, & Sprachman, 1997; NICHD ECCRN, 2002a). The environment in care and education consists of multiple components and therefore is the quality of care and education a multidimensional construct and should be assessed using various kinds of metrics and differing units of analysis.

### **2.7.1 How to measure educational quality?**

The school learning environment of young children can be analyzed at different levels. Phillips and Lowenstein (2011) mention three tiers of high-quality early care and education (ECE). They call them: a) *process quality* concerning the child-adult relationship and interactions, mostly evaluated focusing on the emotional support, behavioral support, and

instructional support in the classroom, thus how the learning is happening in the classroom; b) the *structural features of care*, for example the child-to-staff ratio, cost of care, and professionalization of the staff, and c) the *surrounding community and policy context*. Phillips and Lowenstein further define high quality in ECE as a setting with features that foster positive developmental outcomes.

Outcomes of studies about structural features of care are mixed and inconsistent, and they vary between countries regarding whether structural features influence the development of children (e.g., Early et al., 2006; Howes et al., 2008; Mashburn et al., 2008; Montie, Xiang, & Schweinhart, 2006; Slot, Lerkkanen, & Leseman, 2015b). Early et al. (2006) for example investigated the association between teachers' education levels and gains in children's outcomes in the United States. They concluded that education, training, and credentialing are not consistently related to classroom quality in programs that serve 4-year olds or other academic gains for these children.

Montie et al. (2006) compared results from 15 different countries and found that the structural feature group size at age 4 was not related to language or cognitive outcomes at age 7. However, after further analysis they concluded that the relation "between group size and adult-child ratio and process characteristics are country-specific rather than universal" (Montie et al., 2006, p. 329).

Slot et al. (2015b) found that single structural quality features of ECE are not consistently related to process quality. However, some configurations of structural characteristics did influence process quality. One of the five countries in their analysis was the Netherlands, and they found some main effects and interaction effects in their data about ECE for children from 2 to 6 years old. A smaller group size in combination with more unfavorable child-to-staff ratio was related to higher scores on emotional and behavioral support. Furthermore, more work experience of the teachers was related to higher quality in all domains of process quality and curriculum quality. Additionally, when a center provided more professional development opportunities, this was related to higher curriculum quality.

Nevertheless, process quality has been found to be a better predictor of child outcomes than indicators of structural quality (see Mashburn et al., 2008; Whitebook, 1989). Process quality or program quality (Pianta et al., 2005; Mashburn et al., 2008) refers to the direct experiences of children while they are enrolled in ECE. These experiences are proximal-level interactions and transactions among teachers, children and also materials (Bronfenbrenner & Morris, 1998; Pianta, 2003). It includes the ways in which teachers organize routines and implement activities and lessons. Furthermore, it concerns how teachers make interesting materials available to children. Finally, it is about the quality of teachers' interactions with children (Hamre & Pianta, 2007; Harms, Clifford, & Cryer, 1998). According to Van Schaik, Leseman, and De Haan (2018) "high process quality is defined as reflecting a setting where teacher-child relationships and interactions are warm,

sensitive, and supportive” (p. 897). To investigate process quality in education, researchers primarily use classroom observations.

However, linear associations between indicators of quality and child outcomes are mostly small (Burchinal, Kainz, & Cai, 2011) and sometimes even non-significant (e.g., Weiland, Ulvestad, Sachs, & Yoshikawa, 2013). Pianta, Barnett, Burchinal, and Thornburg (2009) argue that this is due to a large variation in program designs, curriculum, staffing, and level of educational aims. According to Mashburn et al. (2008) there is on the other hand not much variation in for example observations with the ECERS-R (Harms et al., 1998) – an observation tool for overall quality of the environment in pre-kindergarten – which makes it difficult to establish effects as well. Mashburn et al. argue that due to regulation all facilities have improved the quality of the classroom reducing the variation between the classrooms and with that, attenuating predictions of the ECERS-R to child outcomes. Hatfield et al. (2016) investigated the small effects of process quality on child outcomes in more detail and found that there are certain thresholds for quality measurements: an effect is only visible when scores are above a certain threshold (see also Zaslow et al., 2010).

All of these aspects must be considered when measuring classroom quality and looking for effects of quality on child development. An essential issue for this study is whether these quality measures are suitable for measuring the environment of second language learners. Peisner-Feinberg et al. (2014) reviewed ten studies to see how well measures of quality in early child care education would also reflect the needs of dual language learners. They concluded that the widely used measures for quality discussed above function similarly for dual language learners when compared to mainstream populations. However, they encourage researchers to further disentangle the aspects of quality that stand out for dual language learners, for example examining whether interactions are culturally and linguistically responsive. Furthermore, it is also important to keep in mind that even though similar methods are used for observing process quality, the context of the observation – for example the population in the classroom, regional differences, or implemented programs – make it hard to compare outcome measures (Buysse, Peisner-Feinberg, Páez, Scheffer Hammer, & Knowles, 2014; Buell, Han, & Vukelich, 2017).

In the present study, process quality will be studied at two levels: (1) teacher behavior and (2) focal pupils’ experiences. The next two sections will discuss some findings from previous research regarding these two levels.

### **2.7.2 Focus on teacher behavior**

Quality of education largely depends on the processes in the classroom such as socio-emotional climate and instructional support of the teacher (La Paro, Pianta, & Stuhlman, 2004). Teachers who are sensitive towards the pupils and who create a positive climate in

their classroom tend to be more familiar with the academic needs of the individual children in their classroom (Helmke & Schrader, 1988). Furthermore, studies show that the quality of the social and emotional interactions in the classroom predicts pupils' performance on standardized test (Pianta et al., 2005) and it predicts pupils' engagement in the classroom (Bryant et al., 2002; NICHD ECCRN, 2002b).

Regarding high quality instructional support, teachers monitor pupils' performance and they provide additional explanations and ideas (Meyer, Wardrop, Hastings, & Linn, 1993). Already in 1986 Brophy concluded based on a large body of research that when teachers emphasize academic objectives and use effective management strategies pupils in their class achieve more. These teachers provide their pupils with feedback through scaffolding and support (Yates & Yates, 1990). Due to increased pupil engagement in lessons with high instructional quality these pupils function higher academically (NICHD ECCRN, 2004). Furthermore, Hamre and Pianta (2005) found that high-quality instructional support had a positive effect on pupils who were at-risk and improved their academic performances.

Focusing on actual teacher behavior during instruction is more important than looking at the implementation of a specific program, as Justice, Mashburn, Hamre, and Pianta (2008) concluded. They looked at the relation between *procedural fidelity* and quality of language and literacy instruction. Procedural or curriculum fidelity is whether a teacher implements the procedures of a curriculum in an accurate, efficient, and appropriate way. Even though most participants in the study of Justice et al. (2008) were able to implement the lesson plans as they were intended, the quality of the instruction was considered low. Focusing on language development, teachers did not provide instruction with strategies such as asking open-ended questions, repeating and extending children's utterances, or modeling advanced vocabulary. Also, for literacy instruction the quality could improve if the instructions were more explicit, systematic, and purposeful. Research proved that with professional development of teachers the quality of the classroom can be enhanced (e.g. Hamre et al., 2012; Pianta et al., 2014).

### **2.7.3 Focus on focal pupils experiences**

When the focus is on teacher behavior mostly the classroom as a whole is observed, without specific attention to specific children, called the focal pupils. When the focus is on specific targeted pupils, a different pattern might be observed compared to whole group observations. First, it is interesting to investigate which activities a pupil is engaged in during a day. It could be that a certain pupil has different activities than the larger part of the group, or that the teacher is not engaged in the activity of that particular pupil. In a study by Kutnick et al. (2007) several European countries were compared from this perspective. They found a high degree of similarity across the countries: "a majority of children's activities were undertaken with peers and with relatively little planning or

support by practitioners” (Kutnick et al., 2007, p. 402). These group settings, in which children play and collaborate with peers, may offer children multiple opportunities to develop their social skills and cognitive skills (Diamond & Lee, 2011; Elias & Berk, 2002; Howes et al., 2011).

In addition to which activities the pupils are engaged in, the interactions (and following from that the language use in the classroom), are also of interest. According to Rydland, Grøver, and Lawrence (2014), “differences in vocabulary development can to some extent be attributed to variability in the talk offered to young children” (Rydland et al., 2014, p. 214). Note that having interaction with teachers or peers does not necessarily imply that languages being used or words are being learned. Blum-Kulka and Grobett (2014) report on the longitudinal observation of young migrant pupils in the age of 3 to 7 years old and show that in the first months in school there were actually limited conversations, with mainly non-verbal interactions.

It is expected that in the beginning the interactions between second language learners and their peers might be of limited help for language learning. According to Cekaite and Aronsson (2014), migrant pupils need to first learn a certain amount of communicative behavior before entering into verbal interactions with linguistically competent peers. Before that threshold is reached, interaction mainly consists of playful keying (using simple words or gestures in their play to get attention) and ranging from light teasing to ceremonial verbal rituals in their play (Cekaite & Aronsson, 2014; Tabors & Snow, 1994). Blum-Kulka and Snow (2004) summarize research on peer learning as follows: “Thus, while peers can be powerful sources of language input and information, they are not helpful to learners in the earliest stages of acquiring the target language. Some social engineering, both of the peer capacities and of the newcomer’s communicative attempts, is necessary to ensure optimal effectiveness” (Blum-Kulka & Snow, 2004, p. 296)

Mashburn, Justice, Downer, and Pianta (2009) found that pupils’ development of receptive and expressive language during pre-kindergarten was stronger in a classroom with high peer expressive language abilities (see also Justice, Petscher, Schatschneider, & Mashburn, 2011). However, they also found that initial level of language skills and better classroom management influenced this effect and thus emphasized the importance of investigating language development in a more ecological way.

#### **2.7.4 Pedagogical practices in the Netherlands**

The general pedagogical practices in the Dutch educational setting have been studied extensively. De Haan, Elbers, Hoofs, and Leseman (2013) investigated whether class composition and teacher-managed activities had an effect on disadvantaged children’s emergent academic literacy and mathematical skills. These disadvantaged children came from ethnic-minority families where at least one of the parents had a weak educational background or these children were from Dutch families where both parents had a weak

educational background. The overall result from the study by De Haan et al. was that educationally disadvantaged children benefit from a more balanced socioeconomic and ethnic-cultural classroom composition. The authors speculated that this finding was due to their interactions with peers with better expressive abilities and larger vocabularies. Furthermore, they suggested that a classroom with a more balanced composition would lower the workload for the teacher, which means she or he could spend more time guiding and instructing the children. De Haan et al. (2013) also found that using a special education program did not have a positive effect on disadvantaged children's literacy and math development. However, this may point to a lack of implementation fidelity, since trained and non-trained teachers both provided language, literacy, and math activities in the classroom. Nevertheless, there were many differences in time spent in these different activities, which seemed to be crucial. They could not find a positive effect of having a special developed education program, but they did find a relation between time spent in language, literacy, and math activities on the one hand and child outcomes on the other hand. Thus the implementation of a certain program should be carefully be evaluated and actual time spent in activities should be measured.

In 2009 a cohort study named "Pre-COOL" was started to investigate the societal benefits of the investments of ECE in the Netherlands (Veen et al., 2012a; 2012b). The intention of the study was to follow different cohorts of children attending early child care (or not) and see what impact attendance had on different outcomes. Furthermore, multiple variables of the educational facilities as well as child characteristics were included. Different studies on these cohorts will be reviewed here even though they do not specifically concern newly arrived migrant pupils. That is, second language learners were included in these cohorts, but *newly arrived* second language learners were not looked at separately. Nevertheless, we still discuss these studies since they can illustrate the general practices in the Netherlands; moreover, there could have been newly arrived migrant pupils included in these Pre-Cool studies, although not intentionally or explicitly.

Leseman and Slot (2013) investigated the quality in two kinds of early child care and education facilities in the Netherlands to see how this quality influenced the language development of the children. First, they found an overall difference with international studies: even though instructional support scores were lower compared to emotional support in international studies, in the Dutch facilities this instructional support was even lower. Second, they saw little differences between facilities in the Netherlands concerning the emotional quality of the group processes; however, there were differences in the educational quality. The differences in outcome on emotional quality could be explained by the structural characteristic of group size and child-staff ratio. Level of education of the teacher did not seem to have a consistent relation with process quality. However, the amount of time spent in creative, educational, and play activities compared to time spent in routines could explain differences in both emotional as well as educational quality.

Leseman and Slot recommend implementation of a system for quality care and continuing professionalization. Furthermore, they recommend the systematic planning of the daily activities over a longer period of time with a good balance between (guided) play and educational activities.

De Haan, Elbers and Leseman (2014) investigated children's daily experiences in preschool classrooms and found that teacher-managed academic activities did not occur often. Despite the low occurrence, young children (ages 2 to 4) exhibited language-literacy and math development that was positively associated with the amount of teacher-managed activities. For kindergarten children (ages 4 to 6) only a positive association of the amount of math activities on math development was found and not for language activities.

Slot, Broekhuizen, Leseman, and Veen (2015a) continued with the Pre-COOL data on which previous researchers already had concluded that the emotional and behavioral support was of medium to high quality, while the instructional support was of low to medium quality (Slot, 2014; Leseman & Slot, 2013; Veen et al., 2013). Slot et al. (2015a) looked at effects of quality on cognitive development and they found small positive effects of emotional support on vocabulary development. Informal and personal conversations with the child about everyday experiences which are related to the interest of the children seem to contribute to vocabulary growth. Slot et al. did not find an effect of instructional support on vocabulary. Their explanation was that the instrument they used, the Peabody Picture Vocabulary Task (Dunn & Dunn, 2005) was not sufficiently sensitive because it measures general vocabulary and not the specific vocabulary that was actually taught in class. Instructional support does seem to have an effect on selective attention.

Similarly, Verhagen et al. (2016) looked at the influence of certain activities on vocabulary development. In environments where teachers encouraged fantasy play and where free play has been enriched by the teachers a positive relation has been found for vocabulary development. Contrary to De Haan et al. (2013) Verhagen et al. (2016) did find an effect of the use of an early care program on vocabulary development.

To summarize, the quality of Dutch early educational facilities could be optimized for both L1 and L2 learners, especially in the area of instructional support (e.g., Slot et al., 2015a; Slot, Leseman, Verhagen, & Mulder, 2015). However, the quality of educational facilities for newly arrived migrant pupils has not yet been investigated specifically. Therefore, the influence of this quality on the development of this target group cannot be established yet.

In Chapter 1 the Dutch educational context for newly arrived migrant kindergarteners was described, distinguishing two types of organization: separate DL2-schools or Mainstream schools. It is relevant to know whether these two types show differences in quality. However, it might also be possible that different teachers within one and the same educational facility act differently and consequently influence child outcomes, as De Haan et al. (2013) have shown for individual teachers using the same program. It is

for this reason that it is important to investigate process quality in the school learning environment focusing on teacher behavior as well as on focal pupils' daily experiences to establish actual differences and similarities.

In the next chapter the extended research questions will be presented. The questions reflect an ecological and sociolinguistic perspective on language development and justify an interdisciplinary research approach to assessing multiple aspects of language development and taking into account different aspects of the educational setting of newly arrived migrant pupils, as was discussed in this theoretical chapter.



## CHAPTER 3

---

### **Present study: extended research questions and methodology**

---

### **3.1 Introduction**

The main research question of this dissertation is: To what extent do pedagogical practices contribute to the second language development of newly arrived migrant kindergarteners in the first two-and-a-half years after arrival in the Netherlands? The present chapter will introduce the extended research questions based on the theoretical background of Chapter 2 (Section 3.2). Furthermore, in the current chapter the methodology of the dissertation as a whole will be discussed (Section 3.3). Finally the participating schools, teachers, and pupils will be introduced (Section 3.4). The participants will be introduced in this chapter because they the same pupils participated in all four sub-studies of this dissertation. Therefore, in Chapter 3 the participating schools, teachers, and pupils will be described in detail, and in the subsequent chapters they will only be briefly discussed with specific tables summarizing the necessary descriptive statistics for that particular chapter.

### **3.2 Extended research questions**

Considering the theoretical background of Chapter 2, this dissertation intends to describe multiple aspects of the second language development of 42 newly arrived migrant pupils, aged 4 to 6, in two different school contexts: a) Separate Dutch as Second Language schools (DL2-schools) and b) Mainstream schools. First, because there is not much information about the language development of newly arrived migrant kindergarteners. Second, even though there seems to be clear organizational and logistical issues between schools that provide separate schooling or not, the pedagogical practices have not yet been thoroughly investigated and compared. Third, there is need for research on the relation between the school learning environment of newly arrived migrant kindergarteners and their second language development. This is the gap of knowledge this dissertation wants to fill.

In this dissertation both a linguistic and a pedagogical perspective were adopted. By means of assessments of their second language development linguistic data was gathered and an overview of the pupils' second language development could be provided. Additionally, two observation instruments were used to investigate the pedagogical approaches in the educational facilities these pupils visit. Those two separate data collections were brought together by investigating the relationship between the educational setting and the second language development of the newly arrived migrant pupils.

The second language development of newly arrived migrant kindergarteners was investigated by taking into account the influence of child characteristics as well as the characteristics of the school learning environment. The three major aims were: (1) to gain insights into the second language development of newly arrived kindergarteners; (2) to gain insights into the pedagogical differences and similarities of the different types of schooling for newly arrived migrant pupils; and (3) to investigate to what respect these differences in

the educational setting affect the development of the second language of newly arrived migrant kindergarteners in the first stages of their schooling in the Netherlands.

Therefore, to reach an answer to the general research question “*To what extent do pedagogical practices contribute to the second language development of newly arrived migrant kindergarteners in the first two-and-a-half years after arrival in the Netherlands?*” five sub-questions have been formulated. All questions reflect the theoretical perspective of an ecological and sociolinguistic view on language development and justify an interdisciplinary research approach on multiple aspects of language development.

First, two language skills of the pupils’ second language, Dutch, the language of instruction at school, were assessed and their development was evaluated:

- (a) How does the *receptive vocabulary* (in Dutch) of newly arrived migrant kindergarteners develop during the first two-and-a-half years of schooling in the Netherlands in relation to school type?
- (b) How does the *narrative ability* (in Dutch) of newly arrived migrant kindergarteners develop during the first two-and-a-half years of schooling in the Netherlands in relation to school type?

The development of receptive vocabulary and narrative ability is related to school type in order to see whether pupils attending a separate DL2-school in the first year have a different kind of second language development compared to pupils who immediately start in a Mainstream school. For these two research questions we assume that the differences in organizational structure as such between DL2-schools and Mainstream schools have consequences for the pedagogical practices. Without further distinction school type is included as a variable in the Chapters 4 and 5. However, we should stress that the school types of the pupils only differed in the first year of education since later on all pupils attended a Mainstream school.

In order to evaluate the division of school type between DL2-schools and Mainstream schools the similarities and differences between these different educational facilities for newly arrived migrant kindergarteners is studied. Specifically, the school learning environment was investigated in more detail by looking at the pedagogical practices in the classrooms by means of observations. Two questions will be answered:

- (c) What are the differences in characteristics of the school learning environment *regarding teacher behavior* between DL2-schools and Mainstream schools?
- (d) What are the differences in characteristics of the school learning environment *from the point of view of focal pupils’ experiences* between DL2-schools and Mainstream schools?

Finally, all the above questions come together, and we will study the relationship between the pedagogical practices in the school learning environment and the development of pupils' receptive vocabulary and narrative ability. The final sub-question is therefore:

- (e) To what extent do *differences in the characteristics of the school learning environment* during the first year after arrival relate to receptive vocabulary development and to narrative ability development of newly arrived migrant kindergarteners during the first two-and-a-half years of schooling in the Netherlands?

Sub-question (a) will be answered in Chapter 4. In Chapter 5 sub-question (b) will be answered. Chapter 6 answers sub-question (c) and Chapter 7 sub-question (d). Finally Chapter 8 concerns sub-question (e), see Table 3.1.

Table 3.1: Overview of the Research Questions over the Data Chapters 4 to 8.

	Chapter 4	Chapter 5	Chapter 6	Chapter 7	Chapter 8
<b>RQ (a)</b> Receptive vocabulary	X				X
<b>RQ (b)</b> Narrative ability		X			X
<b>RQ (c)</b> Teacher behavior			X		X
<b>RQ (d)</b> Focal pupils' experiences				X	X
<b>RQ (e)</b> Relation learning environment and language development					X

### 3.3 Methodological considerations

#### 3.3.1 Second language assessment

To track the development of receptive vocabulary and narrative ability of the participating newly arrived migrant kindergarteners, their second language must be assessed. There are multiple aspects involved in language and therefore choices must be made what to assess and how. Specifically concerning the young group of participants in the present study, choices had to be made. In Chapter 2 the theory on communicative competence was explained. The theory of communicative competence emphasizes that language is meant for communication, but that there are many different facets to communication, for which a

person needs different aspects of language. Language therefore should also be assessed in such a way that it captures multiple essential features.

From my experience as a teacher, language in Dutch school settings is assessed mostly with receptive and productive vocabulary tasks in which pupils must point to pictures after hearing a target word or name the picture, assessing the breadth of vocabulary. Word description is also a common task in which pupils must describe a certain word, for example “what is a chair?” A more communicative task according to communicative competence would also involve discourse and thus narratives. However, due to the time-consuming aspect of the evaluation of spontaneous discourse this is a test which is not often performed by teachers. Language sample analyses are mainly used for further diagnostic research only when a teacher has concerns about a pupil’s language development.

The initial goal of this dissertation was to investigate how second language vocabulary developed and thus we wanted to incorporate receptive and productive vocabulary tasks. A receptive vocabulary task was easy to find, but a productive task was harder to decide on. When we encountered a narrative task, we first wanted to analyze the stories solely on the use of vocabulary, but this was later extended to the investigation of general narrative ability. Thus, for the present study both a receptive vocabulary task (the Peabody Picture Vocabulary Task; PPVT; Dunn & Dunn, 2005) as well as a narrative task (Multilingual Assessment Instrument for Narratives; MAIN; Gagarina et al., 2012) were included in the assessment of pupils’ second language proficiency. With these two assessments a more comprehensive picture could be given of the second language development of the newly arrived migrant kindergarteners.

Even though it is acknowledged in the present study that looking at language development through a communicative competence lens will benefit language development assessment, this PhD research project could not cover all segments from the theory of communicative competence due to practical limitations. The choice therefore has been made to include more detailed information on only two segments of communicative competence by means of longitudinal data of the participants: (1) linguistic competence will be discussed in Chapter 4 about receptive vocabulary development and Chapter 5 about narrative ability. (2) Discourse competence will also be discussed in Chapter 5 about narrative ability. Additionally, sociocultural competence is briefly mentioned in Chapter 7. Based on the data that has been collected in the present study many segments of communicative competence could have been discussed, however, time limitations prevented us from analyzing the data on more segments of communicative competence. In Chapter 9 Section 9.5.3 we will include some points of departure how to fully assess communicative competence, including strategic competence.

### 3.3.2 Research design and data collection

Ortega and Iberri-Shea (2005) remark that research on second language development should include “time” and “learning,” this could be done via data collection that is *longitudinal* or *cross-sectional*. Data collection in longitudinal studies used to be carried out with one single subject or a small group of participants, over a prolonged period of time with regular intervals. Longitudinal data usually consist of several aspects of language development and other cognitive development. Although a lot of data is collected in longitudinal studies, it is sometimes difficult to generalize with such a small group of participants. An alternative for that is to do cross-sectional research. In cross-sectional research data is usually collected by means of large groups of participants with different ages at a single point in time. In these studies the focus is mostly on one specific aspect of language development. The idea is that with such a large group one is “able to see a slice of development, which is used to piece together actual development” (Gass & Selinker, 2008, p. 56). However, it is always problematic to make sure that the different age groups are comparable. Both methodologies have advantages as well as disadvantages; fortunately, the distinction is not as rigid as it seems. There is flexibility in categorizing research as cross-sectional or longitudinal (Gass & Selinker, 2008) as is, for example, the case with the Pre-Cool data collection (e.g., Veen, van der Veen, Heurter, & Paas, 2012a) and with our data collection. The Pre-Cool cohort study, which started in 2011, is an example of a longitudinal study with a large group of participants (N = 3000).

The present study is a longitudinal study, with a fairly large group of participants: different kinds of data about the language use of the participants were gathered. The data could be described qualitatively but with additionally specific quantitative analyses, by means of statistical measures.

Gilmore (2016) considers a mixed methods approach most appropriate for classroom-based research projects, like our study, because in classroom-based research an attempt is made to measure changes in a complex construct, composed of multiple, interacting sub-components, and emerging across multiple layers of a complex learning context over an extended period. In our case the complex construct is the second language development, of which we assessed the interacting sub-components receptive vocabulary and narrative ability. Furthermore, this second language development is emerging across multiple layers of a complex learning context, namely the language classroom and we followed the participants over an extended period of two-and-a-half years. To summarize, the present study is longitudinal, with a descriptive, evaluative, and correlational design.

To longitudinally describe the second language development of newly arrived migrant pupils, information was gathered using four primary data sources: (1) performance tests on receptive vocabulary development in Dutch and (2) performance tests on a productive narrative task. Data set (1) and (2) were both conducted with the 42 kindergarteners who participated in the study. Data set (3) includes observational data from

ten teachers whose teacher behavior was observed. (4) The final set of data includes the observations of the pupils' experiences in class. In doing so, this study is a response to Ortega and Ibberrri-Shea's call, who stated that "[...] the diversity and accumulation of recent and future longitudinal research will help chart the development of advanced L2 capacities and help us understand the appropriate timing, duration, and content of optimal educational practices for L2 learning across educational settings and multilingual contexts" (Ortega & Ibberrri-Shea, 2005, 42).

### **3.3.3 Further methodological considerations**

Investigating the learning environment and the second language of young newly arrived migrant pupils longitudinally implies multiple methodological considerations. First, it was decided to use observations in the classroom itself because many educational researchers have argued that such a method is more valid than studying policy documents, relying on program design, relying on teacher qualifications, or using self-report of teachers (e.g., Mashburn et al., 2008). Observing actual practices in the classroom reveals the implemented curriculum instead of the intended curriculum (van den Akker, 2003). Inferring the quality of the school learning environment from individual teaching qualifications also seems less reliable, because it is questionable whether teachers actually act on specific cornerstones of their training (e.g., Henrichs et al., 2017; De Haan, Leseman, & Elbers, 2011; Bulters & Vermeer, 2007). Therefore, the choice in this dissertation has been to conduct observations in the classroom to investigate the school learning environment. The focus of the observations is on the pedagogical practices in the classroom based on observations of teacher behavior and based on the observations of the experiences of the focal pupils: investigation the specific interactions these pupils had with their teachers and peers.

A second consideration was that two levels of narrative ability were included with which the productive abilities of the pupils were analyzed which made the picture of the second language development of the participants more complete: (1) the *microstructural level*: focusing on word-level. What kind of words did pupils use to tell the story? What was the lexical diversity of the words used? And (2) the *macrostructural level*: were pupils able to tell a story with a coherent structure? Did they use all relevant story grammar elements? (e.g., Stein & Glenn, 1979). The pupils were assessed four times over a period of two-and-a-half years with these instruments to capture their development on these various aspects of the second language.

Finally, besides being a researcher, I am a primary school teacher with eight years of experience with newly arrived migrant pupils, including kindergarteners. These experiences also guided me to take into account the workload that teachers experience. I was able to carefully choose instruments which would not add to their workload and with which I would not disturb the routines in the classroom. The pupils were quickly

accustomed to me being part of their class and they had no difficulty with me assessing them in a separate room. As a teacher-researcher, I was confident that I could approach pupils and teachers in such a way that it was appealing for all parties to participate in my study.

### **3.4 Participating schools, teachers, and pupils**

In this section the participating schools, teachers, and pupils will be introduced. These participants will be introduced now because they are the same participants in all four sub-studies of this dissertation. Here we will discuss the participating schools, teachers, and pupils in detail, while in the subsequent chapters we will only briefly discuss the participants with specific tables summarizing the necessary descriptive statistics for that particular chapter.

#### **3.4.1 Participating schools and teachers**

To recruit schools, all schools with Dutch as second language classes for newly arrived migrant pupils (around 300) listed on the website of a national organization for supporting the education for newcomers (named in Dutch LOWAN) were analyzed. It was rather difficult to analyze school guides and websites to figure out whether the school had kindergarten classes for newly arrived migrants or not. We were able to identify around ten of these 300 listed DL2-schools as having separate language facilities for kindergarteners in 2014. All these DL2-schools were approached and asked whether their kindergarteners could participate. Additionally, mainstream schools listed on the website having newly arrived migrant pupils were approached and asked whether they had young newly arrived migrant pupils in their kindergarten classes which would be willing to participate. Multiple schools replied, some schools replied in anticipation of receiving newly arrived migrant pupils; however, when this study began these children had not arrived yet, so these schools did not ultimately participate.

In the beginning of 2014 kindergarten classes in thirteen schools all over the Netherlands, DL2-schools and mainstream schools were visited and preliminary observations were carried out in order for the researcher to get familiar with the different settings and schools. Due to practical reasoning the study continued with a total of ten schools: five DL2-schools and five mainstream schools (see Table 3.2). The schools were located in urban settings in different parts of the Netherlands that varied in population from 70,000 to over 600,000 inhabitants. The DL2-schools were a good representation since almost all known DL2-schools with kindergarteners participated, except for the DL2-schools adjacent to an asylum seeker center. These schools were left outside this study because we could not rely on the continuous participation of these children since their stay in that asylum seeker center was unpredictable. Practical reasons played a role in the

selection of the participating schools because there were limitations to the geographical reach of the investigator, and other time-consuming limitations.

Seventeen female teachers from the ten schools were asked to take part in the study and all consented. Eleven of these seventeen teachers worked at a DL2-school, and six at a Mainstream school. Of the 11 DL2-school teachers 9 worked in an independent segregated DL2-school, two in a DL2-class within a Mainstream school (see Table 3.2). It would have been interesting to see if there were differences between the teachers in the two types of DL2 educational settings, but the distribution was too skewed, therefore both separate schooling types have been taken together in our analysis. The low number of participating teachers in DL2-classes was caused by the fact that it was more likely to have separate kindergarten classes for newly arrived migrant pupils at a DL2-school than at a Mainstream school. Furthermore, DL2-schools most often had multiple kindergarten classes which made it easier to recruit teachers.

Table 3.2: Distribution of Participating Teachers (N = 17) over the two School Types (N = 10).

	Number of Teachers
<b>DL2-schools</b>	
School 1 (class within mainstream school)	1
School 2 (segregated school)	3
School 3 (class within mainstream school)	1
School 4 (segregated school)	3
School 5 (segregated school)	3
Total DL2-schools (5)	11
<b>Mainstream schools</b>	
School 6	2
School 7	1
School 8	1
School 9	1
School 10	1
Total Mainstream schools (5)	6

Two more teachers from two different DL2-schools were excluded from the data collection. One teacher became ill and only returned to the school after all observations were made. The other teacher who was excluded was the direct colleague of the main researcher who would be doing the observations. It was decided not to make observations in the classroom she herself taught two days a week. We expected the pupils in the classroom to behave differently if their regular teacher was present in the classroom but now in the role of researcher and making notes. It would also have been difficult for the researcher herself to focus on what is going on at the moment of observation and not include assumptions based on behavior of pupils on days she was the teacher.

Four out of the eleven teachers at DL2-schools had followed a specific in-service training after their bachelor's in education on second language learning. Six other DL2-school teachers followed different courses on vocabulary teaching/learning<sup>9</sup> within their own school and received coaching. One teacher in a DL2- school did not receive any specific training for teaching newly arrived migrant pupils. Two of the six Mainstream school teachers had followed a similar course like the one from the DL2-schools on vocabulary teaching/learning. The other four did not receive specific education on second language learners/teaching.

At the end of the study more than ten schools were involved in this study, since the participants from DL2-schools transferred to mainstream schools. However, these new mainstream schools were not analyzed on their characteristics of the school learning environment. Around the time of the new round for data collection we approached the former teacher of the participating pupil and asked for details of the new school. We then telephoned or emailed the new school, explained our intentions, and made an appointment for the next round of data collection. All new schools allowed us to come.

### **3.4.2 Demographic characteristics of the participating pupils**

Forty-two pupils (21 girls) participated in the study. They were enrolled in the ten different schools described above. All parents signed a consent form for the participation of their child, which also included a few background questions. At the time of the first data collection the mean age of the participants was 5;1 years old (range 4;0–6;4). Three children started their Dutch education at the end of school year 2013–2014, but the other 39 children started in 2014–2015. At the beginning of the study all pupils were less than a year in the Netherlands. On average they had been at the Dutch school for two months when the data collection began. The number of months a pupil has attended a Dutch school (excluding summer months) is used to measure the Exposure to Dutch in school (see Table 3.3).

---

<sup>9</sup> The course is called “*Met woorden in de weer*” and is based on the methodology by van der Nuft and Verhallen.

Table 3.3: Descriptive Characteristics of the Participants (N = pupils) at the First Data Collection in Session 0 (Age, Gender, and Exposure to Dutch at School).

	N	Age		Gender		Exposure to Dutch at School (in months)	
		Mean	<i>sd.</i>	Boys %	Girls %	Mean	<i>sd.</i>
DL2-school	32	5;1	0;9	53.1	46.9	1.81	1.49
Mainstream school	10	5;2	0;7	40.0	60.0	2.90	1.79
Total	42	5;1	0;8	50.0	50.0	2.07	1.61

Table 3.4 shows the distribution of the participants over the schools in the two school types in this study. As can be seen, there was skewedness in the distribution of the participants over school types and over schools, as a consequence of our ecological setting. Whenever there was a DL2-school willing to participate in this study they could provide multiple pupils from mostly multiple classes. The participating Mainstream schools could only offer small number of pupils. Including more schools, and thus more classes and pupils would have been too time-consuming for the single researcher in this study. In the end, only a fourth of the participants were schooled in a mainstream context. It has to be said that this is not a reflection of the actual situation for newly arrived migrant kindergarteners. As mentioned in Chapter 1, most newly arrived migrant kindergarteners are mainstreamed after arrival, however for practical reasons it was difficult to include more kindergarteners from Mainstream schools.

Table 3.4: Distribution of Participants over the two School Types.

	Number of Classes	Number of Pupils
<b>DL2-schools</b>		
School 1	2	3
School 2	4	7
School 3	1	3
School 4	3	6
School 5	3	13
Total DL2-schools (5)	14	32
<b>Mainstream schools</b>		
School 6	3	7
School 7	1	1
School 8	1	1
School 9 & 10 <sup>b</sup>	1	1
Total Mainstream schools (5)		10

<sup>b</sup>This participant changed schools.

The consent form parents filled in at the start of the study included some background questions about the children, their languages, and their migration history. We collected data on their birth date, country of birth, and the age of arrival in the Netherlands. We also asked which languages the child spoke with his/her mother, father, and siblings. We also asked to provide us with a short migration history stating in which countries the child has lived. Some participants had a refugee background fleeing from war; others were for example labor migrants. The children came from twenty-four different countries (see Table 3.5).

Table 3.5: Country of Origin, the Country the Pupil Lived Longest Before Moving to the Netherlands.

	Frequency	Percent
Poland	8	19.0
Syria	4	9.5
China	3	7.1
Somalia	3	7.1
Latvia, Libya, Romania, Spain	8 (2 each)	19.0 (4.8 each)
Aruba (Dutch Antilles), Belarus, Brazil, Bulgaria, Curacao (Dutch Antilles), Egypt, Germany, Greece, Honduras, Hungary, Indonesia, Italy, Kenya, Philippines, Portugal, and South Korea.	16 ( 1 each)	38.0 (2.4 each)
Total	42	100.0

The main language of communication at home was a language other than Dutch. At home twenty-six different language combinations were found, involving twenty-two different languages (see Table 3.6). Thirteen pupils were reported as having a monolingual language situation at home. The two major languages heard or spoken by the children were Arabic and Polish. For ten pupils it was reported that the mother, father, siblings, or extended family member spoke Dutch (in addition to another language).

Table 3.6: The Different Languages Spoken or Heard by the Participants at Home with the Frequency in the Sample.

Language family	Languages (frequency)
A. Afro-Asiatic	1. Arabic (11) 2. Berber (1) 3. Somali (3)
B. Austronesian	4. Indonesian (1) 5. Filipino (1)
C. Indo-European	6. French (Romance) (1) 7. Italian (Romance) (1) 8. Portuguese (Romance) (2) 9. Papiamentu (Portuguese Creolean) (2) 10. Romanian (Romance) (2) 11. Spanish (Romance) (3) 12. Dutch (Germanic) (10) 13. English (Germanic) (3) 14. German (Germanic) (1) 15. Bulgarian (Balto-Slavic) (2) 16. Latvian (Balto-Slavic) (1) 17. Polish (Balto-Slavic) (8) 18. Russian (Balto-Slavic) (3)
D. Koreanic	19. South Korean (1)
E. Niger Congo	20. Kiswahili (1)
F. Sino Tibetan	21. Chinese (4)
G. Uralic	22. Hungarian (1)

For most of the population, 35 children, Dutch was their L2. Four children learned Dutch as their L3. Additionally, three children learned it as L4, which means that seven children were already multilingual before learning Dutch.

The skewness and the diversity of the population is a concern for this study, but it is also the reality for school populations nowadays. In the next two chapters, Chapter 4 and 5, the language development of these participants will be reported. The language scores for a receptive vocabulary task and a narrative task will be reported over a period of two-and-a-half years in relation to school type, as based on organizational structure. In Chapter 6 and 7 the actual pedagogical practices in the learning environment in the ten schools will be reported.



## CHAPTER 4

---

### **Development of receptive vocabulary of newly arrived migrant kindergarteners**

---

## 4.1 Introduction

This dissertation was set up to evaluate the possible impact of the organization of educational settings on the second language development of newly arrived migrant kindergarteners. In order to do so, first the pupils' language development had to be investigated. In this chapter, the focus is on receptive vocabulary, and Chapter 5 focusses on narrative ability. The combination of these two methods of assessment is related to the concept of communicative competence. Communication Competence aims to assess language on the basis of four segments: linguistic competence, sociocultural competence, discourse competence, and strategic competence (following from Canale, 1983; Celce-Murcia, Dörnyei, & Thurrell, 1995; see Chapter 2). Receptive vocabulary relates to the first segment, linguistic competence.

Receptive vocabulary is defined as the number of different words (verbs, nouns, and adjectives) a child can identify, as measured by the Peabody Picture Vocabulary Test for Dutch (PPVT-III-NL; Schlichting, 2005). The PPVT is a standardized receptive vocabulary test that measures vocabulary size. In the PPVT, the participant hears a target word and has to choose the correct referent out of a set of four pictures. Productive vocabulary, in contrast, is the number of different words a child can produce orally, or that older pupils can write. In Chapter 5, productive vocabulary will be measured by assessing the narrative ability of forty-two pupils using the Multilingual Assessment Instrument for Narratives (abbreviated as MAIN; Gagarina et al., 2012). Following the protocol, the participants need to tell a story by means of a six-picture sequence.

The focus of the present chapter is the newly arrived migrant pupils' receptive vocabulary development and the possible differences in this development between pupils in DL2-schools and Mainstream schools. The following research question is central:

How does the Dutch *receptive vocabulary* (in Dutch) of newly arrived migrant kindergarteners develop during the first years of schooling in the Netherlands related to school type?<sup>10</sup>

The present chapter is structured as follows: first, we give a brief review of the literature of what is already known about receptive vocabulary development of newly arrived migrant pupils; then we discuss the data collection; the methodology and results of data collected from forty-two participants is presented. The chapter ends with a discussion, which will be extended in Chapter 9.

---

<sup>10</sup> In Chapter 3 this research question was written as sub-question (a).

## **4.2 Receptive vocabulary**

The studies we outlined in Chapter 2 on receptive vocabulary development indicate that vocabulary is essential for academic development. However, second language learners often have a vocabulary gap compared to their peers because they must divide their language learning time over multiple languages. This vocabulary gap implies an academic catch up phase. Research has indicated that this catch up phase may take 5 to 7 years (Collier & Thomas, 1989; Cummins, 1981). In fact, several studies into second language development carried out in the Netherlands show that second language learners of Dutch have a smaller Dutch vocabulary than first language learners. Bilingual children consistently scored below monolingual children on Dutch vocabulary test, and even though the bilingual children showed a faster Dutch vocabulary development than the monolingual Dutch children, they still lagged behind at the age of six (Scheele, 2010; See also other Dutch studies like Strating-Keurentjes, 2000; Leseman, Henrichs, Blom, & Verhagen, 2019).

Additionally, according to Driessen (1996) and Verhoeven and Vermeer (1996), the difference in Dutch vocabulary persists until at least the end of primary education – and according to some researchers the difference in Dutch vocabulary size between L1 and L2 learners even increases throughout their educational career.

To understand how vocabulary develops in second language learners, and how they accelerate the catch up phase it is important to focus on vocabulary development in young second language learners and from very early on. Furthermore, it is important to understand which factors influence vocabulary development to support the pupils in their academic catchup phase. When variables are discovered that influence the vocabulary development of newly arrived migrant kindergarteners these variables can be taken into account in education.

## **4.3 Method**

### **4.3.1 Participants**

In this study 42 pupils (21 girls) participated who, in the beginning, were enrolled in ten different schools (see Chapter 3 section 3.4 for details about the participants and schools). Due to their transfer from DL2-schools to mainstream schools the participants were enrolled in 35 schools at Session 3. The participants were tested four times. All parents signed a consent form for the participation of their child and filled in a questionnaire with a few background questions (see Chapter 3, Section 3.4.2).

Three variables concerning child characteristics were considered: Age, Exposure to Dutch at School, and Educational Facility in order to find out whether the development of pupils differed between school types. The Age of the pupils and Exposure to Dutch at School are correlated, given that every month a child grows older the Exposure to Dutch at

School grows. Nevertheless, this variable allows us to understand differences between children with similar amounts of exposure but different ages and between children with a similar age but with different amounts of exposure. The variable Educational Facility has two values: “segregated,” meaning a DL2-school (segregated independent DL2-school or a DL2-class within a mainstream school) or “inclusive,” meaning a Mainstream school.

Tables 4.1 and 4.2 provide descriptive characteristics of the participants in this study, per session. Session 0 is the first assessment, but since it was fairly in the binning of the language learning process we called it Session 0 to indicate the null-measurement. The division of participants is skewed, with 32 participants at DL2-schools compared to 10 participants at Mainstream schools. This skewedness is caused by the fact that it was easier to recruit pupils at DL2-schools compared to Mainstream schools since there is a concentration of newly arrived migrant pupils at DL2-schools. At Mainstream schools, in contrast, most of the time only one pupil was eligible to participate in the study.

Table 4.1: Descriptive Characteristics of the Participants (N = number of pupils) in the Receptive Vocabulary Assessments. Mean (Standard Deviation) and Age Range (in years; months) per Session.

	Session 0			Session 1			Session 2			Session 3		
	Age Range		Mean ( <i>sd</i> )	Age Range		Mean ( <i>sd</i> )	Age Range		Mean ( <i>sd</i> )	Age Range		Mean ( <i>sd</i> )
	Min	Max		Min	Max		Min	Max		Min	Max	
DL2-schools (N = 32)	4;2	6;4	5;1 (:9)	4;7	6;10	5;6 (:9)	5;2	7;8	6;4 (:9)	6;2	8;8	7;4 (:9)
Mainstream schools (N = 10)	4;0	5;9	5;2 (:7)	4;6	6;2	5;7 (:7)	5;6	6;10	6;4 (:6)	6;6	7;11	7;4 (:6)
Total (N = 42)	4;0	6;4	5;1 (:8)	4;6	6;10	5;7 (:8)	5;2	7;8	6;4 (:8)	6;2	8;8	7;4 (:8)

Note: the shaded cells indicate that participants were attending a mainstream school. The participants in the DL2-schools group started in a DL2-school, but Session 2 took place approximately 6 months after their transfer from a DL2-school to a mainstream school.

Table 4.2: Descriptive Characteristics of the Participants in the Receptive Vocabulary Assessments at Session 0.

	Gender		Exposure to Dutch at School (in months)	
	Boys (%)	Girls (%)	Mean	<i>sd.</i>
DL2-school (N = 32)	53.1	46.9	1.81	1.49
Mainstream school (N = 10)	40.0	60.0	2.90	1.79
Total (N = 42)	50.0	50.0	2.07	1.61

### 4.3.2 Measurement

In the present study the PPVT-III-NL (Schlichting, 2005) was used to measure pupils' receptive vocabulary development. The PPVT, a standardized test, is available in many languages, and is used worldwide to evaluate word knowledge of participants from the age of 2;3 to 90 years old. The PPVT can be used with monolingual and bilingual children; however, the standardization has been based on monolingual norms only. The PPVT consists of 204 test sheets with four pictures each. The participant hears a word and is asked to choose the corresponding picture.

In order not to interfere with other language tests used in school, we chose a test that had not been administered to the participants before. The PPVT was chosen in our study because it is a test that is not commonly used in Dutch schools and the PPVT seemed to be suitable as a curriculum-independent instrument.

### 4.3.3 Procedure

All participants were assessed four times using the PPVT: Session 0, 1, 2, and 3. The first two sessions took place within the first year after the child had arrived in the Netherlands. The mean time interval between Session 0 and Session 1 was approximately 6 months, between Session 1 and 2 approximately 9 months, and between Session 2 and 3 it was 12 months (see Table 4.1). For participants at DL2-schools, Session 2 took place about 6 months after their transfer to a mainstream school. Thus, in the final two sessions no participant attended a DL2-school anymore. For participants already at a mainstream school the date of Session 2 had been set on about 1.5 years after the beginning of their school career in the Netherlands.

All pupils were assessed individually in a separate room at their school. The PPVT was administered using a laptop computer with touch screen (see Figure 4.1 for the set-up). For each item the pupils saw four pictures on the lap top screen, and after hearing the target word they were asked to point to the picture that matched the target word best.

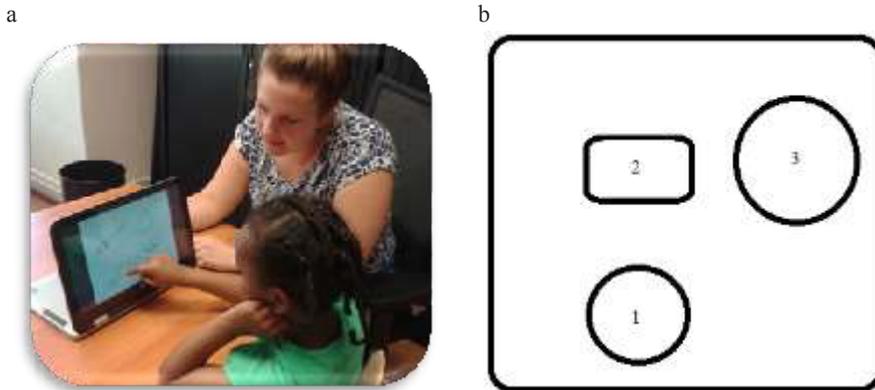


Figure 4.1a and 4.1b: Picture and Graphical Representation of the Assessment Set-up. 1: participant; 2: laptop; 3: researcher.

The Dutch version of the PPVT has 180 items divided into 15 sets of 12 words increasing in difficulty to adapt to different age groups. Following the standard approach, a test session starts with the set appropriate for the pupil's age at the time of testing. When the pupil has 5 or more incorrect answers on the starting set, a lower, easier set is administered until less than 4 mistakes are made. After that, the test session continues with the next higher set. The test stops when a pupil has made 9 or more mistakes in the last set. For example: a participant of 4;3 years old needs to start in Set 4 (the age appropriate level for participants between 4;0 and 4;5). However, when he or she makes five mistakes Set 3 is administered. In this set the child makes 3 mistakes and then the assessment is continued with Set 5. The test ends after Set 6 because the participant made 9 mistakes.

However, because the pupils in the present study had less than a year of exposure to Dutch, we began our assessment in Session 0 for all pupils regardless of their age with the set with the lowest level, normally used for participants aged 2;3-2;5. We then continued with the next higher set until they made nine or more mistakes, the break off point from the standard protocol. Nine of the forty-two pupils made 9 errors or more before their age appropriate set was reached. At the second assessment, we used age-appropriate set as starting set.

#### **4.3.4 Analyses**

##### **Statistical analysis**

The PPVT provides a raw score, all correct answers, which can be converted to a standardized score from a table providing age-corrected normative scores. In our analysis of the receptive vocabulary development however, we used the raw scores of the PPVT instead of the standard scores, following the reasoning of Golberg, Paradis, and Crago

(2008). Normally, raw scores are adjusted for different ages however, newly arrived migrant kindergarteners could have the same amount of exposure to Dutch even though their ages differ, and thus standard scores adjusted for age will put older children at a greater disadvantage (see also Jacobs, 2016).

Our measurements are nested within pupils, and observations of the same pupil are of course more alike than observations of different pupils. Due to these dependencies, the most appropriate method of analysis is considered multilevel modeling (see Quené & van den Bergh, 2004). Our 42 participants provided us with 168 data points. For this purpose, multilevel modeling of repeated measures data procedures (MLwiN software version 2.36; Rasbash, Charlton, Browne, Healy, & Cameron, 2016) was carried out.

The mean development of PPVT scores was modelled by fitting different polynomial functions to the data. In the first model it was assumed there is no growth, in the second model it is assumed that scores change (on average) linearly with age, and in the third model it is assumed that the relation between scores and age deviates from linearity. In addition to these differences in mean scores, we also modelled the variance within and between pupils. Multilevel modeling allows for heteroscedasticity of variances at each level, therefore the variance within and between pupils was modelled in a series of subsequent models.

In general polynomials are very flexible functions and can take almost any shape, depending on the number of parameters. In the models, both the fixed and random components increased in complexity. If  $y_{ij}$  is the score on the  $i$ th age (the age in months) of the  $j$ th individual, then a polynomial can be written as:  $y_{ij} = f_j(\text{age}_{ij})$ . This function can be written as a regression model, which assumes that the latencies depend on powers of age:

$$y_{ij} = \beta_{0ij} + \beta_1 * \text{Age}_{ij}^1 + \beta_2 * \text{Age}_{ij}^2 + \dots$$

It is an empirical matter to determine which order a polynomial will take in a given data set. Generally, the most parsimonious is chosen (Van den Bergh, Schmittmann, Hofman, & van der Maas, 2015), the model that explains the variation in outcome scores the best.

The development of our receptive vocabulary score was tested with a maximum of 7 models. Model 0 was the basic null model in which the PPVT score was allowed to vary within and between pupils. In Model 1, Age at testing was added as a fixed main effect to test whether average scores differed over time. Whether the differences within and between pupils depend on Age was tested in respectively Model 2 and 3. Model 4 tested the main effect of  $\text{Age}^2$  and Model 5 and 6 tested whether there was variance within or between pupils in the exponential effect of Age.

Finally one of the seven models was assigned as the best fit and seen as the General Development Model. Each score of the fit of each model, along with the difference

in fit between consecutive models are presented in the tables with the Likelihood Ratio Tests and expressed by  $-2 \log$  likelihoods.

In the subsequent sections and chapters, the General Development Model will be extended to include the child characteristics and the characteristics of the school learning environment. The explanatory variables in this study are added separately one by one, and not in combination, because this way we can make our models more reliable. Due to our sample size with a small number of participants and a limited number of observations per participant we need to be extra careful.

#### 4.4 Results

The results of the receptive vocabulary development are presented as follows: first, the developmental data from PPVT is presented to answer the question: How does the *receptive vocabulary* (in Dutch) of newly arrived migrant kindergarteners develop during the first two-and-a-half years of schooling in the Netherlands in relation to school type? The basis of the analysis is the General Development Model, which includes only Age at which receptive vocabulary was measured as an explanatory variable. Second, the variables Exposure to Dutch at School and Educational Facility are introduced into the model.

To describe the development of the receptive vocabulary scores, several models were fitted. Models will be built but no tables with mean scores will be presented since the participants differ in too many aspects. That is, an overview of descriptive statistics will not be informative, whereas the models will be. In the tables that are presented, the increase in fit in relation to the previous model is shown.

From the comparison between the consecutive models (see Table 4.3) it is apparent that a model with a fixed linear component – allowing for differences in Age – fit the data better than a model with only an intercept ( $\Delta\chi^2$  (PPVT<sub>1</sub>) = 125.87;  $df$  = 1;  $p$  < .001). The variance within individuals depends on the age of participants ( $\Delta\chi^2$  (PPVT<sub>2</sub>) = 24.54;  $df$  = 2;  $p$  < .001). The variance between individuals is a (linear) function of age as well ( $\Delta\chi^2$  (PPVT<sub>3</sub>) = 4.00;  $df$  = 1;  $p$  = .045). Finally, adding Age<sup>2</sup> to the model improved the fit significantly ( $\Delta\chi^2$  (PPVT<sub>4</sub>) = 8.18;  $df$  = 1;  $p$  = .004). Hence, in the final model (PPVT<sub>4</sub>) a fixed effect of Age<sup>1</sup> and Age<sup>2</sup>, as well as variance within and between pupils components which depends on Age needs to be included, and with this model we continued the analysis.

Table 4.3: Fit of Different Polynomials (-2LL) for Changes in PPVT Score (168 cases) as well as the Comparison of Consecutive Models.

Model	-2LL	Comparison			
		Models	$\Delta\chi^2$	<i>Adf</i>	<i>p</i>
PPVT <sub>0</sub> : $\beta_{0ij}cons^a$	1540.51				
PPVT <sub>1</sub> : PPVT <sub>0</sub> + $\beta_1 Age^1_{ij}{}^b$	1414.63	PPVT <sub>0</sub> vs PPVT <sub>1</sub>	125.87	1	<.001
PPVT <sub>2</sub> : PPVT <sub>1</sub> + $e_{1ij} Age^1_{ij}$	1390.09	PPVT <sub>1</sub> vs PPVT <sub>2</sub>	24.54	2	<.001
PPVT <sub>3</sub> : PPVT <sub>2</sub> + $u_{10j} Age^1_{ij}$	1386.09	PPVT <sub>2</sub> vs PPVT <sub>3</sub>	4.00	1 <sup>c</sup>	.046
PPVT <sub>4</sub> : PPVT <sub>3</sub> + $\beta_2 Age^2_{ij}$	1377.92	PPVT <sub>3</sub> vs PPVT <sub>4</sub>	8.18	1	.004
PPVT <sub>5</sub> : PPVT <sub>4</sub> + $e_{2ij} Age^2_{ij}$	1375.17	PPVT <sub>4</sub> vs PPVT <sub>5</sub>	2.74	3	.43ns
PPVT <sub>6</sub> : PPVT <sub>5</sub> + $u_{20j} Age^2_{ij}{}^d$					

<sup>a</sup> In addition to the intercept, variance components for differences within and between individuals are estimated.

<sup>b</sup> In order to circumvent numerical issues in parameter estimation, the explanatory variable age was centered around the grand mean.

<sup>c</sup> Only the covariance-coefficient was estimated.

<sup>d</sup> No convergence within 250 iterations.

Based on this General Development Model both the average development and the differences within and between individuals are represented in Figure 4.2 (see Table 1 in Appendix 1 for the parameter estimates). The average receptive vocabulary at an age of 72 months was estimated as 64.39. With each month a child grew older, his receptive vocabulary increased by 1.60, but the quadratic function of Age decreased the score by 0.14.

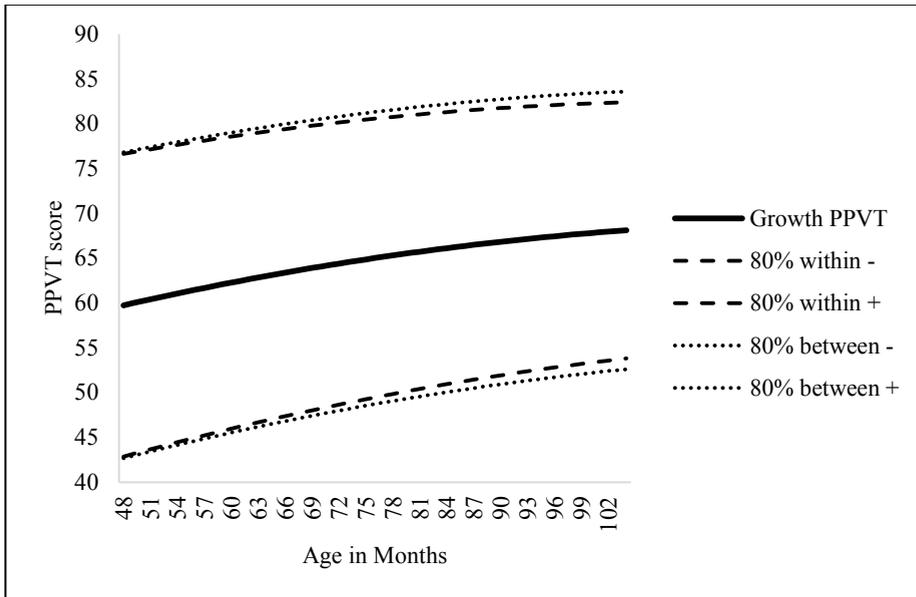


Figure 4.2: Graphical Representation of the General Development Model of PPVT with 80% Reliability<sup>11</sup>.

Figure 4.2 shows that the effect of Age on PPVT scores is significant. The growth shows a quadratic slope which means that the increase in vocabulary scores is larger for younger pupils compared to older pupils. The variance within pupils (represented by the striped lines in Figure 4.2) also depends on age; we can give a more precise estimate of the vocabulary development for older pupils as compared to younger pupils. The variance between individuals (represented by the dotted lines in Figure 4.2) is a function of Age as well; the variance between older pupils is smaller than the variance between younger pupils.

Then, Exposure to Dutch at School, measured as the number of months a pupil has attended a Dutch school (excluding summer months), was added to the General development Model in order to see what the effect of Exposure to Dutch at School was. A likelihood ratio test showed that the main effect of Exposure to Dutch at School contributed significantly to the fit of the model to the observed data ( $\Delta\chi^2$  (PPVT<sub>7</sub>) = 24.71;  $df = 1$ ;  $p < .001$ ). The same holds for the interaction between Age and Exposure to Dutch at School ( $\Delta\chi^2$  (PPVT<sub>8</sub>) = 50.17;  $df = 1$ ;  $p < .001$ ; see Table 4.4).

<sup>11</sup> The striped lines indicate the within pupils variance of the score. The dotted lines indicate the between pupils variance of the score. With an 80% reliability the scores would fall within these areas.

Table 4.4: Fit of Different Models (-2LL) for Changes in PPVT score (168 cases) with Addition of Months of Exposure to Dutch at School (EDS) as Explanatory Variable.

Model	-2LL	Comparison			
		Models	$\Delta\chi^2$	$\Delta df$	$p$
PPVT <sub>4</sub> : $\beta_{0ij}cons + \beta_{1ij}Age_{ij}^l + \beta_{2ij}Age_{ij}^2$	1377.92				
PPVT <sub>7</sub> : $PPVT_4 + \beta_3EDS_{ij}$	1353.21	PPVT <sub>4</sub> vs PPVT <sub>7</sub>	24.71	1	<.001
PPVT <sub>8</sub> : $PPVT_7 + \beta_4Age*EDS_{ij}$	1303.04	PPVT <sub>7</sub> vs PPVT <sub>8</sub>	50.17	1	<.001

In Figure 4.3 the general development for pupils with different amounts of Exposure to Dutch at School is presented. Figure 4.3 might be difficult to interpret since Exposure to Dutch at School and Age are related: every month a child grows older it also has had more exposure to Dutch. However, there was a fixed main effect of Exposure to Dutch at School. Overall, pupils with a larger Exposure to Dutch at School had higher PPVT scores, indicated by the solid black line. Furthermore there was an interaction effect between Age and Exposure to Dutch at School, which means that the relation between Age and PPVT scores differs per amount of Exposure (see Table 1 in Appendix 1 for the estimated parameters). The negative interaction between Age and Exposure to Dutch at School means that the effect of Exposure to Dutch is larger for younger pupils than for older pupils. In Figure 4.3 the general development of PPVT with Exposure to Dutch as explanatory variable for pupils with different ages is presented (see Table 1 in Appendix 1 for the estimated parameters). There was a fixed main effect of Exposure to Dutch at School and an interaction effect between Age and Exposure to Dutch at School. Therefore, all lines in Figure 4.3 show an increase (remember that when pupils grew older, their Exposure to Dutch also increased one month), however, since the effect of Exposure is smaller for older pupils, the striped and dotted line for the oldest pupils shows less difference with the solid black line (the hypothetical 0 Exposure line) compared to the lines of the youngest and average aged pupils.

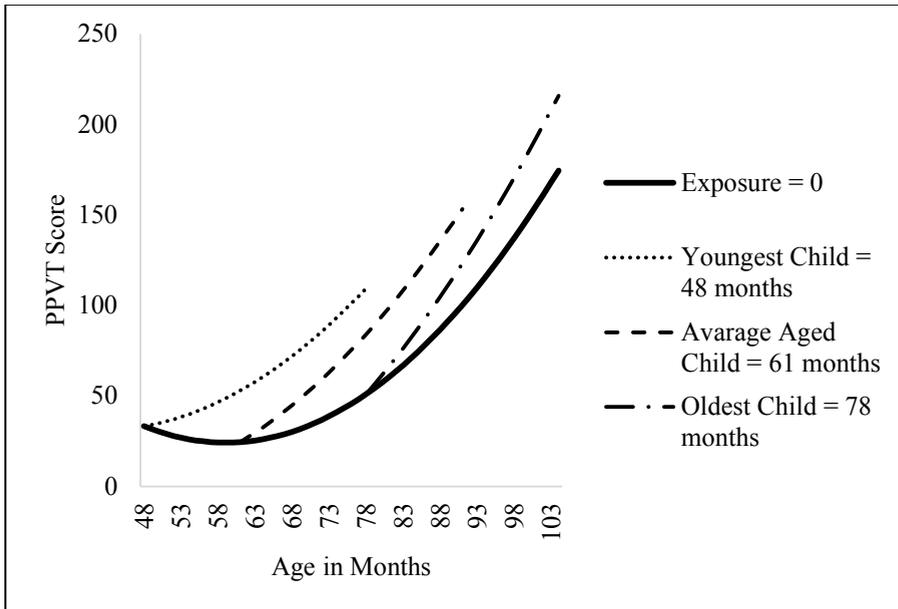


Figure 4.3: Development of PPVT Score with Exposure to Dutch at School as Explanatory Variable.

When the variable Educational Facility was added to the general model of the PPVT development, neither a fixed main effect of Educational Facility, nor an interaction between Age or Age<sup>2</sup> and Educational Facility improved the fit of the model. This means that we could not find differences between scores, nor between development in scores of pupils from DL2-schools and pupils from Mainstream schools.

#### 4.5 Summary and conclusion

The main goal of this chapter was to analyze how the Dutch receptive vocabulary of young newly arrived migrant kindergarteners developed in the first two-and-a-half years after arriving in the Netherlands, answering the research question: How does the Dutch *receptive vocabulary* (in Dutch) of newly arrived migrant kindergarteners develop during the first years of schooling in the Netherlands related to school type?

Table 4.5 summarizes the significant effects of the four variables. The answer to the research question is that, as expected, the receptive vocabulary of the newly arrived migrant pupils grows in the first two-and-a-half year in the Netherlands. Results show that the increase in receptive vocabulary is not linear. The model of growth of receptive vocabulary shows at younger ages a significant steeper growth, but this growth then levels off.

Exposure to Dutch at School proved to improve the General Development Model significantly. The results show, as expected, that the more Exposure to Dutch at School, the higher a pupil's receptive vocabulary score. This effect of Exposure to Dutch at School was larger for younger pupils compared to older pupils.

Table 4.5: Overview of Significant Child Characteristics on the PPVT score.

Age	+
Age <sup>2</sup>	+
Exposure to Dutch at School	+
Educational Facility	-

The other variable included in the receptive vocabulary development model, the educational facility, did not significantly improve the model. We did not find that attending a DL2-school or a Mainstream school from the beginning of preschool per se influences the development of the second language of newly arrived kindergarteners, as measured by the PPVT.

Thus, the school learning environment, whether segregated or inclusive, does not seem to influence the development of receptive vocabulary of newly arrived kindergarteners in the research time frame. If this result were confirmed, this measure, with its narrow definition, would therefore not justify the separation of these children from mainstream education. It is of course necessary to consider the duration of the research, which certainly impacts the results. Maybe in the longer term a difference could appear.

However, it is possible that other aspects of the school learning environment may have an impact on the development of the target language in newly arrived pupils. One of them might be the impact of the teacher. During our visits to the different schools, we noted the adoption of different pedagogies by teachers who did not seem to depend on the learning environment (segregated or inclusive). This is therefore the factor that we will study in Chapter 7. Additionally, we noticed that the population of pupils in one inclusive school was not comparable to the population of pupils in the other inclusive school, which made us realize that we should also zoom into the school learning environment on the level of the pupils' experiences with teachers and peers. This will be further investigated in Chapter 6.

Now that we know how receptive vocabulary develops we will continue in Chapter 5 with the analysis of the development of another aspect of the second language. In Chapter 5 we will investigate the development of narrative ability of newly arrived migrant pupils and also look at whether this development differs per school type. We did not find an effect of school type on receptive vocabulary, but there could be an effect of school type on narrative ability. Furthermore, including another aspect of language into the assessment of the newly arrived migrant kindergarteners, answers the call of the theory of communicative competence. The main idea of that theory is to assess language based on the four segments

defined by Canale (1983; see Section 2.3 in Chapter 2) to get a more complete picture of the second language of newly arrived migrant pupils. While this chapter on receptive vocabulary focused on linguistic competence, the next chapter on narrative ability, Chapter 5, will focus on both linguistic competence and discourse competence.



## CHAPTER 5

---

### **Development of narrative ability of newly arrived migrant kindergarteners**

---

## 5.1 Introduction

The overall goal of our overarching study is to explore the possible influence of the school learning environment on the second language development in the early years of schooling in the new country. First, the second language development itself is investigated in more detail. In the initial stages the focus of this dissertation was on vocabulary development. The selection of a productive vocabulary task was challenging, as we aimed at incorporating a truly communicative task. Therefore, in the sub-study in this chapter, we investigated the development of narrative ability of the same forty-two newly arrived migrant kindergarteners as in Chapter 4. By adding data on narrative ability to receptive vocabulary development data, we aim to cover the whole communicative competence theory as defined by Canale, (1983) and Celce-Murcia et al., (1995; see Chapter 2) namely these four segments: linguistic competence, sociocultural competence, discourse competence, and strategic competence. While receptive vocabulary relates to linguistic competence, narrative ability taps not only into linguistic competence but also into discourse competence, sociocultural competence, and strategic competence.

Moreover, when including children with multilingual repertoires, researchers should be careful in the selection of their tasks. According to different researchers (e.g. Cleave, Girolametto, Chen, & Johnson, 2010; Boerma, Leseman, Timmermeister, Wijnen, & Blom, 2016; Paradis, Genesee, & Crago, 2011) narrative tasks are less biased for multilingual participants because narratives do not only tap into language specific knowledge; Narratives ask for general skills such as cognitive, social, and pragmatic skills (Liles, 1993).

Narrative ability can be measured by different types of tasks. Namely, narratives can be recorded in different ways: spontaneous or elicited. Since this present study aims to investigate the development of narrative ability there was a need for a relatively controlled setting which would enable a comparable analysis of the stories between participants and over time within one participant. Therefore, the data for this study on narrative development was collected using the Multilingual Assessment Instrument for Narratives (abbreviated as MAIN; Gagarina et al., 2012). The MAIN consists of picture sequences which are used to elicit speech in a more controlled way compared to spontaneous non-guided speech. In the present study the Dutch version of the MAIN is used.

The foci of the present chapter are (1) the newly arrived migrant pupils' narrative ability development and (2) the possible differences in this development between pupils in DL2-schools and Mainstream schools. The following research question is central:

How does the *narrative ability* (in Dutch) of newly arrived migrant kindergarteners develop during the first two-and-a-half years of schooling in the Netherlands in relation to school type?<sup>12</sup>

The present chapter is structured as follows: before discussing the data collection, a brief review is given of what is already known about narrative ability development of second language learners using the MAIN. Then the methodology and results of data collected from 42 participants is discussed. The chapter ends with a discussion, which will be extended in Chapter 9.

## **5.2 Narrative ability development**

### **5.2.1 Microstructure and macrostructure**

In Chapter 2 narrative ability development was discussed extensively; what follows here is a summary. Narrative ability is the ability to tell a story, and it can be analyzed from two perspectives: the microstructural level and the macrostructure level. Microstructure refers to how words and sentences work together to build a story. Semantic and syntactic productivity as well as complexity and accuracy are the basis on which microstructure is determined (Appose & Karuppali, 2018). Narrative macrostructure refers to the global organization of a story beyond the word, sentence, or utterance level (Blom & Boerma, 2016). Macrostructure can be analyzed using so-called a story grammar, which identifies components such as the setting and a logical event structure (e.g., initiating event, internal response, plan, action, consequence, reaction; Stein & Glenn, 1979).

Macrostructural performance is seen as less language dependent than microstructure performance, because it is suggested that macrostructure is partly dependent on cognitive schemas (Iluz-Cohen & Walters, 2012; Pearson, 2002). Microstructure needs to be newly acquired for every language because every language has different vocabulary necessary to tell a story.

### **5.2.2 Measuring narrative ability with the MAIN**

With the MAIN protocol, data on microstructure level as well as macrostructure level can be obtained at the same time. Since the publication of the MAIN, several studies have been published using this instrument. However, caution is necessary when comparing the results of studies using the MAIN, as is demonstrated in the overview in Pesco and Kay-Raining Bird (2016) of the first eight published studies using the MAIN. The studies were difficult to compare because there was a lack of uniformity in the use of the instrument and there were differences in the populations. Even though studies using the MAIN are hard to

---

<sup>12</sup> In Chapter 3 this research question was written as sub-question (b).

compare we will now report on some findings which can be used as references for our study.

### **Difference between narrative ability in L1 and L2**

Most studies using the MAIN as an instrument investigated the difference between macrostructure in L1 and L2 stories. In some studies, a difference was found between the two languages of the participants. For example, Kapalková, Polišíenská, Marková, and Fenton found that Slovak-English children's (age 5 to 6) macrostructure scores were higher in their L1 compared to their L2, but not for the number of goals and attempts (Kapalková et al., 2016;). Others did not find a difference between L1 and L2 (For Hebrew-English children aged 5;6–6;6: Altman, Armon-Lotem, Fichman, & Walters, 2016; For Finnish-Swedish children aged 5;0–6;7: Kunnari, Välimaa, & Laukkanen-Nevala, 2016). Altman et al. (2016) found that there was no difference in macrostructure score between L1 and L2, furthermore, length of exposure to the new language showed no influence on macrostructure. The fact that most macrostructure components do not differ between the two languages of a participant confirms the claim of the authors of the MAIN (Gagarina et al., 2012) that a narrative task is less biased for multilingual pupils, which is reassuring for the present study.

One study that did find differences in macrostructure was Gagarina (2016). Gagarina reported differences in Structural Complexity between the German and Russian stories of Russian-German bilinguals. Russian-German bilinguals told stories that were more complex in Russian compared to their German stories. Gagarina argued that since the children in Russian received explicit instruction about all aspects of Story Structure, they scored better in that language. For some reason the transfer of knowledge on macrostructure in Russian to German seemed to be delayed. Gagarina (2016) stressed the fact that narrative ability should be investigated through multiple macrostructure measures, because it seems that different components do not have an identical developmental trajectory. Furthermore, Gagarina used story length as a microstructural measure, but she concluded that a more in-depth investigation of microstructural elements would be appropriate.

### **Development of narratives**

Whereas the studies in the previous section focused on differences between stories in L1 and L2, other studies using the MAIN focused on the development of different narrative ability components with a cross-sectional design. For example, Bohnacker (2016) compared stories from Swedish-English 5-year-old bilinguals with stories from 6- to 7-year olds, with a minimum of two years of exposure to the L2. The younger group mainly produced attempts, and internal state terms were used rarely. The older participants did not perform on ceiling level either: goals were still not mastered, but the macrostructure did differ between these age groups, the older children told more complex stories. Gagarina

(2016) also showed significant improvement of macrostructure scores but in her study for Russian-German participants. She used a cross-sectional design with preschoolers, first graders and third graders. There were significant differences between all the age groups, although the difference between the preschoolers and first graders was larger than the difference between first graders and third graders. Gagarina thus showed that narrative ability still improves between these ages.

Also, Maviş, Tunçer, and Gagarina, (2016) found that their younger group of Turkish-German children (age 2;11–3;11) and middle group (age 4;0–5;11) had significantly lower scores on story complexity than the older group (6;0–7;11). Interestingly, Story Structure and Internal State Terms did not show significant differences between the age groups. In the second study of Maviş et al. there was no age effect visible, but in that study, there were only two age groups with an age range of 5;5–7;0 and 7;1–7;11, respectively. Thus, it seems that between the ages of five and seven narratives reach some kind of ceiling level.

Based on the theoretical perspectives and previous studies concerning narrative ability development we hypothesized that given the age of the participants ( 4 to 6 years old) and the fact that we followed them for two-and-a-half years, we will see development of the macrostructure, as Bohnacker (2016) and Gagarina (2016) have shown. The results of our final assessment will show if the participants will reach a ceiling level at the end of our study, as was suggested by Maviş et al. (2016). However, it might be that the participants in the present study will show a different pattern of development, since their exposure to the second language is smaller than that of the participants in the previous studies with the MAIN.

Furthermore, including multiple aspects of macrostructure in the present study might reveal whether the different macrostructural measures have identical developmental trajectories or not, as was suggested by Gagarina (2016). The inclusion of multiple microstructural components also contributes to a more in-depth study of pupils' linguistic competence.

All in all, the present study will use the MAIN as a measure for productive second language development, since it is considered to be a valid measure for second language learners. However, the goal of this study is not to replicate any previous study using the MAIN, but to contribute to the growing database of studies using the MAIN. The present study may be a valuable addition because we used the MAIN longitudinally, and it was used with pupils with only a small amount of exposure to the language of assessment.

## **5.3 Method**

### **5.3.1 Participants**

In this part of the study concerning the narrative ability of newly arrived migrant kindergarteners the same 42 pupils (21 girls) took part as in Chapter 4, which reported their receptive vocabulary development. At the beginning of the study, at Session 0, they were enrolled at 10 different schools. Due to their transfer from DL2-schools to mainstream schools the participants were enrolled in 35 different mainstream schools at the time of Session 3. All parents signed a consent form for the participation of their child and filled in a questionnaire with a few background questions.

Table 5.1 provides the descriptive characteristics of the participants in this sub-study on narrative ability per session. The statistics for Session 0 and Session 1 in table 5.1 are slightly different from Table 4.1 in Chapter 4 with the descriptive statistics of the participants during the receptive vocabulary assessments of the same pupils, because the first two assessments of the narrative ability were not on the same day as the assessments of the receptive vocabulary. The descriptive statistics of Sessions 2 and 3 from the narrative assessment and the receptive vocabulary assessment are directly comparable because these two assessments took place on the same day.

Table 5.1: Descriptive Characteristics of the Participants During the Narrative Ability Assessments. Mean Age (standard deviation) and Age Range (in months) per Session.

	Session 0			Session 1			Session 2			Session 3		
	Age			Age			Age			Age		
	Min	Max	Mean ( <i>sd</i> )	Min	Max	Mean ( <i>sd</i> )	Min	Max	Mean ( <i>sd</i> )	Min	Max	Mean ( <i>sd</i> )
DL2-school (N = 32)	4;4	6;6	5;3 (:9)	4;6	6;1	5;6 (:9)	5;2	7;8	6;4 (:9)	6;2	8;8	7;4 (:9)
Mainstream school (N = 10)	4;3	6;1	5;3 (:8)	4;6	6;2	5;6 (:7)	5;6	6;10	6;4 (:7)	6;6	7;11	7;4 (:7)
Total (N = 42)	4;3	6;6	5;3 (:8)	4;6	6;2	5;6 (:8)	5;2	7;8	6;4 (:8)	6;2	8;8	7;4 (:8)

Note: the shaded cells indicate that participants were attending a mainstream school. The participants in the DL2-schools group started in a DL2-school, but Session 2 took place approximately 6 months after their transfer from a DL2-school to a mainstream school.

### 5.3.2 Measurement

The data for this study on the development of narrative ability was collected using the Multilingual Assessment Instrument for Narratives (MAIN; Gagarina et al., 2012). In the present study a Dutch version of the MAIN<sup>13</sup> was used. The MAIN consists of four comparable stories: two for story retelling and two for story generation. In a retelling task the participant looks at the picture sequences while the researcher models the story. After listening to the model story the participant is encouraged to retell the story. In the story generation task the participant has to tell his or her own story based on a new picture sequence. All the stories are controlled for cognitive and linguistic complexity. Furthermore, the stories are parallel in microstructure and macrostructure and the stories are controlled for cultural appropriateness and robustness according to the authors (Gagarina et al., 2012).

The different stories of the instrument were created by the authors of the MAIN to be able to test each participant in more than one language. Since the current study only tested narrative ability in one language, the different versions were used instead to minimize learning effects over the longitudinal study. Each participant saw the stories twice over the period of two-and-a-half years. Figure 5.1 and 5.2 show the two versions of the story generation stories that were used in the current study (the picture sequences of the retell stories are not displayed since they are not the focus of the present study).

---

<sup>13</sup> The Dutch version of MAIM has been translated by Elma Blom and Jan de Jong and is called: *MAIN: Nederlandse versie (Meertalig Assessment Instrument voor Narratieven)*.



Figure 5.1: Picture Sequence for the Baby Birds Story (Taken from Gagarina et al., 2012).



Figure 5.2: Picture Sequence for the Baby Goats Story (Taken from Gagarina et al., 2012).

Each story in the MAIN protocol contained three episodes, each with a Goal (G) Attempt (A) and Outcome (O). After each story, ten comprehension questions were asked to test the pupil's understanding of the goals and the internal states of the characters in the pictures. For further description of the MAIN, see Gagarina et al. (2012; 2015).

### 5.3.3 Procedure

The data collection for this study was carried out over a period of about two-and-a-half years. Each participant was assessed four times using the MAIN. The first two sessions took place within the first year after the child had arrived in the Netherlands. The mean time interval between Session 0 and 1 was approximately 3 months, between Session 1 and 2 approximately 9 months, and between Session 2 and 3 it was 12 months. For participants at DL2-schools, Session 2 took place about 6 months after their transfer to a mainstream school. Thus, in the final two sessions no participant attended a DL2-school. For participants already at a mainstream school the date of Session 2 had been set on about 1.5 years after the beginning of their school career in the Netherlands.

All pupils were assessed individually in a separate room at their school (see Figure 5.3a and 5.3b for the set-up of the assessments). Each session was conducted in Dutch, which was the language of instruction at the participants' schools. One researcher conducted all the assessments adhering to the same test protocols and procedures in order to ensure the high reliability.

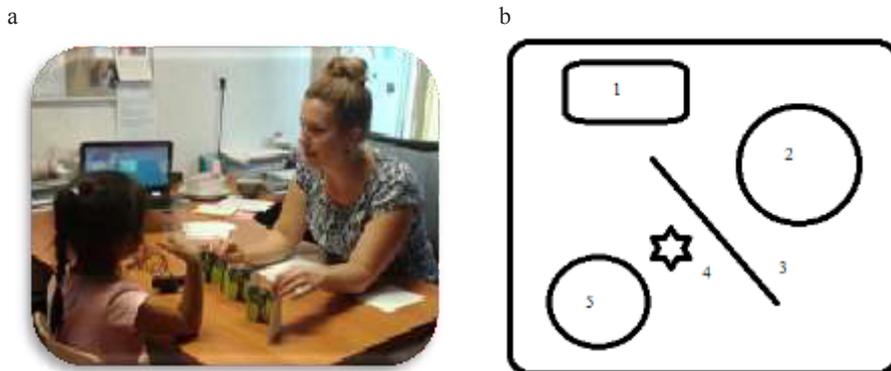


Figure 5.3a and 5.3b: Picture and Graphical Representation of the Assessment Set-up. 1: video-camera; 2: researcher; 3: picture sequence; 4: microphone; 5: participant.

A productive test like the MAIN could cause the participants to become frustrated if they only have little knowledge of the language used in the assessment. Likewise, the collected data could be difficult to interpret if the speaker's command of Dutch is too weak. Therefore, the decision was made to assess the newly arrived migrant pupils for the first

time with the MAIN two to three months after the beginning of the study, when they had had at least two months of Exposure to Dutch at School.

During the assessments we followed the MAIN protocol. The sessions started with the picture sequence of one of the MAIN retelling stories (Cat story or Dog story). This story was modelled by the researcher and the participants had to retell the story. Following this procedure, the narrative structure was modelled for the participants, as well as the specific vocabulary for that story. This retelling story was followed by ten comprehension questions. After that, the participants were presented with a second picture sequence (version Baby Birds or Baby Goats, see Figures 5.1 and 5.2). This story was not modelled by the researcher (although the retell story could be used as a model for this story as well); the children were encouraged to tell the story themselves. After the second story the participants had to answer another ten comprehension questions. The version of the story in Session 0 was randomly selected and then the versions were used alternately (see Table 5.2).

Table 5.2: Overview of a Possible Order of the Stories used for one Participant.

	Story retelling task (warm-up session)	Story generation task
Session 0	Cat story	Baby Goats story
Session 1	Dog story	Baby Birds story
Session 2	Cat story	Baby Goats story
Session 3	Dog story	Baby Birds story

Even though two stories were used in each session during the present study (one model story for story retelling and one for story generation) only the stories for story generation were analyzed. We choose only to analyze the story generation task because we were specifically interested in the more spontaneous productive proficiency of the participants instead of their capacity of being able to retell a story. The vocabulary during the retell part was modelled whereas the researcher did not prompt any of the words in the story generation part. Likewise, the answers to the comprehension questions after all stories, which are part of the MAIN protocol, were not analyzed.

### **5.3.4 Analyses**

#### **Child characteristics**

Three factors concerning child characteristics were taken into account in the analysis: Age, Exposure to Dutch at School, and Educational Facility.

Table 5.3 summarizes the descriptive characteristics of the participants. Table 5.3 is comparable with Table 4.3 from Chapter 4 except for Exposure to Dutch at School, since

the first assessments of the narratives were later than that of the receptive vocabulary assessment, and hence the amount of Exposure to Dutch at School was different.

Table 5.3: Descriptive Characteristics of the Participants During the Narrative Assessments at Session 0 (N = participating pupils).

	Gender		Exposure to Dutch at School (in months)	
	Boys (%)	Girls (%)	Mean	<i>sd.</i>
DL2-school (N = 32)	53.1	46.9	4.50	1.29
Mainstream school (N = 10)	40.0	60.0	5.50	2.32
Total (N = 42)	50.0	50.0	4.74	1.62

### Microstructure and macrostructure

Each narrative was digitally recorded (audio and video) for later transcription and analysis. The samples were transcribed using a sample of the conventions of CLAN/CHAT (MacWhinney, 2000). The analysis of the story will be in twofold: both the microstructure and the macrostructure of the stories were analyzed.

#### Microstructure

For the analysis of the microstructure, different lexical measures can be investigated (e.g. Iluz-Cohen & Walters, 2012; Muñoz, Gillam, Peña, & Gulley-Faehnle, 2003; Pearson, 2002; Reuterskiöld, Hansson, & Sahlén, 2011; Reuterskiöld-Wagner, Sahlén, & Nettelbladt, 1999; Uccelli & Páez, 2007). Gagarina et al. (2015) suggested ten features that could serve as the basis for analyzing the microstructure of narratives. These suggested features included length and lexis, syntactic complexity and discourse cohesion, and code-switching. For our purpose, however, we only used one of these suggested (namely, Number of Different Words) features and added two (namely, Guiraud Index Score and Measure of lexical Richness), since we focused more on vocabulary than on sentences. Microstructure was analyzed in three parts, namely two measures for lexical diversity (Number of Different Words and the Guiraud Index Score) and one for lexical richness (Measure of lexical Richness).

#### Lexical diversity

Lexical diversity was measured by means of Number of Different Words (NDW) and the Guiraud Index Score (GIS). NDW was used because this was reported in many studies using the MAIN (e.g., Lindgren, 2018; Altman et al., 2016; Kapalková et al., 2016; Tsimpli, Peristeri, & Andreou, 2016).

However, the NDW is disputed because it is highly influenced by the length of a story. A measure of lexical diversity which seems to reduce the impact of text length is the Guiraud Index Score (GIS; Daller, Van Hout, & Treffers-Daller, 2003). Therefore, in addition to NDW we calculated the GIS by dividing the number of types (different words) by the square root of the total of tokens (total number of words) of the story. The GIS was chosen since it was likely that the population of the present study would, at least in the first narratives, produce stories with a limited text length, the GIS was expected to be a more precise measure than NDW.

### *Lexical richness*

Another way of looking at lexical items is to assess their richness. Two pupils could have a similar score of NDW or GIS but in practice use different kind of words. For example, one pupil could use the frequently used word “tree” while another pupil would use the less frequent word “oak”. To make a distinction between participants using general, more frequent words and participants using specific low frequent words a measure of lexical richness was added to the analysis. The Measure of Lexical Richness (MLR) takes into account the frequency band of the words. An online program<sup>14</sup> which compares the words from the story with a frequency list of Dutch words was used to calculate the lexical richness of the words in the participants’ stories. This program divides the words in nine frequency bands (Vermeer, 2016). Words not recognized by the program were checked and mostly changed from colloquial speech to their written counterpart. For example, “’t” (the reduced form for it) was changed into “het”. Also, incorrectly inflected verbs were rewritten in the correct form, for example “gevliegen” (instead of “gevloegen,” meaning has flown). The words from the higher frequency bands were checked specifically for whether the participant actually meant what the program had assigned as a meaning to the word.

### *Macrostructure*

Macrostructure was investigated in three parts: through Story Structure, Structural Complexity, and Internal State Terms (ISTs). All three parts of macrostructure are present in the protocol of the MAIN (Gagarina et al., 2012), however, our analysis slightly deviates from that protocol. Table 5.4 illustrates the macrostructure of an example story. Story Structure, Structural Complexity, and ISTs are in column one, four, and five of Table 5.4, respectively.

### *Story Structure*

For the analysis of Story Structure, the story grammar model (e.g. Mandler, 1979; Stein & Glenn, 1979) was used, but in a slightly adapted version (following Gagarina et al., 2012).

---

<sup>14</sup> <https://lukasvermeer.github.io/mlr/>

The total score for Story Structure was calculated as follows: the mentioning of the setting (a reference to the time and place or a context introduction) could score up to 2 points. Further, each full episode could receive up to 5 points, with each part of an episode scoring 1 point. A full episode contained: An Internal State Term (IST) as an initiating event, a Goal (G), an Attempt (A), an Outcome (O), and an IST as reaction at the end of the episode. Each story included three episodes, therefore the participants were awarded in total up to 17 points for all the story structure elements they would mention.

Table 5.4: Macrostructural Framework of the Multilingual Assessment Instrument for Narratives (Taken from Gagarina et al. 2012, p. 20, with two Additional Columns on the Right).

Story Structure elements	Description	Example	Part of Structural Complexity	Part of Internal State Terms
Setting	Reference to time and place (considered to be outside the episode itself; 2pts).	<i>One day in the forest, there was a mother bird with three little babies.</i>		
IST as initiating event	An event or an internal state that sets the events of the story in motion (1pt.).	<i>The baby birds were crying, and the mother bird saw that the babies were hungry.</i>		x
Goal (G)	A statement of an idea of the protagonist to deal with the initiating event (an indication of goal-directed planning) (1pt.).	<i>“Oh, my babies are so hungry”, said the mother bird and she decided to get some worms.</i>	x	
Attempt (A)	An indication of action to obtain the goal (1pt.).	<i>The mother bird flew away to look for food.</i>	x	
Outcome (O)	The event(s) following the attempt and causally linked to it (either one or several outcomes, either successful or not; 1pt.).	<i>The mother bird came back with a big worm and the baby birds got some food.</i>	x	
IST as reaction	A statement defining how the protagonist(s) feel or think about the outcome. It can also include an action resulting from an emotional response (1pt.).	<i>And the baby birds were so happy.</i>		x

### *Structural Complexity*

Structural Complexity pertains to whether the participants mentioned the Goal (G), the Attempt (A), and the Outcome (O) of an episode (Westby, 2012; Gagarina et al., 2012). Structural Complexity is thus part of Story Structure, but the ISTs are left out of this

measure of complexity, leaving only the GOA sequence. The Structural Complexity score ranges from high to low, based on the relationships between the macrostructural elements. Different combinations of these three elements are seen as more or less complex (Bonifacci, Barbieri, Tomassini, & Roch, 2017). Different authors have used different methods for calculating Structural Complexity. Some authors counted all elements, others assigned points to different combinations of elements (a.o., Altman et al., 2016; Gagarina, 2016). Another possibility is to include the setting (Peristeri, Andreou, & Tsimpli, 2017). Other researchers assigned a level to the complexity, varying from absence of complexity to a high level of complexity. Some used these levels and did their analysis with only the highest level assigned to one of the episodes of the story (Bonifacci et al., 2017). In the present study we used four levels of complexity which will be explained below. Furthermore, we assigned points according to the level of complexity to each episode of the story to calculate a total score as a measure of complexity.

An episode could have one of the following four levels of complexity: “Absence of complexity” implies none of the elements or only an Attempt or an Outcome. “Low complexity” implies the presence of only the Goal, or both Attempt and Outcome but no Goal. “Medium complexity” implies the presence of two elements, of which at least one is the Goal. “High complexity” implies the presence of all three elements, thus a complete sequence of Goal, Attempt, and Outcome. The levels “Absence,” “Low,” “Medium,” and “High” complexity were respectively scored as 0, 1, 2, and 3 (see Table 5.5). Finally, the scores of all three episodes were added up, resulting in a maximum score of 9 for Structural Complexity.

Table 5.5: The four Levels of Structural Complexity

Level of complexity	The presence of Goal, Attempt, or Outcome			Points assigned
	Goal	Attempt	Outcome	
Absence of complexity	-	-	-	0
	-	+	-	0
	-	-	+	0
Low complexity	+	-	-	1
		+	+	1
Medium complexity	+	+	-	2
	+	-	+	2
High complexity	+	+	+	3

Our analysis of complexity is similar to Bonifacci et al. (2017), with one adjustment, namely that also the mentioning of only the Goal was assigned as having a LOW COMPLEXITY, instead of counting the mentioning of only the Goal as absence of

complexity. Contrary to Bonifacci et al., in the present study all scores were added up, instead of using the highest score for complexity of one of the episodes.

*Internal State Terms*

The third and last level of the macrostructure of narratives is the use of Internal State Terms (ISTs). With the use of ISTs narrators provide important information about their own awareness and understanding of the story characters’ mental states: what they know, what they value, and what they believe (Nippold, Ward-Loneragan, & Fanning, 2005).

The development of the internal state language of the pupils can be measured by means of the frequency of occurrence of ISTs. The words included as ISTs were taken from the Dutch version of the MAIN with the explicit addition of “to laugh,” “to be shocked,” and “to cry” (see Table 5.6). These three terms were used often in the present data set and were explicitly added to the protocol to make sure that the two transcribers would count them as ISTs. All frequency lists generated with CLAN were checked for these terms and counted to get a total number of IST tokens in the story.

Table 5.6: Internal State Term Categories and Examples (Taken from Gagarina et al., 2012).

<b>Internal State Term category</b>	<b>Examples</b>
Perceptual state terms	see, hear, feel, smell
Physiological state terms	thirsty, hungry, tired, sore
Consciousness terms	alive, awake, asleep
Emotion terms	sad, happy, angry, worried, disappointed, laugh*, cry*, to be shocked*
Mental verbs	want, think, know, forget, decide, believe, wonder, have / make a plan
Linguistic verbs/ verbs of saying/ telling	say, call, shout, warn, ask

\* These three terms are added to the examples of Gagarina et al. (2012).

ISTs can be scored based on all tokens (lexical items) of ISTs the narrators used in their story. On the other hand, ISTs can also be scored by calculating only the IST elements of the episodes in the Story Structure thus: one internal response at the beginning of the episode and one at the end of the episode. Following the procedure of Gagarina (2016), the percentage of ISTs out of all the word tokens in the narrative was then calculated. Including all IST tokens in a narrative exceeds the two IST parts in the Story Structure. To illustrate this: the word “said” was counted in the present analysis as an IST even though it was uttered in part of the Goal in the following: “mother said that she had to go away in order to get some food.” Likewise, “happy” was counted in the present analysis as IST, even though it was uttered in part of the Outcome: “the happy mother returned with the food.” In the Story Structure analysis there is a limit of 6 points for ITs. However, when all ISTs

uttered in a story are included, there is no maximum because pupils can use as much ISTs as they wish, in any part of the story.

Analyzing the stories at both the microstructural and macrostructural levels enabled us to discuss the narrative ability development of newly arrived migrant kindergarteners in both the linguistic competence as well as the discourse competence areas of communicative competence. Table 5.7 summarizes all the six measurements in the present study.

Table 5.7: Summary of the Six Measurements with the MAIN.

Microstructure	Lexical Diversity	1. Number of Different Words (NDW)
		2. Guiraud Index Score (GUI)
	Lexical Richness	3. Measure of Lexical Richness (MLR)
Macrostructure		4. Story Structure (SS)
		5. Story Complexity (SC)
		6. Internal State Terms (IST)

### Statistical analysis

To answer our research questions, multilevel modeling is applied to take into account the hierarchical structure of the data due to repeated measures per participant. Measurements of a pupil's outcome variables (level 1,  $N = 168$ ) were temporally ordered within pupils (level 2,  $N = 42$ ). For this purpose, multilevel modeling of repeated measures data procedures (MLwiN software version 2.36; Rasbash, Charlton, Browne, Healy, & Cameron, 2016) was carried out. See for further details on the statistical analysis Section 4.3.4.

### Intra-rater reliability

#### *Microstructure*

The transcription of the stories was done by one transcriber due to practical limitations. Nevertheless, we calculated the intra-transcriber reliability in order to see if the transcriptions could be considered reliable. We randomly selected 16 stories, 4 from each session. These 16 stories were transcribed twice by the same researcher, with a minimum of five months a part. The word-by-word agreement for the transcriptions was 89.9% (range 68.8–100%) with a Cohen's kappa of .80, which is considered a substantial strength of agreement (Landis & Koch, 1977). The mean reliability in Session 3 was the highest (95.0%) and in Session 1 the lowest (82.9%). Mean Cohen's kappa scores of the sessions ranged from .68 to .90, substantial to almost perfect agreement. This high intra-transcriber

reliability assured us that we could use the transcriptions for the calculation of NDW, GIS, and MLR.

### *Macrostructure*

For this study the scoring protocol of the MAIN (Gagarina et al., 2012) was used, but in an extended version. Because there were young participants in this study who just started to learn Dutch, it was difficult to follow the strict MAIN protocol. These pupils were creative, and it was difficult to predict what they would tell in the stories; therefore, the protocol was extended with more examples and guide lines, to make the scoring of the protocol more consistent over the raters.

For macrostructure, inter-rater reliability was only calculated for Story Structure, because Structural Complexity was based on the story grammar components of Story Structure and the IST tokens were already included in the reliability of the transcriptions.

We again randomly selected 16 stories (four stories per session) for this inter-rater reliability analysis. The author of this dissertation and a trained research assistant both scored the 16 stories independently and the point-to-point agreement for the ratings was 88.7% (range 75.1–100%) with a Cohen's kappa of .70, which was considered a substantial strength of agreement (Landis & Koch, 1977). The mean reliability in Session 2 was the highest (96.9%) and in Session 3 the lowest (84.4%). Mean Cohen's kappa scores of the sessions ranged from .49 to .93, meaning moderate to almost perfect agreement. In Session 1, one child did not score any points according to one rater which resulted in a Cohen's kappa of 0 (the use of only one category does not deviate much from chance), even though the agreement was 81%. This low kappa reduced the overall kappa. When this kappa of 0 was left out of the analysis the overall kappa increased from .70 to .74. This substantial inter-rater reliability assured us that we could continue our analysis with the scorings of the narratives by one rater.

## **5.4 Results**

The sample comprised of 164 narratives collected during four sessions over a period of two-and-a-half years. Due to technical difficulties, four stories from four different participants are missing in the data set. The results of the narrative ability development are presented as follows: first, the developmental data from the MAIN is presented to answer the question: How does the *narrative ability* (in Dutch) of newly arrived migrant kindergarteners develop during the first two-and-a-half years of schooling in the Netherlands in relation to school type? The basis of the analysis is the General Development Model, which includes only Age at which narrative ability was measured as an explanatory variable. Second, child characteristics are introduced into the model. Exposure to Dutch at School will reveal whether exposure has a different influence on the

development than Age. Furthermore the variable Educational Facility is added to the model to see whether there are differences between pupils in DL2-schools and Mainstream school.

The development of the narrative ability of young newly arrived migrant pupils is described on two levels: the microstructural level and the macrostructural level. First the three measures of microstructure (Number of Different Words, the Guiraud Index Score, and the Measure of Lexical Richness) will be presented. Second the three measures of macrostructure (Story Structure, Structural Complexity, and the Internal State Terms) will be presented.

### 5.4.1 Microstructure

Three types of microstructure measures were analyzed: Number of Different Words (NDW), the Guiraud Index Score (GIS), and the Measure of Lexical Richness (MLR). In this section the General Development Models of these three measures were build.

#### *Number of Different Words*

Example (1) and (2) illustrate our variable Number of Different Words. Both stories are from the same girl and they show how the Baby Bird story evolved from Session 0 to Session 3. In over a month the NDW score of this girl's stories increased from 8 to 25. In all coming examples we will first give the Dutch text and then an English translation. It is not a word-by-word translation in order to make the English translation more understandable. The wrong verb inflections are from the original text in order for the English reader to understand the incorrect use of grammar.

Example (1): Subject number 0252; age 5;1; around 2 months in school; home language: Chinese; Session 0; NDW score 8.

mama. baby. deze van de baby. deze van mama. poes. poes deze poes eten drinken. mama.

Mama. Baby. This one of the baby. This one of mama. Puss. Puss this one puss eats drinks. Mama.

Example (2): Subject number 0252; age 6;6; around 19 months in school; Session 3; NDW score 25.

de moedervogel ga de eten zoeken. en de mamavogel is weg. maar toen kwam de poes aan. en de poes klimt. maar de moeder is er. en de poes pakt het een vogel. en de hond ziet. en pakt het poes. en dan de hond ga poes doen wegwezen.

The mother bird am going to look for the food. And the mother bird is gone. But then the puss came. And the puss climbs. But the mother is there. And the puss grabs it one bird. And the dog sees. And grabs it puss. And then the dog am going to get away cat.

The NDW is the first measure of lexical diversity in this study. From the comparison between the consecutive models for the NDW (Table 5.8) it is apparent that a model with a fixed linear component – allowing for differences in Age – fit to the data better than a model with only an intercept ( $\Delta\chi^2$  (NDW<sub>1</sub>) = 79.44;  $df = 1$ ;  $p < .001$ ). The variance within (NDW<sub>2</sub>) or between (NDW<sub>3</sub>) individuals did not depend on the age of the participants. The inclusion of Age<sup>2</sup> (NDW<sub>4</sub>) did not improve the fit significantly. Hence, in the final model (NDW<sub>1</sub>) a fixed effect of Age is needed, and with this model we continued the analysis.

Table 5.8: Fit of Different Models (-2LL) for Changes in Number of Different Words (168 cases) as well as the Comparison of Consecutive Models.

Model	-2LL	Comparison			
		Models	$\Delta\chi^2$	$\Delta df$	$p$
NDW <sub>0</sub> : $\beta_{0ij}cons$ <sup>a</sup>	1365.18				
NDW <sub>1</sub> : $NDW_0 + \beta_1 Age^1_{ij}$	1285.75	NDW <sub>0</sub> vs NDW <sub>1</sub>	79.44	1	<.001
NDW <sub>2</sub> : $NDW_1 + e_{1ij}Age^1_{ij}$	1285.51	NDW <sub>1</sub> vs NDW <sub>2</sub>	0.23	1	.63ns <sup>b</sup>
NDW <sub>3</sub> : $NDW_2 + u_{10j}Age^1_{ij}$	1283.61	NDW <sub>2</sub> vs NDW <sub>3</sub>	1.90	2	.39ns
NDW <sub>4</sub> : $NDW_3 + \beta_2 Age^2_{ij}$	1282.47	NDW <sub>3</sub> vs NDW <sub>4</sub>	1.15	1	.28ns

<sup>a</sup>NDW<sub>0</sub>: In addition to the intercept, variance components for differences within and between individuals are estimated

<sup>b</sup> Only the covariance-coefficient was estimated.

Based on this General Development Model we constructed Figure 5.4, in which both the average development as well as the differences within and between individuals are represented (see Table 2.1 in Appendix 2 for the parameter estimates). The average NDW score at an age of 73 months was estimated as 36.27. Each month a pupil grew older the NDW increased by 0.73.

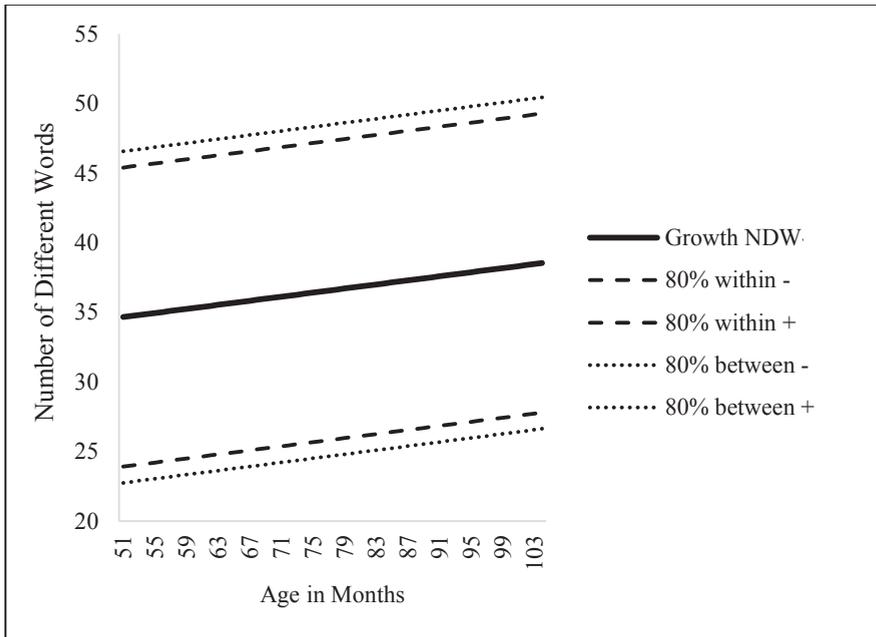


Figure 5.4: Graphical Representation of the General Development Model of NDW with 80% Reliability.

Figure 5.4 shows that of the effect of Age on the NDW is significant. The differences within individuals (represented by the striped lines in Figure 5.4) and between individuals (represented by the dotted lines in Figure 5.4) did not depend on Age. The variance within and between pupils is large and constant which means that over time there are equal large differences in scores within participants and between participants.

Exposure to Dutch at School, measured as the number of months a pupil has attended a Dutch school (excluding summer months), was added to the General Development Model of the NDW. A likelihood ratio test showed that the main effect of Exposure to Dutch at School contributed significantly to the fit of the model to the observed data ( $\Delta\chi^2(\text{NDW}_5) = 5.73; df = 1; p = .02$ ; see Table 5.9): the longer the Exposure of Dutch at School was, the higher the score of the NDW. No interaction between Age and Exposure to Dutch at School was found ( $\text{NDW}_6$ ).

Table 5.9: Fit of Different Models (-2LL) for Changes in Number of Different Words (168 cases) with Addition of Exposure to Dutch at School (EDS) as Explanatory Variable.

Model	-2LL	Comparison			
		Models	$\Delta\chi^2$	$\Delta df$	$p$
NDW <sub>1</sub> : $\beta_{0ij}cons + \beta_1Age^l_{ij}$	1,285.75				
NDW <sub>5</sub> : NDW <sub>1</sub> + $\beta_2EDS_{ij}$	1,280.01	NDW <sub>1</sub> vs NDW <sub>5</sub>	1	5.73	.02
NDW <sub>6</sub> : NDW <sub>5</sub> + $\beta_3Age*EDS_{ij}$	1,276.80	NDW <sub>5</sub> vs NDW <sub>6</sub>	1	3.22	.07ns

The difference in NDW score when Exposure of Dutch at School was taken into account is graphically presented in Figure 5.5. There was a fixed main effect of Exposure to Dutch at School, but there was no interaction effect between Age and Exposure to Dutch at School (see Table 2.1 in Appendix 2 with the estimated parameters). Pupils with a longer Exposure to Dutch at School score higher on the NDW than pupils with shorter Exposure to Dutch at School. The effect of Exposure to Dutch on NDW was similar for younger pupils and older pupils.

Importantly, the main effect of Age became insignificant when Exposure to Dutch at School was added: the main effect of Age was overshadowed by Exposure to Dutch at School. This is plausible since Exposure to Dutch at School goes hand in hand with growing older. So, in this respect Age and Exposure to Dutch at School are (partly) the same variable (1 month increase in Exposure to Dutch at School equals 1 month increase in Age). However, it is visible in Figure 5.5 that Exposure to Dutch at Schools exceeds Age, represented by the line “Exposure = 0” which indicates general growth. At any age, the additional effect of Exposure of Dutch at School was similar, but significant.

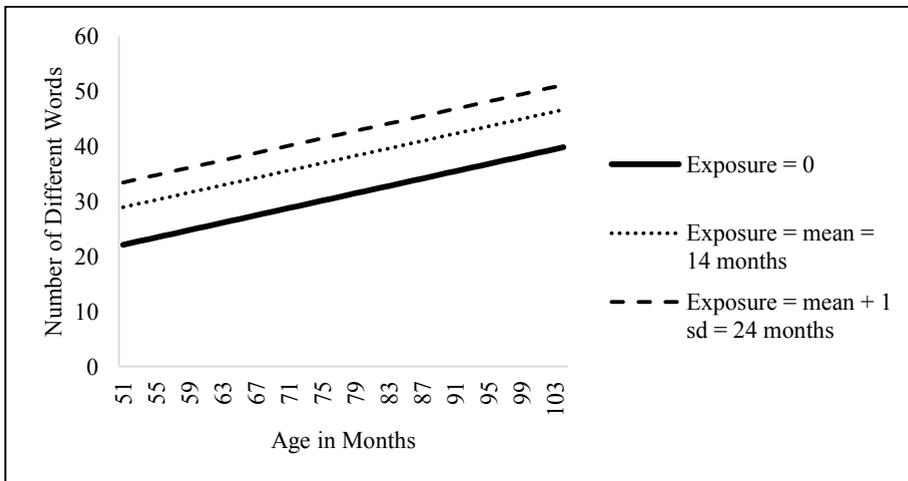


Figure 5.5: Development of NDW with Exposure to Dutch at School as Explanatory Variable.

The addition of Educational Facility as a fixed main effect did not improve the General Development Model of NDW score (NDW<sub>7</sub>; see Table 5.10), however the interaction between Age and Educational Facility did ( $\Delta\chi^2$  (NDW<sub>8</sub>) = 5.88;  $df = 1$ ;  $p = .02$ ).

Table 5.10: Fit of Different Models (-2LL) for Changes in Number of Different Words (168 cases) with Addition of Educational Facility (EduFac) as Explanatory Variable.

Model	-2LL	Comparison			
		Models	$\Delta\chi^2$	$\Delta df$	$p$
NDW <sub>1</sub> : $\beta_{0ij}cons + \beta_1Age^l_{ij}$	1,285.75				
NDW <sub>7</sub> : NDW <sub>1</sub> + $\beta_2EduFac_{ij}$	1,284.36	NDW <sub>1</sub> vs NDW <sub>7</sub>	1.39	1	.24ns
NDW <sub>8</sub> : NDW <sub>9</sub> + $\beta_3Age * EduFac_{ij}$	1,278.48	NDW <sub>7</sub> vs NDW <sub>8</sub>	5.88	1	.02

There was no fixed main effect of Educational Facility, but there was an interaction effect between Age and Educational Facility (see Table 2.1 in Appendix 2 for the estimated parameters). The influence of Educational Facility is larger for younger pupils than for older pupils. We do however refrain from plotting the model in a graph because this would be difficult to interpret. A larger sample is necessary to confirm the effect whether a certain school type is better for younger pupils with regard to NDW than the other.

### **Guiraud Index Score**

Example (3) and (4) illustrate our variable Guiraud Index Score. Both stories are from the same boy w and they show how the Baby Bird story evolved from Session 0 to Session 3. In a year the GIS of this boy's stories increased from 1.18 to 4.09.

Example (3): Subject number 0701; age 5;10; around 5 months in school; home language: Polish; Session 0; GIS 1.18.

die vlieg. en die vlieg. poes vlieg en poes. een vlieg en poes. vlieg en poes en hond. vliegt. vlieg en poes en vlieg en poes en hond kijk. vliegt poes en hond, vliegt poes en hond. die die die. ja vlieg, ja hond. en poes.

That one fly [is a bird?]. And that one fly. Puss, fly and puss. A fly, and, puss. Fly and puss and dog. Flies. Fly and puss and fly and puss and dog look. Flies puss and dog, flies puss and dog. That one that one that one. Yes, fly, yes dog. And puss.

Example (4): Subject number 0701; age 6;10; around 18 months in school; Session 3; GIS 4.09.

mam heb teruggevliegt en nog een keer gaan weg. en mam vliegt zie niet poes, maar is ie wel. poes springt op de boom. mam vliegt heeft baby is vliegt eten. en dan poes springt op de tweede baby. en dan hond komt, poes zie niet. en hond pakt van poes staart. en haalt van de boom. poes gaan wegrennen en hond achter poes.

Mam has flown [wrong Dutch inflection] back and go away again. And mam flies and see not puss, but it is. Puss jumps upon the tree. Mam flies has baby is flies eat [food?]. And then, puss jumps on the second baby. And then dog comes, puss do not see. And dog grabs of the cat's tail. And takes from the tree. Puss go run away and dog behind puss.

The GIS is a second measure of lexical diversity in this study. From the comparison between the consecutive models for the GIS (Table 5.11) it is apparent that a model with a fixed linear component – allowing for differences in Age – fit the data better than a model with only an intercept ( $\Delta\chi^2$  (GIS<sub>1</sub>) = 63.11;  $df = 1$ ;  $p < .001$ ). The variance within individuals did not depend on the age of the participants (GIS<sub>2</sub>). However, the variance between individuals was a (linear) function of age ( $\Delta\chi^2$  (GIS<sub>3</sub>) = 4.90;  $df = 1$ ;  $p = .03$ ). The inclusion of Age<sup>2</sup> did not improve the fit of the model (GIS<sub>4</sub>). Hence, in the final model (GIS<sub>DEF</sub>) a fixed main effect of Age, as well as a variance between pupils component which depends on Age needs to be included, and with this model we continued the analysis.

Table 5.11: Fit of Different Models (-2LL) for Changes in Guiraud Index Score (168 cases).

Model	-2LL	Comparison			
		Models	$\Delta\chi^2$	$\Delta df$	$p$
GIS <sub>0</sub> : $\beta_{0ij}cons$ <sup>a</sup>	432.48				
GIS <sub>1</sub> : GIS <sub>0</sub> + $\beta_1 Age^1_{ij}$	369.37	GIS <sub>0</sub> vs GIS <sub>1</sub>	63.11	1	<.001
GIS <sub>2</sub> : GIS <sub>1</sub> + $e_{1ij} Age^1_{ij}$	365.75	GIS <sub>1</sub> vs GIS <sub>2</sub>	3.62	2	.16ns
GIS <sub>3</sub> : GIS <sub>2</sub> + $u_{10j} Age^1_{ij}$	360.86	GIS <sub>2</sub> vs GIS <sub>3</sub>	4.90	1	.03 <sup>b</sup>
GIS <sub>4</sub> : GIS <sub>3</sub> + $\beta_2 Age^2_{ij}$	359.68	GIS <sub>3</sub> vs GIS <sub>4</sub>	1.18	1	.28ns
GIS <sub>DEF</sub> : $\beta_{0ij}cons + \beta_{1j} Age^1_{ij}$	362.10				

<sup>a</sup> GIS<sub>0</sub>: In addition to the intercept, variance components for differences within and between individuals are estimated.

<sup>b</sup> Only the covariance-coefficient between the intercept- and the age-residuals was estimated.

Based on this General Development Model we constructed Figure 5.6, in which both the average development as well as the differences within and between individuals are represented (see Table 2.2 in Appendix 2 for the parameter estimates). The average GIS at an age of 73 months was estimated as 3.63. Each month a child grew older, his GIS increased by 0.04.

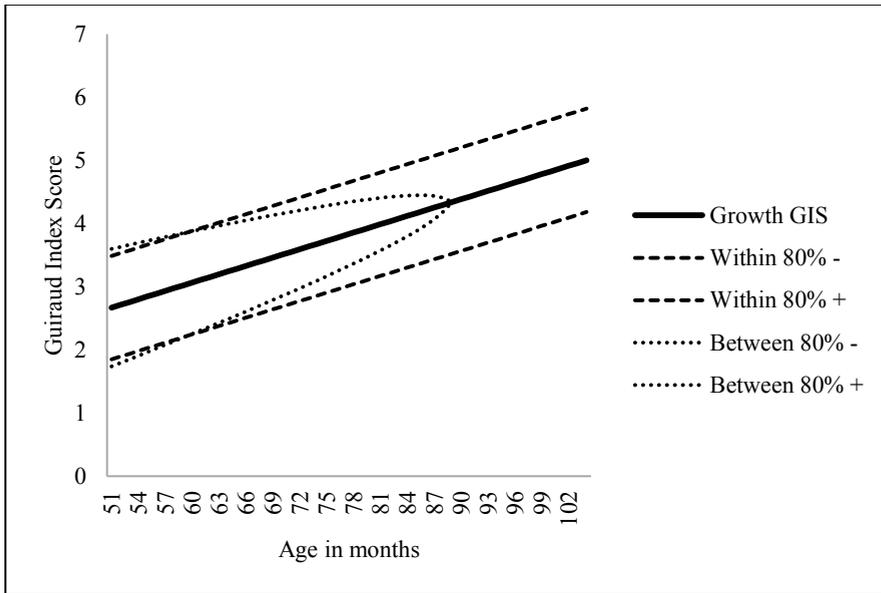


Figure 5.6: Graphical Representation of the General Development Model of GIS with 80% Reliability.

Figure 5.6 shows that in the effect of Age on GIS is significant. Figure 5.6 also shows that differences between individuals (represented by the dotted lines in Figure 5.6) are a function of Age; differences between younger children proved to be relevant (and significant), however the differences between children appeared to diminish with age.

First, Exposure to Dutch at School was added to the model. A likelihood ratio test showed that the main effect of Exposure to Dutch at School contributed significantly to the fit of the model to the observed data ( $\Delta\chi^2$  (GIS<sub>5</sub>) = 12.38;  $df = 1$ ;  $p < .001$ ; see Table 5.12): the longer the Exposure to Dutch at School the higher the score of GIS. No interaction between Age and Exposure to Dutch at School was found (GIS<sub>6</sub>). The difference in GIS when Exposure to Dutch at School was taken into account is graphically presented in Figure 5.7 (see Table 2.2 in Appendix 2 for the estimated parameters).

Table 5.12: Fit of Different Models (-2LL) for Changes in Guiraud Index Score (168 cases) with Addition of Exposure to Dutch at School (EDS) as Explanatory Variable.

Model	-2LL	Models	Comparison		
			$\Delta\chi^2$	$\Delta df$	$p$
GIS <sub>DEF</sub> : $\beta_{0ij}cons + \beta_{1j}Age^{l}_{ij}$	362.10				
GIS <sub>5</sub> : GIS <sub>DEF</sub> + $\beta_2EDS_{ij}$	349.72	GIS <sub>DEF</sub> vs GIS <sub>5</sub>	12.38	1	<.001
GIS <sub>6</sub> : GIS <sub>5</sub> + $\beta_3Age*EDS_{ij}$	346.41	GIS <sub>5</sub> vs GIS <sub>6</sub>	3.31	1	.07ns

Importantly, the effect of Age became insignificant when Exposure to Dutch at School was added: the main effect of Age was overshadowed by Exposure to Dutch at School. This is plausible since Exposure to Dutch at School goes hand in hand with growing older. So, in this respect Age and Exposure to Dutch at School are (partly) the same variable (1 month increase in Exposure to Dutch at School equals 1 month increase in Age). However, it is visible in Figure 5.7 that Exposure to Dutch at Schools exceeds Age, represented by the line “Exposure = 0” which indicates general growth. At any age the additional effect of Exposure of Dutch at School was similar but significant.

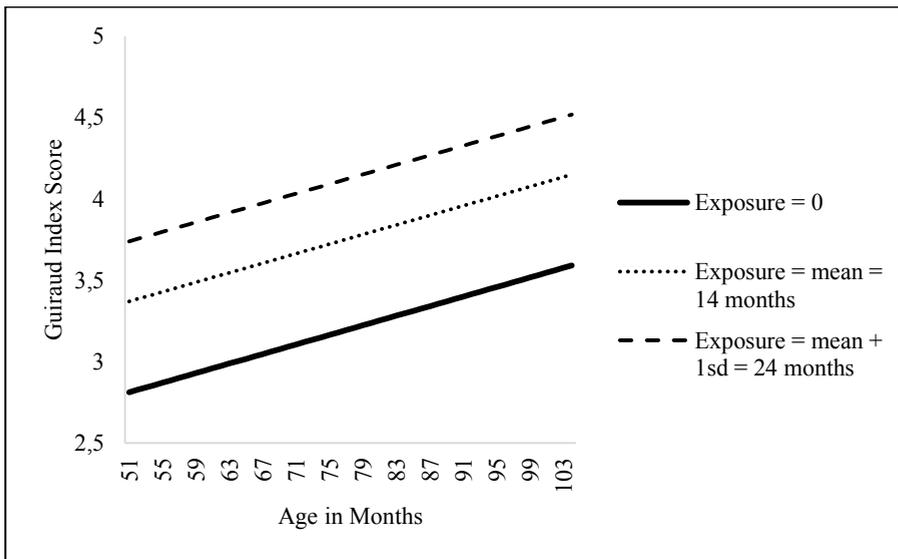


Figure 5.7: Development of GIS with Exposure to Dutch at School as Explanatory Variable.

The addition of the variable Educational Facility did not improve the general model of growth in Guiraud Index Score. This means that we could not establish a relationship between being at a DL2-school or a Mainstream school and the development of pupils' GIS.

### **Measure of Lexical Richness**

Example (5) and (6) illustrate our variable Measure of Lexical Richness. The stories in (5) and (6) both have a NDW score of 43, but the story in Example (5) has a MLR score of 9.42 (low frequent words are indicated in bold). For comparison, the story in Example (6) has a MLR score of 1, only high frequent words were used for this story.

Example (5): subject number 1032; age 6;5; around 14 months in school; home languages: Filipino, English, and Chinese; Session 2; NDW score 43; MLR score 9.42.

de mama-vogel kijkt naar de twee **babyvogel**. de twee babyvogel wil eten. de vogel gaat **wegvliegen** te eten pakken. en de poes vindt de twee **babyvogel** lekker. hij vindt het eten maar hij kan niet klimmen. twee **babyvogel** gaat een **slak** eten. toen de poes gaat klimmen de twee **babyvogel** te pakken en te eten. maar hij was toen daar dan heb de een **babyvogel** gepakt. en mama heb niet toen gekeken. toen de hond schrikt toen de poes was zo boos en hij wil de vogel eten. toen de hond klimt dan de hond heb de staart van de poes **geknepen**. en de vogel schrikt want dan de hond heb de poes bang gemaakt. en de vogel was zo schrikt en zij vindt de twee babietjes zo fijn.

The mother bird looks at the two **baby-bird**. The two **baby-bird** want food. The bird is going to **fly away** to get food. And the puss likes the two **baby-bird**. He likes it to eat, but he cannot climb. Two **baby-bird** goes eat a **snail**. Then the puss goes climbing the two **baby-bird** to catch and to eat. But he was then there, then has the a **baby-bird** grabbed. And mama has not then looked. Then the dog is scared, then the puss was very angry and he wants the bird to eat. Then the dog climbs, then the dog **has pinched** the puss' tail. And the bird is scared because then the dog has frightened the puss. And the bird was so scared and she likes the little babies really much.

Example (6): subject number 0243; age 7;5; around 19 months in school; home language: Chinese; Session 2; NDW score 43; MLR score 1.

een grote vogel hebben twee vogeltjes en de ze kind. en de poes wil deze kleine vogel pakken. en ze moeder vliegt weg. zo [?] kan maar ze poesje kan naar boom klimmen. en vogel kijkt naar poes ze is heel bang. en dan poesje wil de vogeltjes pakken. en ze gaat de boom klimmen wil de vogel pakken. ze moeder is aan de andere vogel kijken zo kan ze niet zien deze poes de vogeltjes pakken. de hond heb gezien de poesje ga vogel pakken en ze moeder ga andere ze de vogeltjes kan niet zien en de poesje gaat andere vogeltjes pakken. en de hond gaat de poesjes staart eventjes gaat niet eten maar tand poesje. en poesje ze vindt pijn. en de poesje rent weg en de hond ga ook weg. en dan vogeltjes moeder kijkt naar ze kindje en ze blij.

A big bird have two birdies and the she [her?] child. And the puss wants to catch this little bird. And she [her?] mother flies away. Like this [?] can but she pussy can climb to the tree. And bird looks at puss she is very afraid. And then pussy wants to catch the birdies. And she goes the tree climb wants the bird to catch. She mother is looking on the other bird, that is why she cannot see this puss grab the birdies. The dog has seen the pussy am going to grab bird and she mother am going to other she, the birdies cannot see and the pussy is going to catch the other birdies. And the dog is going the puss' tail, for a little while goes not eat but tooth [bite?] pussy. And pussy has pain. And the pussy runs away and the dog is also going away. And then birdies mother looks at the little child and she happy.

MLR indicates the use of infrequent words. From the comparison between the consecutive models for the MLR score (see Table 5.13) it is apparent that a model with a fixed linear

component – allowing for differences in Age – fit the data better than a model with only an intercept ( $\Delta\chi^2$  (MLR<sub>1</sub>) = 21.74;  $df = 1$ ;  $p < .001$ ). The variance within individuals depends on the age of participants ( $\Delta\chi^2$  (MLR<sub>2</sub>) = 59.20;  $df = 2$ ;  $p < .001$ ). The variance between individuals is a (linear) function of age as well ( $\Delta\chi^2$  (MLR<sub>3</sub>) = 14.64;  $df = 1$ ;  $p < .001$ ). Finally, adding Age<sup>2</sup> to the model improved the fit significantly ( $\Delta\chi^2$  (MLR<sub>4</sub>) = 7.34;  $df = 1$ ;  $p = .01$ ). Hence, in the final model (MLR<sub>4</sub>) a fixed effect of Age<sup>1</sup> and Age<sup>2</sup>, as well as variance within and between pupils components which depends on Age needs to be included, and with this model we continued the analysis.

Table 5.13: Fit of Different Models (-2LL) for Changes in Measure of Lexical Richness (168 cases) as well as the Comparison of Consecutive Models.

Model	-2LL	Comparison			
		Models	$\Delta\chi^2$	$\Delta df$	$p$
MLR <sub>0</sub> : $\beta_{0ij}cons^a$	759.14				
MLR <sub>1</sub> : $MLR_0 + \beta_1 Age^1_{ij}$	737.41	MLR <sub>0</sub> vs MLR <sub>1</sub>	21.74	1	<.001
MLR <sub>2</sub> : $MLR_1 + e_{1ij}Age^1_{ij}$	678.20	MLR <sub>1</sub> vs MLR <sub>2</sub>	59.20	2	<.001
MLR <sub>3</sub> : $MLR_2 + u_{10i}Age^1_{ij}$	663.56	MLR <sub>2</sub> vs MLR <sub>3</sub>	14.64	1 <sup>b</sup>	<.001
MLR <sub>4</sub> : $MLR_3 + \beta_2 Age^2_{ij}$	656.21	MLR <sub>3</sub> vs MLR <sub>4</sub>	7.34	1	0.01
MLR <sub>5</sub> : $MLR_4 + e_{2ij}Age^2_{ij}$	649.31	MLR <sub>4</sub> vs MLR <sub>5</sub>	2.74	3	.08ns
MLR <sub>6</sub> : $MLR_5 + u_{20i}Age^2_{ij}^c$					

<sup>a</sup> In addition to the intercept, variance components for differences within and between individuals are estimated

<sup>b</sup> Only the covariance-coefficient between the intercept- and the age-residuals was estimated.

<sup>c</sup> No convergence within 250 iterations.

Based on this General Development Model we constructed Figure 5.8, in which both the average development and the differences within and between individuals are represented (see Table 2.3 in Appendix 2 for the parameter estimates). The average MLR score at 73 months was 2.25. Each month a pupil grew older his or her MLR score increased by .07, but the quadratic function of age decreased the score by .003 (since this is a small effect it is not visible in Figure 5.8).

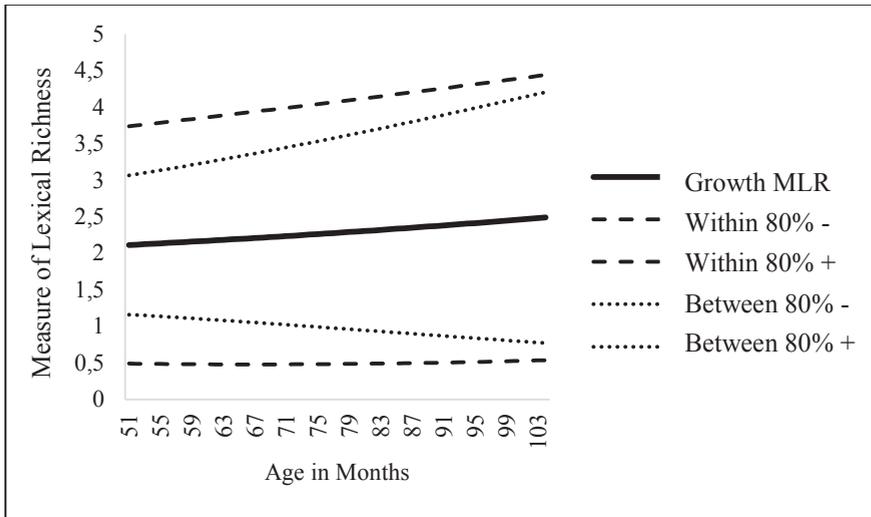


Figure 5.8: Graphical Representation of the General Development Model Measure of Lexical Richness with 80% Reliability.

Figure 5.8 shows that the effect of Age on MLR scores is significant. This is a quadratic relation. Furthermore, the differences within individuals (represented by the striped lines in Figure 5.8) also depend on Age; we can give a more precise estimate of the Measure of Lexical Richness for younger pupils as compared to older pupils. The differences between individuals (represented by the dotted lines in Figure 5.8) are a function of Age as well; differences between younger individuals are smaller than differences between of older pupils.

A likelihood ratio test showed that the main effect of Exposure to Dutch at School contributed significantly to the fit of the model to the observed data ( $\Delta\chi^2$  (MLR<sub>7</sub>) = 9.25;  $df$  = 1;  $p$  = 0.002; see Table 5.14): the longer the Exposure to Dutch at School, the higher the score on the MLR. No interaction between Age and Exposure to Dutch at School was found (MLR<sub>8</sub>).

Table 5.14: Fit of Different Models (-2LL) for Changes in Measure of Lexical Richness (168 cases) with Addition of Exposure to Dutch at School (EDS) as Explanatory Variable.

Model	-2LL	Models	Comparison		
			$\Delta\chi^2$	$\Delta df$	$p$
MLR <sub>4</sub> : $\beta_{0ij}cons + \beta_{1ij}Age^1_{ij} + \beta_{2ij}Age^2_{ij}$	656.22				
MLR <sub>7</sub> : MLR <sub>4</sub> + $\beta_3EDS_{ij}$	646.97	MLR <sub>4</sub> vs MLR <sub>7</sub>	9.25	1	.002
MLR <sub>8</sub> : MLR <sub>7</sub> + $\beta_4Age*EDS_{ij}$	646.81	MLR <sub>7</sub> vs MLR <sub>8</sub>	0.16	1	.69ns

The difference in MLR score when Exposure to Dutch at School was taken into account is graphically presented in Figure 5.9. The curve in Figure 5.9 is caused by the estimated parameters of Age and Age<sup>2</sup> (see Table 2.3 in Appendix 2) Since these estimates are unreliable due to large standard errors, the curve itself should not be taken into account by interpreting this graph. What Figure 5.9 does show is the effect of Exposure to Dutch at School on the MLR score: every increase in Exposure increases the MLR score, no matter how old the pupils are.

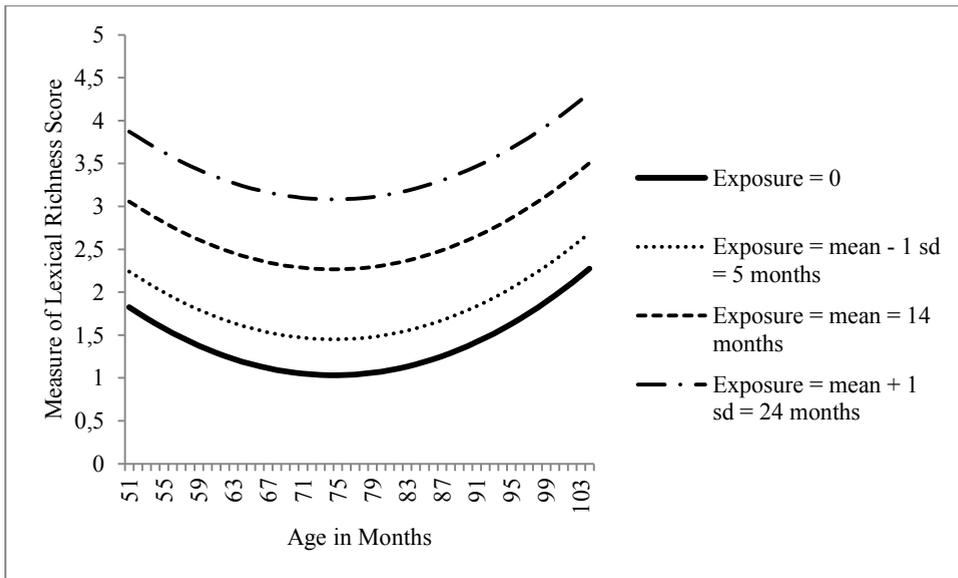


Figure 5.9: Development of MLR with Exposure to Dutch at School as Explanatory Variable.

The addition of type of Educational Facility did not improve the model fit of the General Development Model of Measure of Lexical Richness. This means we could not show that whether the pupil was at a DL2-school or a Mainstream school influenced the MLR score, the use of less frequent words.

#### 5.4.2 Macrostructure

The three elements of macrostructure will now be presented: Story Structure (SS), Structural Complexity (SC), and Internal State Terms (IST). For all three elements General Development Models will be built.

##### *Story Structure*

The Story Structure is the score for the included parts of story grammar in the total narrative. Example (7) and (8) illustrate this SS score variable. The stories in Examples (7)

and (8) have quite similar NDW score, respectively 55 and 52; nevertheless their SS score differs a lot. The story in Example (7) is from Session 3, the SS score is 12 (of maximum 17 points) while Example (8) from Session 0 has a SS score of 3. Even though Example (7) and (8) have almost similar NDW scores, the story in (8) is very unclear. This boy needed a lot of words to describe the story because he did not have enough Dutch vocabulary to make it clear, nevertheless it was hard to figure out what parts of the story structure the boy was telling.

In the transcripts the marking of the SS components is given in the following way: **setting** goal attempt outcome internal *state term as initiating event or as reaction*. The superscript numbers indicate to which episode the component belongs.

Example (7): subject number 0243; age 8;5; around 31 months in school; home language: Chinese; Session 3; NDW score 55; SS score 12.

**er was eens** twee geitjes. een geit valt **in de vijver**. en de andere geitje gaat de gras eten. en de moedergeit ging de geitje van water naar de gras duwen<sup>1</sup>. en er was een vos wil het lammetje eten<sup>2</sup>. en de lammetje zag ze niet. en toen gaat die andere geitjes de gras eten<sup>1</sup>. en *de moeder ben blij*. maar de andere geitje zag niet. en de vos gaat ze toen pakken<sup>2</sup>. en toen ging moedergeitje en de andere lammetje ging water drinken. en de vos had de lammetjespoot gepakt<sup>2</sup>. en *de raaf zag ze bent ook heel verdrietig*. en toen ging de raaf de vos staart happen pijn doen. en die moedergeit en die andere lammetje ging schrokken omdat de raaf ging toen gewoon rechtdoor de vos eten. en *toen vond vos heel pijn dus ze ga niet meer die kleine lammetjes eten*. en toen leefden ze lang en gelukkig.

**Once upon a time** there was two goats. One goat falls **into the pond**. And the other little goat is going to eat the grass. The mother goat went to push the little goat from the water to the grass. And there was a fox who wanted to eat the little lamb. And the little lamb did not see her [the fox?]. And then went the other little goats eat the grass. *And the mother am happy*. But the other little goat saw not. And the fox went to get she. And then went little mother goat and the other little lamb drink water. And the fox had taken the little lambs' paw. *And the raven saw she am also very sad*. And then went the raven the fox's tail bite hurt. And the mother goat and the other little lamb went scared because the raven went just straight-ahead eat the fox. And then *the fox was hurt thus she is not going to eat the little lambs any more*. And then they lived happily ever after.

Example (8): subject number 0501; age 4;3; around 2 months in school; home language: Rumanian; Session 0; NDW score 52; SS score 3.

een dier is **in de water**. in [?] een [?] weerde [?] eten [?]. en [?] nog [?] een [?] weerde [?] eten [?]. en nog twee is in de water. een zijn [?] naar de bloemen. waar is de dat wat kan pijn [?] doet. de dit de tong [?] naar buiten in de grote neus. en de vos wil de eten dat wat kan het pijn met die. de geit wil hij eten. maar is twee dat ie weet het. de dat wat hebbe [?] bij dat kleine wat kan het. de kleine geit is wil weg voor de wat heb neus grote. wil weg voor de vos voor de eten met de tanden. in die kleine heb zo. de geit die wat heb neus grote

wil hebbe gepakt voor voet geit, en de geit wil wegracen. ja maar huilen. de geit niet huilen. huilen niet. de geit is moe. in de die twee een grote nog een kleine is ook moe. en allemaal geiten nu is blij. en nu is een vogel met neus wat kan pijn doen. in die wat heb neus grote weg.

An animal is **in the water**. In [?] one [?] wanted [?] eat [?] and [?] another [?] wanted [?] eat [?]. And another two is in the water. And are [?] to the flowers. Where is that what can hurt [?], the this the tongue [?] outside in the big nose [snout?], and the fox wants that to eat that what can hurt it with that. He want to eat the goat. But is two that knows it. the that what has [?] with that little one what can it. The little goat is want away for the what has big nose. Wants away for the fox eats with his teeth. In that little has like this the goat that what has a big nose want to have grabbed the goat's foot. And the goat want to race away. Yes but crying. The goat not crying. Crying not. The goat is tired. In the those two a big one also a little one is also tired. And all goats are happy now. And now is a bird with nose [beak] what can hurt. In that what has snout big away.

From the comparison between the consecutive models for the SS score (Table 5.15) it is apparent that a model with a fixed linear component – allowing for differences in Age – fit the data better than a model with only an intercept ( $\Delta\chi^2$  (SS<sub>1</sub>) = 101.86;  $df = 1$ ;  $p < .001$ ). The variance within individuals depends on the age of participants ( $\Delta\chi^2$  (SS<sub>2</sub>) = 9.25;  $df = 2$ ;  $p = .01$ ). The variance between individuals appeared not to depend on Age (SS<sub>3</sub>). However, adding Age<sup>2</sup> to the model improved the fit significantly ( $\Delta\chi^2$  (SS<sub>4</sub>) = 16.94;  $df = 1$ ;  $p < .001$ , and the variance between individuals was dependent on age after all. Hence, in the final model (SS<sub>4</sub>) a fixed effect of Age<sup>1</sup> and Age<sup>2</sup>, as well as variance within and between pupils components which depends on Age needs to be included, and with this model we continued the analysis.

Table 5.15: Fit of Different Models (-2LL) for Changes in Story Structure (168 cases) as well as the Comparison of Consecutive Models.

Model	-2LL	Comparison		
		Models	$\Delta\chi^2$	$\Delta df$ $p$
SS <sub>0</sub> : $\beta_{0ij}cons^a$	886.11			
SS <sub>1</sub> : $SS_0 + \beta_1 Age^1_{ij}$	784.25	SS <sub>0</sub> vs SS <sub>1</sub>	101.86	1   <.001
SS <sub>2</sub> : $SS_1 + e_{1ij}Age^1_{ij}$	775.00	SS <sub>1</sub> vs SS <sub>2</sub>	9.25	2   .01
SS <sub>3</sub> : $SS_2 + u_{10j}Age^1_{ij}$	774.66	SS <sub>2</sub> vs SS <sub>3</sub>	0.35	1 <sup>b</sup> .56ns
SS <sub>4</sub> : $SS_3 + \beta_2 Age^2_{ij}$	757.71	SS <sub>3</sub> vs SS <sub>4</sub>	16.94	1   <.001
SS <sub>5</sub> : $SS_4 + e_{2ij}Age^2_{ij}$	754.02	SS <sub>4</sub> vs SS <sub>5</sub>	3.68	3   .30ns
SS <sub>6</sub> : $SS_5 + u_{20j}Age^2_{ij}^c$	753.02	SS <sub>5</sub> vs SS <sub>6</sub>	1.01	3   .80ns

<sup>a</sup> SS<sub>0</sub>: In addition to the intercept, variance components for differences within and between individuals are estimated

<sup>b</sup> Only the covariance-coefficient between the intercept- and the age-residuals was estimated

Based on this General Development Model we constructed Figure 5.10, in which both the average development as well as the differences within and between individuals are represented (see Table 2.4 in Appendix 2 for the parameter estimates). The average receptive vocabulary at an age of 73 months was estimated as 7.01. Each month a child grew older, his SS score increased by 0.24, but the quadratic function of age decreased the score by 0.01.

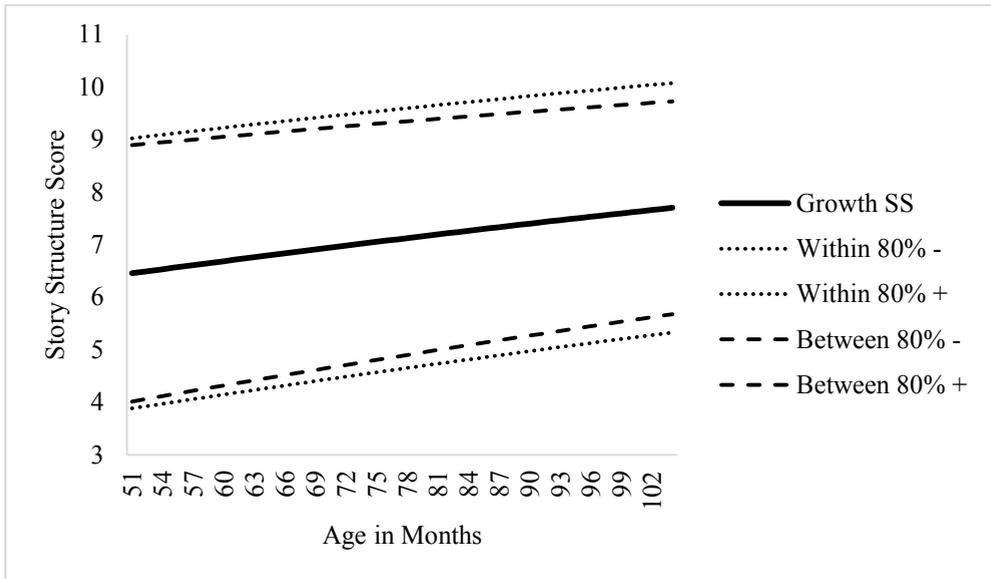


Figure 5.10: Graphical Representation of the General Development Model of Story Structure with 80% Reliability.

Figure 5.10 shows that the effect of Age on SS score is significant; this is a quadratic relation; the increase in SS is larger for younger pupils. The differences within individuals (represented by the striped lines in Figure 5.10) also depend on Age; we can give a more precise estimate of the Story Structure development for older pupils as compared to younger pupils. The differences between individuals (represented by the dotted lines in Figure 5.10) also depend on Age; the differences between younger pupils are larger than differences between older pupils.

Exposure to Dutch at School, measured as the number of months a pupil has attended a Dutch school (excluding summer months), was added to the model of SS. A likelihood ratio test showed that the main effect of Exposure to Dutch at School contributed significantly to the fit of the model to the observed data ( $\Delta\chi^2(SS_7) = 16.35$ ;  $df = 1$ ;  $p < .001$ ; see Table 5.16). The same holds for the interaction between Age and exposure ( $\Delta\chi^2(SS_8) = 23.82$ ;  $df = 1$ ;  $p < .001$ ).

Table 5.16: Fit of Different Models (-2LL) for Changes in Story Structure (168 cases) with Addition of Exposure to Dutch at School (EDS) as Explanatory Variable.

Model	-2LL	Comparison			
		Models	$\Delta\chi^2$	$\Delta df$	$p$
SS <sub>4</sub> : $\beta_{0ij}cons + \beta_{1ij}Age^1_{ij} + \beta_{2ij}Age^2_{ij}$	757.71				
SS <sub>7</sub> : SS <sub>4</sub> + $\beta_3EDS_{ij}$	741.36	SS <sub>4</sub> vs SS <sub>7</sub>	16.35	1	<.001
SS <sub>8</sub> : SS <sub>7</sub> + $\beta_4Age*EDS_{ij}$	717.54	SS <sub>7</sub> vs SS <sub>8</sub>	23.82	1	<.001

In Figure 5.11 the general development of SS with Exposure to Dutch as explanatory variable for pupils with different ages is presented (see Table 2.4 in Appendix 2 for the estimated parameters). There was a fixed main effect of Exposure to Dutch at School and an interaction effect between Age and Exposure to Dutch at School. Therefore, all lines in Figure 5.11 show an increase (remember that when pupils grew older, their Exposure to Dutch also increased one month), however, since the effect of Exposure is smaller for older pupils, the striped and dotted line for the oldest pupils shows less difference with the solid black line (the hypothetical 0 Exposure line) compared to the lines of the youngest and average aged pupils.

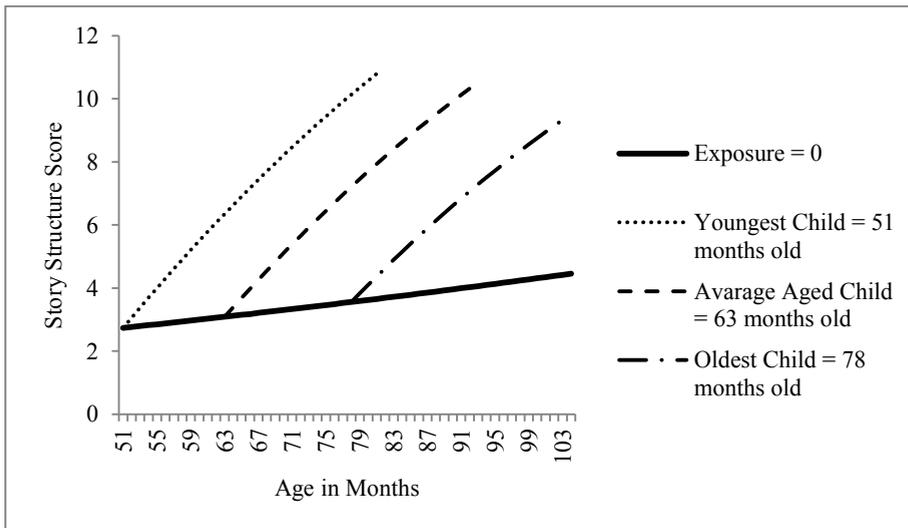


Figure 5.11: Development of Story Structure with Exposure to Dutch at School as Explanatory Variable.

The addition of the variable type of Educational Facility did not improve the fit of the model for Story Structure. This means that we could not show that stories of pupils who

stayed at a DL2-school were more coherent than pupils who stayed at a Mainstream school, respectively.

### **Structural Complexity**

In Examples (9) and (10) both participants score 12 points for the SS. Nevertheless, the story in Example (9) is more complex; since it includes two full Goal-Attempt-Outcome sequences (worth 3 points) and one Attempt-Outcome sequence (1 point). The SS score of the story in Example (10) is equal, but it includes more ISTs. The complexity of the episodes is lower since the first episode only includes an Attempt (0 points for SC), the second a Goal and an Outcome (2 points), and the third an Attempt and an Outcome (1 point). In the transcripts the marking of the SS components is given in the following way: **setting** goal attempt outcome internal *state term as initiating event or as reaction*. The superscript numbers indicate to which episode the component belongs.

Example (9): Subject number 1034; age 8;0; around 30 months in school; home languages: Arabic, French, and English; Session 3; SS score 12; SC score 7.

er was eens twee vogeltjes en een mama. en ze waren heel blij. maar de mama had die waren snel die waren al nu geboren dus daarom gaat mama ze vliegt en ze gaat lekkere wormpjes halen. en toen kwam de poes en poezen houden van vogels, vlees en dus net geboren vogels en van bessen [?] dus daarom toen de mama weg was, was poes gekomen. en toen de mama klaar was had ze een lekker koekje opgegeten. een lekker wormpje. en toen ze klaar was toen was poes aan boompje te klimmen om een vogeltjes te pakken. en toen kwam een hond. en de poes was helemaal boven en die probeerde een vogeltje te pakken. en de hond was een beetje boos. en toen de hond heeft aan de staart van de kat getrekt omdat hij is een beetje boos op hem en dus daarom trekt hij aan zijn staart. en toen schrikt de vogeltjes en de moeder. en toen heeft de hond de poes jaagt. en de poes was kei bang. en de hond was kei boos. en de kleine vogeltjes en de mama leefde lang en gelukkig.

**Once upon a time there were two birdies and a mama.** And they were very happy. But the mama had those were quick those were already now born. thus that is why mama is going to fly<sup>1</sup> and she is going to get jummy earthworms<sup>1</sup>. And then came the puss, and *pusses like bird, meat and, thus newly born birds* and like berries [?] thus that is why when the mother was away the puss came<sup>2</sup>. And then the mother was ready she had eaten a nice cookie. A nice earthworm<sup>1</sup>. And then when she was ready, then was puss climbing little tree to catch a birdie<sup>2</sup>. And then came a dog. And the puss was all the way on top and trying to catch a birdie<sup>2</sup>. *And the dog was a little angry.* And then the dog has pulled to the tail of the cat<sup>3</sup>, because he is a little angry at him and that is why he pulls his tail. And then the birdies are scared and the mother. And then has the dog cashed the puss<sup>3</sup>. *And the puss was very scared.* And the dog was very angry. And the little birdies and the mama lived happily ever after.

Example (10): Subject number 0501; age 5;6; around 18 months in school; home language: Rumanian; Session 2; SS score 12; SC score 3.

de schaaap is in het water. en de mama zag dat maar de broer niet. en toen kwam de mama de kleine schaaap weer uit het water halen. maar toen zag een wolf een schaaap. en toen wou de wolf de schaaap opeten. en toen ging de wolf rennen. en toen was de mama blij maar de broer van de kleine schaaap niet zo blij omdat de wolf die wil opeten. en toen ging de wolf de voet van de schaaap pakken. en toen zag een vogel dat. en toen ging de wolf op de grond liggen want de vogel ging de wolf bijten met zijn bek. en toen zag de kind en de mama had die deed. en toen was de broer een klein beetje verdrietig. toen ging de wolf wegrekken en de vogel achterna rennen. en toen was ze ze allemaal weer blij de schapen.

**The sheep is in the water.** *And the mama saw that but the brother [did?] not. And then came the mother the little sheep again get from the water<sup>1</sup>. But then saw a wolf a sheep. And then wanted the wolf eat the sheep<sup>2</sup>. And then the wolf went to run. And then the mama was happy but the brother of the little sheep not so happy because the wolf wanted to eat that one. And then the wolf went to grab the foot of the sheep<sup>2</sup>. And then saw a bird that. And then went the wolf lay on the ground because the bird was going to bite the wolf with his beak<sup>3</sup>. And then saw the child and the mama had that one did. And then was the brother a little bit sad. Then the wolf went to run away<sup>3</sup>. And the bird run after. And then they were all very happy, the sheep.*

The reliability of the model for SC is estimated as .39<sup>15</sup>. Hence, at least 61% of the observed differences do not relate to SC but can be seen as random noise (error of measurement) instead. Therefore, we refrained from building a General Development Model and did not investigate what the effects of child characteristics are in this General Development Model of SC.

### **Internal State Terms**

ISTs are words mentioning different “states” of the characters of the story. Example (11) and (12) illustrate our variable Internal State Terms. Both examples the participants tell the Baby-Goat story and have relatively high NDW scores, respectively 52 and 45. Nevertheless, the story in Example (11) includes 14 ISTs and Example (12) only 3 ISTs. The bold words indicate the ISTs.

Example (11): subject number 0331; age 6;1; around 10 months in school; home language: Polish; Session 0; NDW score 52; IST 14.

hier ben **verdrietig** daarom deze baby hij is gevalt in de water. deze **zien** niet. en nu hij **wil** help nodig hebben. en hij ook help nodig heb. en papa heb hij zo in de water zo en dan zo maken. en hier heb papa zo gemaakt. en hier is baby nu beetje **blij**. en hij nog eten. dan

---

<sup>15</sup> We estimated the reliability for all models of each of the other languages measures. The reliability of the model was only reported when it was below .50.

komt de vos. en dan deze hoort niet. en deze hoort niet. en deze hoort niet. en hij is ook nog beetje zo **blij**. en hij moet nog nog hij duwen. en hij **wil** eten. en hij is **geschrokken**. en hij is **blij** daarom hij eten. en vos heb gespring naar deze zo deze babietje. en nu deze vos hij pakken van de hij been. hier heb vos gepakt. en deze papa heb hier met de andere jongetje. deze water drink. en die vos heb gepakt op hij been. en hij is **bang**. daarom hij hij eten. en hij zo beetje **verdrietig** deze vogel. en hier deze vogel heb hij au gemaakt zo. en hier deze vos heb nu au. en deze andere heb nu **verdrietig** van deze vos. en deze is ook **verdrietig** van deze vos. en hier deze babietje en jongetje is nu **blij**. en hij ga **kijken** of hij die vos ga niet andere keer hier komen.

Here am **sad** that's why this baby he has fallen [wrong Dutch inflection] in the water. This one **see** not. And now he **want** help need. And he also need help. And daddy have he like this in the water like this and then make like this. And here have daddy made like this. And here is baby now a little bit **happy**. And he still eats. Then comes the fox. And then this one hears not. And this one hears not. And this one hears not. And he is also still a little like this **happy**. And he must still still he [him] push. And he **wants** to eat. And he is **scared**. And he is **happy** that is why he eat. And fox have jumped to this one like this this little baby. And now this fox he grab of the he [his] leg. Here have fox grabbed. And this daddy have here with the other little boy. This one drink water. And that fox have grabbed on he [his] leg. And he is **afraid**. That is why he he eat. And he like this a little **sad** this bird. And here this bird have he done 'au' [hurt him] like this. And here this fox have now 'au' [pain]. And this other have now **sad** of this fox. And this one is also **sad** of this fox. And here this little baby and little boy is now **happy**. And he go **looking** if he that fox do not come here another time.

Example (12): subject number 1012; age 5;8; around 18 months in school; home language; Arabic; Session 2; NDW score 45; IST 3.

hier is de lammetje. die kleine lammetje kan die niet zwemmen en die is die gevallen in de water. en toen gaat die grote lammetje da(de)lijk hem halen naar boven. naar de gras. en toen duwt die kleine lammetje met zijn hoofd. en die andere lammetje gaat ie eten. en er was iemand en die gaat die kleine lammetje opeten, maar van andere. en toen ging die afpakken met zijn benen. en toen **wil** die opeten zijn benen. en toen ging de lammetje **bang** worden. en die ging en iemand hem opeten. en toen ging die beetje **boos** die lammetje. en toen ging z'n die tweeën, die kleine en die grote naar die kleine gaan. en toen ging daar samen vriendjes.

Here is the little lamb. That little lamb can that one not swim and that one is that one fallen in the water. And then goes that big little lamb soon get him up [from the water?]. To the grass. And then pushes that little lamb with his head. And that other little lamb is he going to eat. And there was someone and that one goes to eat that little lamb, but the other. And then went that one take away with his legs. And then **want** that one to eat his legs. And then went the little lamb become **scared**. And that one went and someone eat him up. And then went that one little **angry** that little lamb. And then went that two, that little and that big one to that little one. And then went there friends together.

From the comparison between the consecutive models (see Table 5.17) it is apparent that a model with a fixed linear component – allowing for differences in Age – fit to the data better than a model with only an intercept ( $\Delta\chi^2$  (IST<sub>1</sub>) = 20.13;  $df = 1$ ;  $p < .001$ ). The variance within individuals did not depend on the age of participants (IST<sub>2</sub>). The variance between individuals is a function of Age ( $\Delta\chi^2$  (IST<sub>3</sub>) = 6.10;  $df = 2$ ;  $p = .05$ ). However, the inclusion of Age<sup>2</sup> (IST<sub>4</sub>) did not improve the model. Hence, in the final model (IST<sub>DEF</sub>) a fixed effect of Age, as well as a variance between pupils component which depends on Age needs to be included, and with this model we continued the analysis.

Table 5.17: Fit of Different Models (-2LL) for Changes in Internal State Terms (168 cases) as well as the Comparison of Consecutive Models.

Model	-2LL	Comparison			
		Models	$\Delta\chi^2$	$\Delta df$	$p$
IST <sub>0</sub> : $\beta_{0ij}cons^a$	10.19.84				
IST <sub>1</sub> : IST <sub>0</sub> + $\beta_1 Age^1_{ij}$	999.71	IST <sub>0</sub> vs IST <sub>1</sub>	20.13	1	<.001
IST <sub>2</sub> : IST <sub>1</sub> + $e_{1ij} Age^1_{ij}$	995.98	IST <sub>1</sub> vs IST <sub>2</sub>	3.73	1	.05ns
IST <sub>3</sub> : IST <sub>2</sub> + $u_{1ij} Age^1_{ij}$	989.88	IST <sub>2</sub> vs IST <sub>3</sub>	6.10	2 <sup>b</sup>	.05
IST <sub>4</sub> : IST <sub>3</sub> + $\beta_2 Age^2_{ij}$	989.21	IST <sub>3</sub> vs IST <sub>4</sub>	0.68	1	.41ns
IST <sub>DEF</sub> : IST <sub>0</sub> + $\beta_{1j} Age^1_{ij}^c$	991.35				

<sup>a</sup> In addition to the intercept, variance components for differences within and between individuals are estimated

<sup>b</sup> Only the covariance-coefficient between the intercept- and the age-residuals was significant.

<sup>c</sup> In this model only the between participants covariance-coefficient between the intercept- and the age-residuals was included.

Based on this General Development Model we constructed Figure 5.12, in which both the average development as well as the differences within and between individuals are represented (see Table 2.5 in Appendix 2 for the parameter estimates). The average IST score at an age of 73 months was estimated as 5.45. Each month a pupil grew older his receptive vocabulary increased by 0.14.

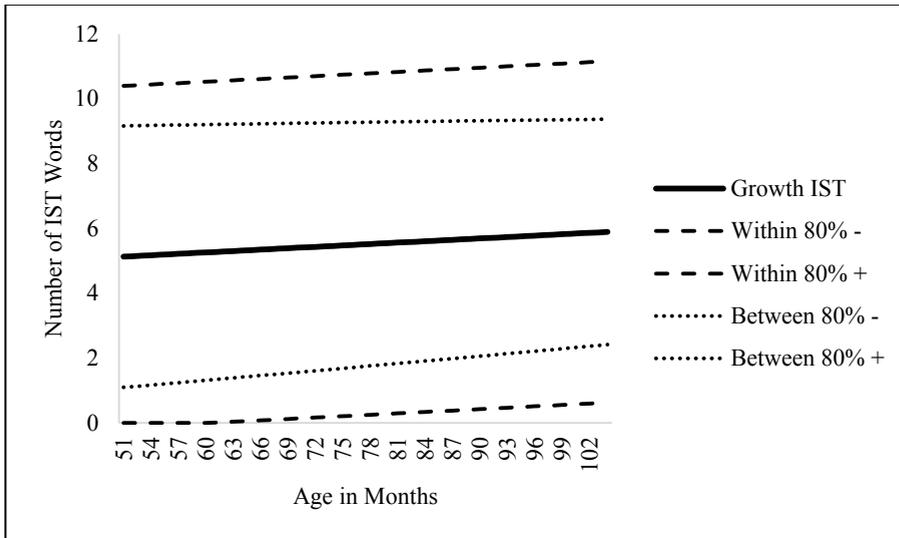


Figure 5.12: Graphical Representation of the General Development Model of IST with 80% Reliability.

Figure 5.12 shows that the effect of Age on IST is significant. The differences within individuals (represented by the striped lines in Figure 5.12) did not depend on Age. The difference between individuals (represented by the dotted lines in Figure 5.12) however was a function of Age; differences between older individuals in Internal State Terms are somewhat smaller than between younger pupils.

The addition of the variables Exposure to Dutch at School and type of Educational Facility did not improve the general model of growth in IST score. This means that we could not show that stories of pupils who had more Exposure to Dutch at School or who stayed at a DL2-school included more ISTs than pupils who had less Exposure to Dutch or who stayed at a Mainstream school.

### 5.5 Summary and conclusion

The main goal of this chapter was to analyze how the narrative ability of young newly arrived migrant kindergarteners developed in Dutch in the first two-and-a-half years after arriving in the Netherlands, answering the question: How does the *narrative ability* (in Dutch) of newly arrived migrant kindergarteners develop during the first two-and-a-half years of schooling in the Netherlands in relation to school type? We wanted to know how each aspect of narrative ability microstructure (Number of Different Words, Guiraud Index Score, and Measure of Lexical Richness) as well as macrostructure (Story Structure, Structural Complexity, and Internal State Terms) developed over time.

It was possible to build a General Development Model for five out of six narrative ability measures. Table 5.18 summarizes the significant effects of the variables on the five narrative aspects. Of the 20 possible significant effects, 12 appeared to be significant.

Table 5.18: Overview of Significant Child Characteristics on the Narrative Ability Models.

	Microstructure			Macrostructure	
	NDW	GIS	MLR	SS	IST
Age	+	+	+	+	+
Age <sup>2</sup>			+	+	
Exposure to Dutch at School	+ <sup>a c</sup>	+ <sup>a</sup>	+ <sup>a c</sup>	+/-	
Educational Facility	- <sup>b</sup>				

NDW = Number of Different Words; GIS = Guiraud Index Score; MLR = Measure of Lexical Richness; SS = Story Structure; IST = Internal State Terms. +: positive effect; -: negative effect; +/-: positive fixed effect and negative interaction effect.

<sup>a</sup> Only the fixed main effect of the variable was significant.

<sup>b</sup> Only an interaction effect with Age was found.

<sup>c</sup> When Exposure is added to the model, Age seemed to be no longer significant.

Age is significant for five out of six narrative measurements, meaning that there is a development with age for these narrative ability measures, as was predicted based on for example Gagarina (2016) and Bohnacker (2016). For NDW, GIS, and IST this development is linear, while the growth rate for MLR and SS decreases over time. The increase in MLR and SS score is larger for younger pupils than for older individuals. For none of the measures a ceiling effect was observed.

Beside Age, Exposure to Dutch at School seemed to be an important factor as well. It influenced all measures except for the number of ISTs. For NDW and MLR the effect of Exposure to Dutch at School even seemed to be of more importance than the effect of Age. The longer a pupil attended a Dutch school, the higher his or her score on the different measures.

For SS there was also an interaction between Age and Exposure to Dutch, thus the influence of Exposure of Dutch at School was larger for younger pupils than for older pupils. Altman et al. (2016) found in their study that length of exposure had no influence on SS. However, it is difficult to directly compare their outcomes with the participants in the current study. That is, in the Altman et al. (2016) study, the mean months of exposure was 45.26 (*sd.* 20), while the participants in our study had a mean of only 29.2 months of exposure (*sd.* 1.9) by the final session. Nevertheless, in the present study it seems that there was an interaction between Age and Exposure to Dutch at School for SS. Therefore, it might have been that if we had followed our participants for a longer period we would have reached the same conclusion as Altman et al. (2016).

With our data we have shown how the narratives of 42 newly arrived migrant pupils in their second language developed. There was only one effect of Educational Facility and that was an interaction between Age and Education Facility for the NDW; for younger pupils the influence of Educational Facility is larger, than for older pupils on the NDW. However, due to our small sample size, this interaction affect should be reproduced by another study in order to be validated.

This chapter with data on a narrative task was included in this dissertation in order to collect information on the productive second language development of newly arrived kindergarteners. With the analysis of the microstructure and the macrostructure of the stories we captured the linguistic competence as well as the discourse competence.

The outcome that the inclusion of type of Educational Facility did not have a clear-cut effect on any of the measures of narrative ability in this chapter, or on the receptive vocabulary as was found in Chapter 4, seems to justify a more detailed investigation of the school learning environment. Being at a DL2-school or at a Mainstream school in itself does not seem to have a large differential effect on the second language development of newly arrived migrant kindergarteners, regarding receptive vocabulary and narrative ability as measured with PPVT (Chapter 4) and MAIN (Chapter 5). However, there could be certain aspects of the schools' environments that differ between the types, and which could be influencing factors on language development. Chapter 6 and 7 will investigate how the school learning environment of newly arrived migrant pupils is composed, with a focus on pedagogical practices. After that, in Chapter 8, the General Development Models of the receptive vocabulary and the narrative ability measures will return and then they will be related to the specific significant characteristics of the school learning environment from Chapter 6 and 7.

## CHAPTER 6

---

### **Pedagogical practices: focus on teacher behavior**

---

## 6.1 Introduction

There is general agreement that the quality of the school learning environment is linked to child development outcomes (see, for example Yoshikawa et al., 2013 for a summary of research on this topic). To measure the quality of a program or a learning environment, many researchers focus on observed social and instructional interactions and on transactions among teachers, children, and materials, known as process quality (Bronfenbrenner & Morris, 1998; Howes & Ritchie, 2002; Pianta et al., 2005; Pianta, 1999; 2003). The exchanges between teachers, children, and materials in early education settings also influence language development (e.g., Howes et al., 2008; Mashburn et al., 2008; Hamre & Pianta, 2007).

Factors of the school learning environment concern among other things, the quality and the quantity of the language input which depends on interactions in the L2. The main source of input in the L2 for most newly arrived migrants comes from the interactions they have at school. Therefore, in the present study the pedagogical practices of the school learning environment of second language learners, and the possible influence they have on second language development is examined in more detail.

In both Chapter 6 and 7 the school learning environment of newly arrived migrant kindergarteners will be reported. In the present chapter the focus is on teacher behavior, while Chapter 7 will focus on focal pupils' experiences. The central question of Chapter 6 is:

What are the differences in characteristics of the school learning environment regarding teacher behavior between DL2-schools and Mainstream schools?<sup>16</sup>

The present chapter is structured as follows: before discussing the data collection, a brief review is given (Section 6.2) about what is already known about the influence of teacher behavior on child-outcomes using the Classroom Assessment Scoring System (CLASS; Pianta, La Paro, & Hamre, 2008). Then the methodology (Section 6.3) and results (Section 6.4) of data collected from 17 participating teachers from 10 different schools is discussed. The chapter ends (Section 6.5) with a summary and discussion, which will be extended in Chapter 9.

## 6.2 Pedagogical practices: focus on teacher behavior

### 6.2.1 Previous research with the CLASS

Since 2008 the CLASS (Pianta et al., 2008) has been used in many different studies concerning analysis of classroom quality, to evaluate the quality of education and care

---

<sup>16</sup> This research question was written in Chapter 3 as sub-question (c).

programs in the United States. Based on the results of the CLASS, policy decisions have been made to further instruct teachers. The different dimensions of the CLASS – emotional support, organizational support, and instructional support – are also used to find relations between teacher behavior and the development of specific skills.

The CLASS instructional scales predict growth in language and literacy skills at the kindergarten level (Howes et al., 2008; Mashburn et al., 2008) and in Grade 1 (Hamre & Pianta, 2007). The instructional support domain has been shown to be especially important for the development of children's languages skills (Mashburn et al., 2008; Cadima, Leal, & Burchinal, 2010). That said, the influence of emotional support should not be underestimated, because in classes with a high level of emotional support children seem to explore more and therefore learn better in such environments (Hamre et al., 2013). Finally, organizational support influences learning, because in well-organized classes pupils seem to be more engaged in learning and teachers tend to make better use of the time (Hamre et al., 2013; Pianta et al., 2014).

The focus of the CLASS measure is the quality of interactions that teachers offer their pupils (Hamre & Pianta, 2007). That the CLASS measures something more than overall classroom quality was demonstrated in a study of Mashburn et al. (2008). They related three aspects of the quality of pre-kindergarten programs to child outcomes on five measures of academic or language skills. (1) They investigated features of program design and infrastructure (for example teacher qualifications and class size), (2) they made a comprehensive observation of the overall quality of the classroom environment (with the Early Childhood Environment Rating Scale- Revisited; Harms, Clifford, & Cryer, 1998), and (3) they investigated the nature and quality of the emotional and instructional interactions between the teachers and the children (with the CLASS). The third measure of quality, that of teacher-child interactions using the CLASS was most consistently and strongly associated with children's development.

The CLASS has been developed for educational settings in the United States. In the study by Downer et al. (2012) it was shown that the CLASS was also a good observational instrument in educational environments with second language learners, based on a study with almost 3,000 children attending state-funded pre-kindergarten programs in different states in the USA. In the classes that they investigated with a high percentage of Latinos or with various levels of dual language learners, higher scores on the CLASS domains were also linked to positive developmental outcomes in mathematics, language/literacy, and social skills of these second language learners.

In order to investigate whether the CLASS was also a useful instrument outside the United States, Pakarinen et al. (2010) used the CLASS in Finland. They provided evidence that the three domains of the CLASS could be applied to educational settings in Finland as well. They concluded that the CLASS has high validity and reliability in the assessment of classroom quality in a cultural and educational setting outside the United

States. It is important that an instrument such as the CLASS has validity worldwide because it allows for comparison between countries.

The use of the CLASS as an observational instrument for classroom quality seems to be also justifiable in the present study with newly arrived migrant pupils in the Netherlands since Slot, Boom, Verhagen, and Leseman (2017) showed that the CLASS was also an adequate measurement of early care and education in the Netherlands. The results of the CLASS in the different classes can be used in this specific context for evaluation of differences and similarities between schools. The CLASS will be used in the present study to see whether DL2-schools differ from Mainstream schools with respect to pedagogical practices focusing on teacher behavior.

### **6.2.2 Dutch studies using the CLASS**

In the Netherlands the CLASS is used in the national Cohort study Pre-COOL, since 2011 (N = 3000 pupils in 250-300 classes; e.g., Veen et al., 2012b). In this cohort study the CLASS was used to determine quality of day care, preschool, and kindergarten education<sup>17</sup>, and to relate the CLASS scores to child development in the domains of language, mathematics, and executive functions. Several publications show relations between the CLASS scores and developmental progress.

For example, Henrichs and Leseman (2016) found that Dutch kindergarten teachers (N = 96) had a mean score of 5 for emotional support, which is at the high end of the medium range on the CLASS scale of maximum 7 points. The mean for instructional support was lower: 3, which is at the low end of the medium range. Within the instructional support domain Henrichs and Leseman report a large variation between the teachers. This pattern of good emotional support, but medium instructional support is comparable with what is described in international research using the CLASS. According to Henrichs and Leseman the medium score for Instructional support shows that on that point the quality of kindergarten education in the Netherlands is higher compared to international studies (compare NICHD ECCRN, 2005; Pakarinen et al., 2010; von Suchodeltz, Fäsche, Gunzenhauser, & Hamre, 2014). Furthermore, the CLASS scores were related to child outcomes: Higher instructional support had a significant, although small to medium, positive effect on vocabulary and mathematical skills.

Veen, van der Veen, van Schaik, and Leseman (2017) found similar patterns in quality as Henrichs and Leseman (2016) with medium scores on emotional support but lower scores on instructional support in Dutch kindergarten classes. Furthermore, they found a small positive effect of behavior regulation and productivity in kindergarten on attention development. For vocabulary they only found a very small effect. For kindergarteners with a non-low-educated mother higher behavior regulation relates to lower

---

<sup>17</sup> See Section 1.4 in Chapter 1 for an explanation of the Dutch educational system.

vocabulary scores. One goal of Veen et al. (2017) was also to see whether there were differences in process quality between teachers who used an officially acknowledged early care program compared to teacher who used a different program: no differences were found.

Other research with the CLASS in the Netherlands focused on younger children compared to this study and therefore used a different version of the CLASS, the Toddler version. In that version of the CLASS, two domains can be distinguished: (1) Emotional and behavioral support and (2) instructional support. For Emotional and Behavioral Support medium and high scores are reported compared to low to medium scores for instructional support (Slot, 2014; Leseman & Slot, 2013; Veen et al., 2013; Van Schaik, Leseman, & De Haan, 2018). The use of a specific program which would stimulate the development of focal pupils was analyzed by Slot and Leseman (2013) in relation to the dimensions of the CLASS Toddler. They found that the use of special program did not have a systematic relation with the emotional or educational process quality.

All in all, what these studies show about Dutch kindergarten classroom is that instructional support should improve the most, compared to emotional support and organizational support. For example teachers give pupils the opportunity to brainstorm on ideas, but they do not stimulate to execute the ideas and to evaluate their predictions. Another example is that teachers do follow-up on a question, but they do not persist in the exchange with a child, after one response a different child is questioned. Even though the scores on language modeling are not the lowest, Henrichs and Leseman (2016) report that the most given advice was on how to improve the way teachers can stimulate language development.

## **6.3 Method**

### **6.3.1 Participants**

In the present chapter concerning teacher behavior the teachers of the 42 participating pupils in Chapter 4 and 5 participated. Seventeen female teachers from ten different schools agreed to be observed with the CLASS, see Table 6.1 (see also Section 3.4.1 in Chapter 3 for more details on the participating schools and teachers).

Table 6.1: Distribution of Participating Teachers (N = 17) over the two School Types (N = 10).

	Number of Teachers	Number of Observations
<b>DL2-schools</b>		
School 1	1	5
School 2	3	12
School 3	1	6
School 4	3	13
School 5	3	15
Total DL2-schools (5)	11	51
<b>Mainstream schools</b>		
School 6	2	10
School 7	1	4
School 8	1	4
School 9	1	5
School 10	1	5
Total Mainstream schools (5)	6	28

Of these 17 teachers, 11 teachers worked at a DL2-school and 6 at a Mainstream school. The 11 DL2-school teachers can be split up into 9 working in an independent DL2-school and 2 in a DL2-class within a Mainstream school. Since the distribution of the DL2-school teachers over schools and classes was too skewed, both separate schooling types are taken together in our analysis (see section 3.4.1 in Chapter 3).

### 6.3.2 Measurement

To study process quality of several dimensions of teacher-child interactions in classrooms, the Classroom Assessment Scoring System (CLASS; Pianta et al., 2008) was used. This assessment instrument is theoretically driven by the CLASS Framework of Children's Learning Opportunities in Early Childhood and Elementary Classrooms (Hamre & Pianta, 2007). The CLASS Framework is on many aspects not different from other descriptions of classroom quality put forth in educational literature; however, the CLASS Framework differs from others because of the unique theoretical grounding and the inclusion of early childhood and elementary classroom settings. Furthermore, the framework is empirically validated. The CLASS is translated to Dutch in cooperation with the American authors, and in the present study this approved Dutch translation was used.

The CLASS is theorized as an assessment of the classroom as a learning environment. It measures "the nature and form of the emotional and instruction climate of the classroom" (Pianta et al., 2005, p. 145). The CLASS was designed to address limitations of other instruments. The creators of the CLASS had the opinion that previous classroom assessment instruments focused primarily on the physical environment and

materials instead of the emotional and instructional aspects of the processes in the classroom (La Paro, Pianta, & Stuhlman, 2004).

The CLASS focuses on a concept of interaction which is known as proximal classroom process (Bronfenbrenner & Morris, 2006). Proximal refers to the assumption that the actions and interactions that are called proximal are direct determinants of children’s learning and development (see Section 2.7.1 in Chapter 2).

The scores of the CLASS represent the experiences of an average pupil in the class. It measures the quality of the teachers’ interactions with the pupils in class in three domains: Emotional Support, Organization Support, and Instructional Support (Hamre et al., 2013; Hamre & Pianta, 2007; Pakarinen et al., 2010; Pianta et al., 2015). These three domains are composed of ten dimensions: (1) *Positive Climate*, (2) *Negative Climate*, (3) *Teacher Sensitivity*, (4) *Regard for Student Perspectives*, (5) *Behavior Management*, (6) *Productivity*, (7) *Instructional Learning Formats*, (8) *Concept Development*, (9) *Quality of Feedback*, and (10) *Language Modeling* (see Table 6.2 and Appendix 3). For each dimension, three to four behavioral indicators are defined that have to be checked during the observations.

Table 6.2: Schematic Representation of the CLASS Quality Domains and Dimensions

<b>The CLASS</b>
<b>Emotional Support</b>
(1) Positive Climate
(2) Negative Climate
(3) Teacher Sensitivity
(4) Regard for Student Perspectives
<b>Organizational Support</b>
(5) Behavior Management
(6) Productivity
(7) Instructional Learning Formats
<b>Instructional Support</b>
(8) Concept Development
(9) Quality of Feedback
(10) Language Modeling

Emotional Support includes the first four dimensions: Positive Climate, Negative Climate (reversed for analysis), Teacher Sensitivity, and Regard for Student Perspectives. Positive Climate reflects the genuine enthusiasm, enjoyment, and respect demonstrated during interactions between the teacher and children, but also among the pupils in the classroom. Negative Climate is the degree to which the classroom has a negative emotional and social atmosphere: whether there are displays of for example anger, aggression, or harshness. Teacher Sensitivity is the extent to which teachers provide comfort, reassurance, help, and

encouragement to the pupils. Regard for Student Perspectives reflects the extent to which classroom activities can be initiated by the pupils and to what extent the teacher includes the ideas of the pupils in the activities.

Organizational Support is measured by three dimensions: Behavior Management, Productivity, and Instructional Learning Formats. Effective Behavior Management includes the teacher's ability to use methods to prevent and redirect children's misbehaviors effectively. Productivity reflects how well the teacher manages instructional time and routines so that there is more time available for learning. Instructional Learning Formats includes the availability of activities, methods of presentation, use of groupings, and the range of materials that teachers use to maximize pupil's engagement.

Instructional Support is also measured by three dimensions: Concept Development, Quality of Feedback, and Language Modeling. Concept Development considers the strategies teachers use to promote pupil's higher order thinking skills and creativity through problem solving, integration, and instructional discussions. Quality of Feedback focuses on the quality of verbal evaluation provided by the teacher to the pupils about their work, process, comments, and ideas. Language Modeling is about the amount of sustained conversations and open questions in the classroom and how rich the language the teacher uses is.

### **6.3.3 Procedure**

All 17 participating teachers were assessed using the CLASS during multiple activities on one morning from 8:30 to 12:00. For most teachers, 4 to 6 activities were observed (which corresponds to 4 to 6 different CLASS observations), making a total of 79 observations with the CLASS (see Table 6.1). One of the teacher had planned to assess pupils individually which did not seem to be a proper activity for a CLASS observation and, therefore, only 2 activities with that teacher could be used for observation with the CLASS during the morning in that classroom.

Following the CLASS manual, the separate observations were done in periods of 20 minutes. After taking notes for the 20-minute observation period, the observer left the classroom, analyzed the notes, filled out the observation form, and calculated the scores. After approximately 10 to 20 minutes, the observer returned to the classroom for a subsequent observation round. All observations for each participating teacher were compiled to compute a mean score for all different CLASS dimensions per teacher, following the prescribed CLASS procedure.

### **6.3.4 Analyses**

#### **The CLASS scores**

The score on each of the ten CLASS dimensions (See Appendix 3) was based on scores on different behavioral indicators for that dimension. These scores were based on detailed descriptions in the CLASS protocol with which indicators were scored “low,” “medium,” or “high.” The dimensions of the CLASS then received a score using a 7-point Likert scale, with 1 and 2 meaning a low score, 3–5 meaning a medium score (where 3 means a low-medium score and a 5 a high-medium score), and 6 and 7 a high score. A low score indicated that behavioral indicators were not or seldom seen during the observation. Furthermore, it could mean that there were many missed opportunities during the observational cycle, meaning that a teacher could have used an opportunity to show indicators of that dimension, but did not. A medium score meant that some indicated behavior had been observed, but not in a consequent manner. A high score meant that the behavior had been observed often and consequently during the observation cycle. One dimension, Negative Climate, had a reversed score, meaning a score of 1 or 2 was a good score and thus a score of 6 or 7 meant there was a very negative climate in the classroom. For the sake of analysis, the scores for Negative Climate were reversed.

#### **Statistical analysis**

The results will be presented for the two different school types (DL2-schools and Mainstream schools). With Linear Mixed models, using SPSS 22, the effect of school type was tested, taking into account differences between teachers within groups. The analysis thus contains a comparison between the scores of the group of 11 DL2-school teachers and a group of 6 Mainstream school teachers.

#### **Rater reliability**

All observations with the CLASS in the present study were completed by one observer; therefore no inter-rater reliability can be calculated. Nevertheless, this observer was trained by a certified trainer of the CLASS. After the training, a test was taken and the observer met the high reliability requirements (80% agreement) with a deviation of one scale point being allowed, following the protocol of the CLASS.

## **6.4 Results**

### **6.4.1 Quantitative analysis of the CLASS**

Looking at Figure 6.1, a comparable pattern in both school types can be seen for the scores on the 10 dimensions of the CLASS (for a description of these dimensions see Appendix 3). In both school types the teachers had the highest quality on Negative Climate (reversed coded, mean = 6.55), meaning that there was very little negativity in the classrooms in this

study. Positive Climate, Teacher Sensitivity, Regard for Student Perspectives, Behavior Management, Productivity, and Instructional Learning Formats were scored in the mid-range of quality (range = 3.67–4.85). Concept Development, Quality of Feedback, and Language Modeling showed low quality (range = 2.36–2.96). All dimensions were normally distributed except for Concept development (Shapiro-Wilk,  $p = .002$ ).

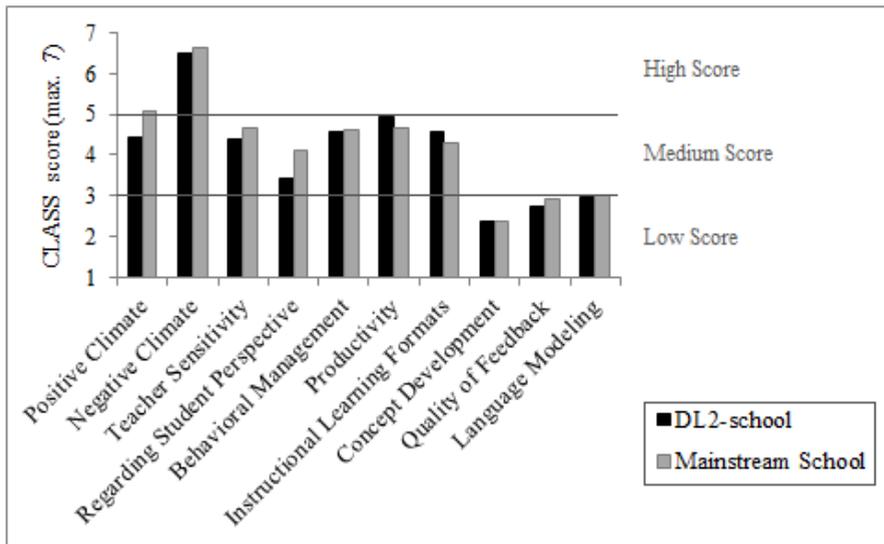


Figure 6.1: Mean Scores for the ten Dimensions of the CLASS.

The boxplot in Appendix 4 shows that there was variation between the teachers in the observed scores. The largest variation was in the dimensions of Instructional Learning Formats for a DL2-school. The scores ranged from 2.60 to 5.60. The smallest variation was found in the Positive Climate dimension for Mainstream Schools. The scores of the Mainstream school teachers ranged from 4.75 to 5.44.

An independent samples t-test was used to compare the means of the two school types on these ten dimensions. Since our sample is small with only 17 teachers we used a significance level of .1 (Mellenbergh, 1976). It appeared that the difference in score was significant for Positive Climate ( $t(15) = 1.93$ ;  $p = .07$ ) and for Regard for Student Perspectives ( $t(15) = 2.12$ ;  $p = .05$ ). For the other variables no difference could be shown ( $t(15) < 1.13$ ;  $p = .28$ ). Teachers at Mainstream schools score higher on both Positive Climate as well as Regard for Student Perspectives compared to DL2-school teachers.

#### **6.4.2 Qualitative analyses of the CLASS**

In addition to the quantitative data in the previous section this section will describe the data more qualitatively, providing some examples of teacher behavior. The following are descriptions of the behavior of the highest and lowest scoring teachers in the three domains: Emotional Support, Organizational Support, and Instructional Support. These descriptions illustrate in more detail the pedagogical practices that were observed in the classrooms.

The teacher who had the highest mean score on Emotional Support (5.44) was a teacher at a Mainstream school. In her classroom, positive climate was the highest. She used positive nicknames for the pupils, there were many shared activities, and the teacher shared the pupils' enthusiasm. Another teacher who scored high (5.40) on Emotional Support was from another Mainstream school. She was sensitive to the pupils and showed high regard for the perspectives of the pupils. Furthermore, she acknowledged their emotions and reactions in class by reacting to them with an explanation or she shared their enthusiasm.

The teachers who had the lowest scores on the domain Emotional Support were two teachers at two different DL2-schools. With their mean scores of respectively 4.10 and 4.13 they scored in mid-range, but for emotional support these are regarded low scores (cf. Buell, Han, & Vukelich, 2017; Pakarinen et al., 2010). On none of the occasions they scored 5 points or higher on one of the dimensions of this domain. Especially the dimension Regard for Pupils' Perspectives could be improved with more flexibility towards the pupils (follow their initiatives) and pupil-centered focus.

Two teachers who had the highest score on Organizational Support were from the same DL2-school (with a score of 5.50 and 5.47, respectively). One was particularly good on the Productivity dimension (6.00). The rules in her classroom were clear and visualized. Furthermore, the transitions during the day were smooth and short. The pupils and the teacher were engaged and there was a great variety of activities to choose from.

The teacher with the lowest score (3.80) on Organizational Support was from a Mainstream School. The scores of this teacher were rated as particularly low on the dimension Productivity. On the dimension Instructional Learning Formats she did not score consistently, with three times a score of 5 and two times a score of 2. The observations with a score of 2 were at the start of the day when the pupils entered the class and sat waiting in a circle; the other one was during eating and drinking. The teacher did not maximize the learning opportunities at these points, she could have provided books, songs, or maybe an assignment to think or talk with a neighbor about a specific topic they would be discussing later.

The Mainstream school teacher who scored highest on the domain Emotional Support also scored highest on the domain Instructional Support (4.00), which is relatively high. Especially her score on Language Modeling was much higher than those of the other

teachers. The difference between her and the other teachers was that she repeated answers and she extended words and concepts more frequently.

The teachers who had the lowest scores on the domain Instructional Support were two teachers at the same DL2-school, with mean scores of 2.17. They scored particularly low on the dimension Concept Development. They did not ask many ‘how’ or ‘why’ questions and in conversations there was little room for making connections with the real world, for problem solving, or for brainstorming on ideas. The quality of feedback could be improved if they would give feedback on the learning or thinking process instead of on the results. Also the persistence of the teacher in given feedback and having a reciprocally exchange with the pupils will improve the quality of the feedback. The language development of the pupils could be stimulated if there would be more back and forth conversation and expansion on pupils’ utterances.

## 6.5 Summary and conclusion

The purpose of this chapter was to investigate what characterizes the school learning environment for newly arrived migrant kindergarteners at the classroom level, especially concerning differences between DL2-schools and Mainstream schools. To study this, the Classroom Assessment Scoring System (CLASS) was used to measure the quality of several dimensions of teacher-child interactions in the different classrooms. The research question which was central to this chapter was: what are the differences in characteristics of the school learning environment regarding teacher behavior between DL2-schools and Mainstream schools? We found 2 out of 10 possible significant effects. Table 6.3 gives an overview of the results.

Table 6.3: Overview of Significant Pairwise Contrasts for Scores on the CLASS.

<b>Emotional support</b>	
Positive Climate	Higher in Mainstream schools
Negative Climate	
Teacher Sensitivity	
Regard for Student Perspectives	Higher in Mainstream schools
<b>Organizational Support</b>	
Behavior Management	
Productivity	
Instructional Learning Formats	
<b>Instructional support</b>	
Concept Development	
Quality of Feedback	
Language Modeling	

The first conclusion of this chapter is about the overall pattern of teacher behavior in the kindergartens in the present study. Overall, the teachers in this study showed a comparable pattern in their scores over the ten dimensions which were found in other studies (cf. Mashburn, Justice, Downer, & Pianta, 2009; Pakarinen et al., 2010; and the Dutch studies: Leseman & Veen, 2016; Henrichs & Leseman, 2016; Veen et al., 2017). The highest score on the CLASS was found on Negative Climate, meaning that there was very little negativity in the classrooms in this study. Buell, Han, and Vukelich state that “in general, classrooms score highest in the Emotional Support domain, and lowest in the Instructional Support domain” (Buell et al., 2017, p. 1636). We found this similar pattern in our data, highest scores on Emotional Support and lowest score for instructional support, with no significant difference in the pattern between the two school types.

Even though the general pattern of the observed quality mirrors that of previous studies, the teachers in our study seem to have overall lower CLASS scores than in international (Mashburn et al., 2009; Pakarinen et al., 2010; Osborn, 2012) and national studies (Henrichs & Leseman, 2016; Veen et al., 2017; Slot et al., 2017). The difference between our study and previous studies was the fact that we looked at classrooms with newly arrived migrant pupils, and thus beginner learners of a second language. A possible explanation for the lower process quality in our study could be that teachers have different, maybe lower, expectations for newly arrived migrant kindergarteners compared to monolingual pupils and therefore that they might set lower goals, ask simpler questions, and want to have more control on the process as a teacher instead of giving pupils the opportunity to show initiatives.

To illustrate this, a necessary precondition for the use of rich language by the teachers, and the occurrence of, for example, back and forth conversations (indicators for Language Modeling), is the expectation that the teachers will be understood by the pupils in class and can elicit output from them. It might be so that when teachers do not expect to be understood by the pupils or when they expect that pupils do not have enough Dutch vocabulary to respond, they simplify their speech, using shorter sentences and more frequent words or demand other replies and actions from the pupils. To score higher on process quality teachers could instead have used language that is rich, with a lot of scaffolding around it to make the language also understood by pupils with lower proficiency in Dutch.

As can be seen in the boxplot of the observational data in Appendix 4, there was a lot of variation in teachers' scores, also within each school type and even between teachers within one school. It would be recommendable that there is constant professionalization based on observations in the classrooms, time for teachers to discuss the observations, and time for teachers to visit colleagues who can show good examples. A tool such as the CLASS can function as a guide line with all the indicators of good practices for each dimension. For example, a school or a teacher can choose one dimension to improve during

a semester and specific observations can be made and feedback can be giving regarding that particular goal.

The second conclusion of our analysis of the characteristics of the school learning environment for newly arrived kindergarteners – looking at the separate scores on the ten dimensions of CLASS – was that Mainstream school teachers scored on average higher on the dimensions Positive Climate and Regard for Student Perspectives than teachers at DL2-schools. Although it has to be said that, even though there was a significant difference between DL2-school teachers and Mainstream school teachers in the dimension of Positive Climate, both scored in the mid-range of the score. The differences that we saw between the two school types mean that we saw more evidence of positive relations between teachers and pupils and among pupils at Mainstream schools. Enthusiasm, affection, and respect to each other were more visible in those classes. Furthermore, in Mainstream schools teachers followed the initiative of pupils more often compared to DL2-schools.

An explanation of this small difference in Positive Climate and Regard for Student Perspectives could be that the classroom organization at DL2-schools tends to be more rigid than at Mainstream schools, based on my own experiences as a teacher. The organization at DL2-schools seems to be more teacher-centered, with more circle time activities and less play, which might have caused more inflexibility of the teachers and less enthusiasm of the pupils. If circle time is more teacher-centered, the teacher might be less flexible and less likely to include the ideas and comments of the pupils; likewise, pupils might have less say in which activities are chosen. A more detailed investigation of the classroom activities used in each type of school will provide a better understanding of how Positive Climate differs between the two (see Chapter 7).

One outcome that should be mentioned is the fact that the score on Positive Climate varied more between DL2-teachers than between Mainstream school teachers. On the one hand this larger variation is not remarkable since there were almost twice as many teachers at DL2-schools in this study compared to the Mainstream schools. On the other hand, there were as many participating DL2-schools as Mainstream schools, and it can be expected that teachers from similar schools share certain practices which would result in less deviation in scores. Further research is needed in order to explain why teachers at DL2-schools varied more in their scores on Positive Climate than Mainstream school teachers.

Taking into account that there is a relationship between classroom quality and language development, all educational facilities would benefit from an investigation into process quality. Based on Hamre et al. (2012) and Henrichs and Leseman (2016) process quality in classrooms can be improved when schools invest in the professional development of the teachers. When teachers receive concrete feedback with tips on their performance, improvements can be made. The course that Hamre et al. (2012) facilitated proved that teachers demonstrated more effective emotional and instructional interactions after participation in their course. Also the results of the professionalization of the teachers

executed within the Henrichs and Leseman (2016) study showed that the scores in the instructional support domain can increase as a result of professionalization and direct feedback after observations.

Based on previous studies we could not predict any differences between DL2-schools and Mainstream schools because these two types had never explicitly been compared. Nevertheless, since the teachers at DL2-schools have a focus on learning the Dutch language, we might have expected differences in, for example, the scores on Language. However, this was not found in our study. An explanation for this could be that both teachers at DL2-schools and Mainstream schools use similar tools for the stimulation of language development of kindergarteners, since all kindergarteners are in the process of acquiring the language. Nonetheless, for both types of schools it is recommended to train the teachers to improve their behavior with respect to Language Modeling, that is, on the whole of the instructional support domain.

The two significant variables from this Chapter 6 (the dimension Positive Climate and the dimension Regard for Student Perspectives) will be used in Chapter 8 in order to look at the effects of learning environment characteristics on the development of the different second language measures from Chapter 4 and 5. But first, Chapter 7 will continue with the investigation of the school learning environment, focusing on the focal pupils' experiences instead of teacher behavior.



## CHAPTER 7

---

**Pedagogical practices: focus on focal pupils' experiences**

---

## 7.1 Introduction

In Chapter 6 the pedagogical practices in (1) DL2-schools and (2) Mainstream schools were observed from the point of view of teacher behavior, focusing on whole class teacher-pupil interactions. Whereas teacher-pupil interactions on the general level in the classroom are an important indication of the school learning environment, the specific focal pupils' experiences might reveal important additional information. Thus in order to get a complete picture of the school learning environment it is important to additionally look at the interactions of our focal pupils with their teacher and their peers in the classroom. Commonly used in studies observing classroom activities is a "snapshot procedure": taking observations at regular intervals of what is going on around a pupil and then coding this. In both Chapter 6 and 7 the school learning environment of newly arrived migrant kindergarteners is investigated; however, while in Chapter 6 the focus was on teacher behavior, the present chapter will focus on specifically the experiences of the newly arrived migrant kindergarteners in the classroom, our focal pupils.

The central question of Chapter 7 is therefore:

What are the differences in characteristics of the school learning environment from the point of view of focal pupils' experiences between DL2-schools and Mainstream schools?<sup>18</sup>

The present chapter is structured as follows: before discussing the data collection, a brief review is given about what is already known about the influence of focal pupil's experiences in education on child-outcomes, specifically focusing on research with a snapshot observation instrument (Section 7.2). Then the methodology (Section 7.3) and results (Section 7.4) of data collected from 42 participants from ten different schools is discussed. The chapter ends (Section 7.5) with a summary and discussion, which will be extended in Chapter 9.

## 7.2 Pedagogical practices: focus on focal pupils' experiences

There are different approaches to measure and observe the quality of education by looking at interactions in a classroom. For example, Peisner-Feinberg et al. (2014) reviewed ten studies on quality in early childcare and half of these studies used some kind of snapshot procedure. However, there are multiple ways to make use of snapshots in the classroom. For example, a structured observation protocol initiated every 15 minutes (Jacoby & Lesaux, 2017), or every 30 minutes (Stegelin, Anderson, Kemper, Wagner, & Evans, 2014) in which anecdotal notes were taken and after which quantitative time sampling measures were recorded in an appropriate category. Also, analysis of whole morning video

---

<sup>18</sup> This research question was written in Chapter 3 as sub-question (d).

recordings is used (Yeager Petalli, Piasta, Justice, & O'Connell, 2014; McDonald Connor et al., 2009).

All in all, snapshot methods are used to assess more objectively and concretely, at a low-inference level, what happens in the classroom, from the point of view of several selected focus children. A snapshot method requires representative sampling and preferably a large number of observation intervals for reliability of the measures. Snapshot measures are usually reported as proportions of intervals in which particular events occurred; the more fine-grained the intervals are, the more accurately the method estimates how time is actually spent in the classroom by the specific focal pupil.

### **7.2.1 Previous studies using snapshot**

Many studies make use of the Emergent Snapshot procedure (Ritchie, Howes, Kraft-Sayre, & Weiser, 2001; Early et al., 2010), taking notes every 10 to 20 seconds, but these are difficult to compare according to Peisner-Feinberg et al. (2014) since each study uses different aspects or versions of that procedure. For example, Zuniga and Howes (2009) used the snapshot procedure to measure provider scaffolding of children's learning in a family child care setting, the responsive involvement, and children's engagement in pre-academic activities. Howes, Shivers, and Ritchie (2004) used a snapshot measure to measure classroom emotional climate, teacher responsive involvement, and social competence with peers. Another example is Chang et al. (2007) who focused with their snapshot procedure on the teacher-pupil interactions, with codes on a continuum varying from simple routine teacher-pupil interactions, to elaborate teacher-pupil interactions. Thus, even though studies use a snapshot method, it is more the technical procedure that is comparable instead of the outcomes. This is in contrast with the instrument used in Chapter 6. The CLASS is an instrument that when it is used the outcomes in different studies can be compared since the CLASS always measures process quality and not something else.

### **7.2.2 Dutch studies using snapshot**

Two studies carried out in the Netherlands using the snapshot procedure are quite similar to the present study. Henrichs and Leseman (2016) also investigated the activities and interactions in Dutch kindergartens (N = 185 participating pupils, of which 94 kindergarteners). The activities occurring most frequently during the observed, fine-grained, 10-second intervals in their study were transition (changing activities, waiting, cleaning, wandering around), conversations, language activities, mathematical activities, and play – all together about two-third of the day's activities. They further analyzed some of the activities. One activity is coded "conversation," meaning talking with each other about more than merely the activity. Only 10% of the time these conversations were among pupils themselves. When a teacher had a conversation with the pupil it was mainly didactic: a sequence of short initiative response sequences. The language and mathematical activities

were mainly teacher led (more than 80%). Teachers used scaffolding half of the time in interactions with pupils during language and mathematical activities. For play the pupils were alone one third of the time, and with their peers half of the time. Around one fifth of the time they interacted with their teacher during play. During these interactions the teacher used scaffolding around 10% of the time.

Another Dutch study about activities in kindergartens (De Haan, Elbers, & Leseman, 2014), with 91 pupils aged 3 to 6, reported that teacher-managed language and literacy activities took place around 15% of the time, which was the highest proportion for pre-academic activities (language, math, or literacy activities). No activity, waiting, or mere transitions took a large proportion according to their observations.

The current study used the same snapshot procedure as Henrichs and Leseman (2016), but with adapted codes to focus on different aspects of activities, interactions and language use. The codes were created to obtain a more complete picture of the school learning environment of the newly arrived migrant pupils in the class, our focal pupils. The focus was on the interactions between the focal pupil and the teacher or their peers and on the use of languages in the classroom because with this information the language input can be analyzed. It was expected that these would be important explanatory factors in the second language development of the kindergarteners.

## **7.3 Method**

### **7.3.1 Participants**

Forty-two pupils (21 girls) participated in this part of the study (see Section 3.4 in Chapter 3 for detailed description of the participants). The 42 pupils were enrolled at 10 different schools; the same schools of which the teachers participated in the observations with CLASS (see Chapter 6). Table 7.1 provides the descriptive characteristics of the participants in this study. Age is the average age over the two observation days within this study.

Table 7.1: Descriptive Characteristics of the Participants (N = 42) in the Snapshot, per School Type (N = 10).

	Number of Classes	Number of Pupils	Age (in months)		
			Range		Mean ( <i>sd.</i> )
			Min	Max	
<b>Total DL2-schools (N = 5)</b>		<b>32</b>	<b>51</b>	<b>81</b>	<b>65 (8.4)</b>
School 1	2	3			
School 2	4	7			
School 3	1	3			
School 4	3	6			
School 5	3	13			
<b>Total Mainstream schools (N = 5)</b>		<b>10</b>	<b>51</b>	<b>82</b>	<b>66 (8.3)</b>
School 6	3	7			
School 7	1	1			
School 8	1	1			
School 9 & 10 <sup>b</sup>	1	1			

<sup>b</sup>This participant changed schools.

### 7.3.2 Measurement

Based on the Emerging Academics' Snapshot method (in short: Snapshot; Ritchie et al., 2001; Early et al., 2010) and the adjusted protocol of this method in Henrichs and Leseman (2016), a snapshot protocol was designed using the E-Prime 2.0 software (Schneider, Eschman, & Zuccolotto, 2002). With this procedure we aimed at capturing in detail the activities and interactions of the specific focal pupils, based on a series of snapshots. It was a cyclic-interval-coding approach which was used to observe the focal pupils. With this a moment-by-moment observation could be executed to describe pupils' experiences within their program.

The outcome of the snapshot procedure is an estimate of the percentage of time a specific pupil spends on certain activities in the classroom. This estimate is increasingly accurate the more fine-grained the observation intervals are. For example, the approach that was developed provides estimates of the percentages of how much time a pupil is working on language activities or in what percent of the cases of teacher-pupil communication there was scaffolding. A researcher can adjust the snapshot method by focusing on specific focal pupils' experiences depending on the focus of the research. Therefore, in the present study, in addition to the standard questions about activities, questions were included about the interactions and the language use by the teacher, the focal pupils, and the peers.

### **7.3.3 Procedure**

With the snapshot method, two observation sessions were performed during the first year that the pupils were in the Netherlands. The observation of the school learning environment from the point of view of the focal pupils' experiences lasted an entire morning, usually from 8:30 to 12:00. In between the two observation sessions there were approximately 3 months. To improve the reliability, the data from the two observations days were combined after confirming that the patterns were stable across the two days.

Per class it differed how many focal pupils were observed. When there was only one participating newly arrived migrant pupil in the classroom only that child was observed. When there were multiple focal pupils in the classroom these were observed during the same morning with a maximum of 6 pupils per observation. The data collection consisted of multiple observation cycles. Observation cycle 1 started by observing the first focal pupil in the classroom for a 10-second observation period (cf. Henrichs & Leseman, 2016), followed by a coding procedure on a laptop. Then, if applicable, the data collector observed each of the other focal pupils in succession. After observing and coding all pupils in the first cycle, the observer started over with the second cycle with the first focal pupil, continuing in this matter for an entire morning.

On average, each pupil was observed and coded 72 times over the two mornings (range = 30–131, see Table 7.2). A total of 2570 observations were coded. The number of pupils who were observed during the same morning in the same classroom varied from 1 focal pupil to a maximum of 6. Therefore, there is a difference of 101 observations between the minimum and maximum number of observations. More observations could be obtained from the pupil who was the only one being observed that morning compared to the pupil who was observed with 5 other peers.

Table 7.2: Number of Observations for the Snapshot Method, per School Type.

	Number of Observations <sup>a</sup>			
	Total	Per Focal Pupil		Mean
		Range		
	Min	Max		
<b>DL2-schools</b>	<b>1779</b>	<b>30</b>	<b>109</b>	<b>64</b>
School 1	191			
School 2	593			
School 3	154			
School 4	345			
School 5	496			
<b>Mainstream schools</b>	<b>791</b>	<b>47</b>	<b>131</b>	<b>92</b>
School 6	434			
School 7	97			
School 8	131			
School 9 & 10 <sup>b</sup>	129			

<sup>a</sup> The number of observations is the total number of observations over the two days, except for one boy at a DL2-school that only took part in one observation day.

<sup>b</sup> This participant changed schools

The snapshot method consists of a list of coding categories with concrete indicators that must be checked as either present or not present during a 10-s period. This 10-s period differs from the 20-s period of Ritchie et al. (2001) but follows Henrichs and Leseman (2016). These indicators are formulated as questions like “in what kind of activities are the pupils engaged?” Sometimes a question will be excluded, and skipped by the E-Prime program, due to the answer to the previous question; therefore the degrees of freedom will differ per category in the results section. Sometimes a category was skipped accidentally, for example, when the enter button was tapped too quickly, which resulted in truly missing data.

The coding categories in the present study were adapted from that of Ritchie et al. (2001), Early et al. (2010), and Henrichs and Leseman (2016) because in the present study we wanted to know more about the interactions of the focal pupils. In previous studies the focus was mainly on pupil-teacher interaction, but in this study, we also included the focal pupil-peer interactions, because pupils are engaged with each other a considerable amount of time in a classroom setting. Furthermore, because we know that language development depends on language input we analyzed that aspect in more detail than in previous studies.

A total of six indicators will be discussed in our study, divided into three categories: Activities, Interactions, and Language Use (see Table 7.3). The different categories will now be presented in short; we would like to refer to Appendix 5 for more details on each indicator.

Table 7.3: The three Observation Categories and their six Indicators.

**Category 1: Activities**


---

 In what kind of activities are the focal pupils engaged?
 

---

**Category 2: Interactions**


---

 What type of interaction does the pupil have?

---

 What kind of teacher interaction is taking place?
 

---

**Category 3: Language use**


---

 What kind of language situation is there?

---

 What kind of language is it?

---

 What type of language is used by the focal pupil, the peer with whom the focal pupil is interacting, and the teacher?
 

---

**Activities**

Pupils in a kindergarten class participate in many different activities per day. Therefore, the first question to be answered is: in what kinds of activities are the focal pupils engaged? The Activities are clustered into nine types: (1) Language Activities, (2) Literacy Activities, (3) Mathematical Activities, (4) Play, (5) Expressive Activities, (6) Physical Activities, (7) Routines, (8) Waiting, and (9) Other. When the activity could fall into more than one category, the most cognitive category is chosen, which is called positive coding (cf. Early et al., 2010).

**Interactions**

The first question in the category interaction is: what types of interactions do the pupils have? This could be (1) Teacher-Focal Pupil Interaction, (2) Peer Interaction, (3) Teacher Instruction, or (4) No Interaction or interaction with a parent. Only in the case of teacher interaction is this interaction is further analyzed. The second question about interaction was therefore: what kind of teacher interactions are taking place? This teacher interaction could be coded as Scaffolding or Didactic. Scaffolding involves a high-quality interaction in which the level at which the pupil is achieving is actively raised to a higher level by the teacher.

**Language Use**

The first question in the category language use is: what kinds of language situations are there? It is scored whether there is language around the pupil during the observation or not. Does the pupil hear language from the teacher involved in a group activity or in a one-on-one situation, or for example does the pupil hear music or sound from a television? The choice could be:

(1) Dialogue, (2) Language Situation with Peers, (3) Teacher Talk, (4) Balanced Language Situation, (5) Self-talk, (6) No language, or (7) Other.

The second question in the category language use is: what kinds of languages are there? The possible answers to this question were: (1) Dutch, (2) Language Other than Dutch, (3) Non-Verbal communication, or (4) Silence.

Finally, the third question in this category is: what types of language are used by the focal pupils, the peers with whom the focal pupils are interacting, and the teachers? (1) Complex language, (2) Simple language, (3) Language Other than Dutch, (4) Non-verbal, (5) No language.

#### **7.3.4 Analysis**

##### **Statistical analysis**

The analysis of the data collected using Snapshot is twofold. First, descriptive statistics are given about the amount of time spent on specific activities, language situations, and interactional situations. Second, using Generalized Linear Mixed models (using SPSS 22; IBM Corp., 2013), fixed effects and fixed coefficients are reported for the effect of school type on the specific variables to answer the question: are the differences in amount of time for a specific variable significant between DL2-schools and Mainstream schools?

##### **Rater reliability**

There was one observer (the author) for all the observations collected with the snapshot method in this study. The observer was trained by Lotte Henrichs, a researcher who has worked intensively with the snapshot method. It would have enhanced the reliability of the data if a second rater had been present, or watched recordings of the observations for at least 10% of the data collection so that inter-rater reliability could have been calculated, but this was not possible, due to, among others, privacy regulations.

### **7.4 Results**

#### **7.4.1 Quantitative analysis of the snapshot data**

The results from the observations across three different categories (activities, interactions, and language use) are presented in this section. First, the different activities that took place in the classroom are reported and analyzed. Then, the kinds of interaction that took place in class are discussed. Finally, the language heard and used in class is reported in detail.

##### **Activities**

The first question to be answered was: in what kind of activities are the focal pupils engaged? Table 7.4 shows the estimated average proportion of intervals in which particular activities are observed, regarded here as a close approximation of the actual time spent on these activities. The time spent on the different activities varied from .01 to .23, averaged over two full observation days. Most frequent were Language Activities (like conversations, instructions, vocabulary lessons, social emotional lessons, and watching educational television). A considerable amount of time was also spent on routines without

clear educational intention (eating and drinking, cleaning-up), on average about one fifth of a day.

Table 7.4: (Estimated) Means of Percentage of Time (Standard Error) Spent on the Nine Activity Categories, per School Type.

	DL2-school	Mainstream school
Language	.23	.18
Literacy	.10	.10
Mathematical	.09	.05
Play	.11	.16
Expression	.06	.09
Physical	.09	.12
Routines	.20	.20
Waiting	.07	.07
Other	.03	.04

Generalized Linear Mixed Models were used to test the statistical significance of the observed differences, based on 42 pupils at the two school types who were individually followed on the two days. Significant effects were found for Mathematical Activities  $F(2,2558) = 6.59$ ,  $p = .01$  and for Language Activities  $F(2,2558) = 5.05$ ,  $p = .03$ . Pupils engaged significantly more often in Mathematical Activities and Language Activities at DL2-schools compared to pupils at Mainstream schools.

### Interaction

In addition to activity, the question “what types of interactions do the focal pupils have?” is answered. Overall, the pupils were for a considerable part of the time not interacting—either because they were working independently or wandering around without interaction—on average for about one third of the time. Another third of the time, they were interacting with their peers, see Table 7.5.

Table 7.5: (Estimate) Mean (Standard Error) of Proportion of Time for Interaction, per School Type.

	DL2-schools		Mainstream schools	
	Mean	SE	Mean	SE
Teacher-Focal Pupil	.21	.02	.15	.02
Peer Interaction	.24	.01	.40	.03
Teacher gives instruction	.19	.01	.16	.02
No Interaction or with parent	.35	.02	.29	.03

The effect of school type was significant for Peer Interaction:  $F(1,2560) = 32.91, p < .001$ . The pairwise contrast showed that pupils at Mainstream schools were more engaged in Peer Interactions than at DL2-schools. The effect of school type was also significant for Teacher-Focal Pupil Interaction  $F(1,2560) = 4.83, p = .03$ . At DL2-schools, focal pupils were more engaged in interactions with only the teacher compared to pupils at Mainstream schools.

When for the previous indicator of interaction “Teacher-focal Pupil interaction” of “Teacher gives instruction” was answered a subsequent question was used: “what kinds of teacher interactions are taking place?” Table 7.6 gives the estimated mean percentages. By far the most teacher interaction was “Didactic.” “Scaffolding” occurred only 5% of the time during teacher interaction in both school types. “No interaction” seems out of place here, but it is scored when a pupil was in interaction with the teacher, but at the exact 10 seconds of observation the teacher was distracted and not focused on the pupil. Additional analysis of the data about the teacher interactions did not reveal significant differences between the types of teacher interaction in the different educational settings.

Table 7.6: (Estimate) Mean (Standard Error) of Percentage of Time for Teacher Interaction, per School Type.

	DL2-schools		Mainstream schools	
	Mean	SE	Mean	SE
Scaffolding	.05	.01	.05	.01
Didactic	.90	.01	.88	.02
No interaction	.07	.01	.08	.02

### Language Use

As the final category the amount and type of language use that involved the individually observed pupils was coded. For language use, first the language situation itself was coded. The question that was answered is: what kinds of language situations are there? The results are presented in Table 7.7. Overall, one fifth of the day there was no language directed at, actively listened to by, or produced by the focal pupils. This does not mean that it was completely silent in the classroom, but it indicates that the pupils did not always pay attention to teachers' and peers' language use. Another fifth of the time teachers and pupils were engaged in a more balanced way in dialogues in a big group or whole class situation, usually when sitting in a circle. If the focal pupils were engaged in verbal interactions, this was most of the time with a peer (on average for the two school types respectively 26% and 42% of the time). One-to-one and small group dialogues between teachers and pupils were less frequent.

Table 7.7: (Estimated) Means of Proportion of Time (Standard Error) for Language Situation, per School Type.

	DL2-schools		Mainstream schools	
	Mean	SE	Mean	SE
A dialogue between teacher and focal pupil	.12	.01	.09	.01
Language situation with peers	.26	.01	.42	.03
Teacher talks	.10	.01	.10	.02
A balanced situation between teacher and pupils	.21	.02	.14	.02
Self-talk of the focal pupil	.06	.01	.05	.01
No language	.21	.02	.19	.03
Other	.04	.01	.04	.01

There were two significant difference between school types for languages situations: for Language Situations with Peers:  $F(1,2286) = 33.40$ ,  $p < .001$ , and for balanced situations between teacher and all pupils:  $F(1,2286) = 6.70$ ,  $p = .01$ . Pupils were more often engaged in language situations among only peers at Mainstream schools compared to DL2-schools. On the other hand, at DL2-schools pupils were more engaged in language situations where there is a balance between the input of the teacher and the pupils compared to pupils at Mainstream schools.

The analysis of language use continued with the question: what kind of language is it? The language that was observed in class was more than three quarters of the time Dutch (see Table 7.8).

Table 7.8: (Estimated) Means of Proportion of Time (Standard Error) for Type of Language, per School Type.

	DL2-schools		Mainstream schools	
	Mean	SE	Mean	SE
Dutch	.81	.03	.77	.05
Other than Dutch	.02	.01	.08	.03
Non-verbal language	.10	.01	.07	.02
Silence	.07	.01	.05	.01

Analysis showed that the only significant difference between the school types was the amount of time spoken in another language than Dutch  $F(1,2027) = 7.07$ ,  $p = .01$ . Pupils at Mainstream schools speak significantly more often in another language than Dutch compared to pupils at DL2-schools. However, pairwise comparison showed that the difference between both groups is very small ( $F(1,2027) = 3.04$ ,  $p = .08$ ). Furthermore, the relative high mean for Mainstream schools is due to only one of the five Mainstream schools. For both reasons we refrain from interpreting this effect any further.

Finally, the language that was observed was analyzed in more detail. The question that was answered was: what type of language is used by the focal pupil, the peer with whom the focal pupil is interacting, and the teacher?' Table 7.9 and Table 7.10 present in detail the kind of language focal pupils, peers, and teachers use in the different school types. Overall, most of the time the nature of the language use was qualified as simple language. Simple language means the use of short sentences – mostly one-word sentences. Furthermore, in simple language there would be no explicit reference and the interlocutors would often use words like “this,” “that,” and “like this.” Diminutives were also used often in simple language as well as mostly high frequent words. Complex Language Use, that is, the use of longer sentences, precise reference, and infrequent words with explanation, was hardly observed. Remarkably, the Language Use of the teachers was overwhelmingly coded as simple at both school types.

Table 7.9: (Estimated) Mean of Percentages of Time (Standard Error) per Type of Language Spoken by Focal Pupils or by Peers while Interacting with the Focal Pupils.

	Focal Pupils							
	DL2-schools		Mainstream schools		Peer DL2-schools		Mainstream schools	
	Mean	SE	Mean	SE	Mean	SE	Mean	SE
Complex	.03	.00	.03	.01	.03	.01	.03	.01
Simple	.37	.02	.35	.04	.56	.03	.45	.05
Other than Dutch	.01	.00	.06	.03	.01	.01	.07	.04
Non-verbal	.27	.02	.26	.03	.24	.02	.23	.02
No language, the pupil is silent	.33	.02	.27	.04	.17	.02	.17	.03

Focal pupils were on average often silent (about one third of the time). The focal pupils' peers were on average less often silent than the focal pupils, and especially teachers were almost never silent when interacting with the focal pupil. The use of a language other than Dutch by the focal pupils or their peers was overall rare, but seemed to occur more often at Mainstream school (at one school in particular). The use of non-verbal forms of communication (e.g., gestures) by focal pupils and peers occurred more frequently than the use of other languages than Dutch.

The differences in language use between the types of schools were tested for statistical significance. For both the focal pupils and their peers, the use of a language other than Dutch was significantly different between the school types, respectively  $F(1,2041) = 8.08$   $p = .005$  and  $F(1, 1433) = 6.41$ ,  $p = .01$ . However, pairwise comparison showed that the difference between both groups is very small. Furthermore, the relative high mean for Mainstream schools is due to only one of the schools. For both reasons we refrain from interpreting this effect any further. Of the 9 schools, at 2 schools, languages other than Dutch were never used; in six schools it was used, but with a maximum of 5%. There was only one school, a Mainstream school, where a significantly larger amount of time other languages than Dutch were spoken in the classroom, with a percentage of almost 18 and 24 for respectively the focal pupils and their peers. The languages that were observed are mainly Arabic, Chinese (as indicated by parents on the confirmation form), and Polish.

Teachers at both school types used simple language most of the time. Regarding the Language Use of the teachers, there was no significant difference between teachers at DL2-schools and teachers at Mainstream schools.

Table 7.10: (Estimated) Mean of Percentages of Time (Standard Error) per Type of Language Spoken by Teachers while Interacting with the Focal Pupils.

	Teacher			
	DL2-schools		Mainstream schools	
	Mean	SE	Mean	SE
Complex	.02	.01	.05	.02
Simple	.91	.01	.88	.03
Non-verbal	.06	.01	.06	.02
No language, the teacher is silent	.03	.01	.03	.01
Missing	.03	.01	.03	.01

#### 7.4.2 Additional field notes

In addition to the observations obtained using the Snapshot, data about the characteristics of the schools was obtained via field notes based on observations and a semi-structured

interview with the teachers of the focal pupils in the ten different schools. Specifically additional information about the pedagogical approaches, strategies, and teaching materials proved valuable for the qualitative comparison. At the national level, there are no guidelines indicating which methods and materials schools should use in the Netherlands. Therefore, teachers in both school types designed and developed their own curriculum based on existing textbook materials complemented by recent didactics and materials. In total, the teachers of the 42 focal pupils—in 10 schools—named eight different methods on which they based their activities.

However, none of the teachers claimed to follow a specific method strictly. Like in most Dutch kindergarten classes, the daily activities were organized within one general theme (which would last a few weeks), such as, shopping, spring, traffic, or being sick. The themes were made explicit by means of crafts and pictures displayed on the walls in both schoolings.

Specifically, all teachers at DL2-schools noted that they adapted the existing methods for mainstream kindergarteners to make them more suitable for newly arrived migrants and complemented them with extra word cards, games, craft activities, worksheets, and digital presentations. The number of new words presented to the pupils varied from lesson to lesson and from school to school (with a minimum of five words per lesson); although one DL2-school made clear that each day twenty new words should be taught. In all DL2-schools word clusters with pictures were visible on the walls, which were not observed in all Mainstream schools.

Two Mainstream schools also reported that they adapted their methods. They claimed that they based their adaptation on goals from the observational methods<sup>19</sup> they use, not specifically designed for the focal group. All Mainstream schools also work with explicit vocabulary instruction, but there are considerable differences between the schools. One school sometimes only teaches two words in a vocabulary lesson. During the Snapshot observation such a vocabulary lesson took place and it was not obvious why these two specific words were chosen, nor did it seem that they were part of a general theme—the words did not reoccur during other lessons of that morning. At only two Mainstream schools, word clusters were made visible in the classroom with pictures and written words.

During the snapshot observations none of the teachers spoke a language other than Dutch. It was noted that the teachers did comment on the use of other languages by the pupils, mostly in a negative way. One of the teachers, for example, told the observer that she must regularly remind the Polish pupils in her Mainstream school “you are at school now,” meaning, at school they must speak Dutch, not Polish. At the time of the observations, deliberately including home languages in the daily activities was not part of the official school policy of any of the schools in the study. In at least two schools of the

---

<sup>19</sup> *KIJK! Groep 1 & 2*, Bazalt, Vlissingen, <https://www.bazalt.nl/expertise-kijk>.

around 35 schools we visited during this study we saw a sign that says: “in this school we speak Dutch.”

Not during the snapshot observations but on another day the observer was present in one DL2-school she heard the teacher say "good morning" in Arabic, but there was no real enthusiastic response to this greeting by the students. The observer later asked the teacher if she uses Arabic regularly and she answered “as a team we have received eight lessons with a short introduction to Arabic, but due to my work load I have not been able to focus on this. I feel that I did not learn or remembered it. I did get the feeling how hard it is to learn such a new total foreign language!” This teacher also said to a pupil (not one of our focal pupils) “say it in Arabic” after which the child translated for a classmate. This teacher furthermore had a folder with some Arabic words translated to Dutch; she used it to say a color in Arabic to help a student pick the right color for her assignment.

One of the teachers made a different remark about the various languages the pupils spoke. She said that even though she can understand and speak English she hardly uses it, because she finds it “unfair” towards pupils with a home language she cannot understand and speak.

It was observed that the pupils spoke languages other than Dutch to each other and it seemed that the focal pupils and their peers with whom they these other language spoke could understand each other. It seemed that the focal pupils knew when and to whom they could speak their home language. An example of this is that at one point a Polish girl was arguing with a Polish boy in Polish and in a split second she turned around and complained about the boy to her teacher in Dutch. This example shows that this child has mastered sociocultural competence.

Besides differences we also found similarities between the two school types. It was especially clear that one of the Mainstream schools used a strict word learning methodology like the one of the DL2-schools. This shows that making a distinction between whether or not second language learners are in a separate class may be less important than investigating the actual practices in the classroom.

## **7.5 Summary and conclusion**

The focus of this chapter was on the school learning environment of newly arrived migrant kindergarteners in the first year of their stay in the Netherlands. The main question to be answered was: What are the differences in characteristics of the school learning environment from the point of view of focal pupils' experiences between DL2-schools and Mainstream schools? Observations of the focal pupils, by means of a snapshot procedure, were made in three categories: Activities, Interactions, and Language Use.

First, concerning the answer to the question “in what kind of activity is the focal pupil engaged?” Most frequent were Language Activities. A considerable amount of time was also spent on routines without clear educational intention, on average about one fifth of

a day. When routines, language, literacy, and mathematic activities are taken together, these activities cover 69% of a day at DL2-schools and 60% of the day at Mainstream schools. This is comparable with the two-third from the Henrichs and Leseman (2016) study for transitions, conversations, and language and mathematical activities.

Our data showed two differences between the DL2-schools and Mainstream schools regarding activities measured with the snapshot method. Specifically, pupils at DL2-schools were engaged in more language and mathematical activities than pupils at Mainstream schools. It seemed that DL2-schools were predominantly focused on academic activities. The overview in Table 7.11 shows the different aspects on which the school types differed from each other.

Table 7.11: Overview of Significant Pairwise Contrasts for Scores on the Snapshot.

<b>Activities</b>	
Mathematical Activities	higher in DL2-schools
Language Activities	higher in DL2-schools
<b>Language Use</b>	
Language Situations with Peers	higher in Mainstream schools
Balanced Language Situation	higher in DL2-schools
<b>Interactions</b>	
Peer Interactions	higher in Mainstream schools
Teacher-Focal Pupil Interaction	higher in DL2-schools

The first question about interactions was: with whom is the interaction taking place? Overall about one third of the time the pupils had no interaction with a peer or a teacher. Another third of the time, they interacted with their peers. The second question about interactions was: what kind of teacher interaction is taking place? Comparable with Henrichs and Leseman (2016), the largest amount of time when a teacher and a pupil had interaction it was considered as “didactic” instead of “scaffolding.”

The snapshot method observations showed differences between DL2-schools and Mainstream schools in the kind of interactions that took place. While at Mainstream schools focal pupils had more interactions with their peers than at DL2-schools, focal pupils at DL2-schools had more interaction with their teachers than at a Mainstream school. The interaction at DL2-schools was more teacher-led; there were many activities in which the pupils sat in a circle around the teacher, while the interaction at Mainstream schools was more pupil-led, with a lot of free play.

The explanation for this could be that the teachers at the DL2-schools follow a more strictly designed curriculum. With more teacher-led and explicit Language Activities, teachers can monitor the input and learning goals for the pupils more closely, which they assume will improve pupils’ vocabulary development. However, during these activities teachers should in fact use high quality interactions, especially conceptually focused

interactions as is suggested by Bonnes Bowne, Yoshikawa, and Snow (2017). Based on the CLASS observations in Chapter 6 it can be asserted that these interactions can be improved in not only DL2-school kindergarten classes, but also in Mainstream schools.

Blum-Kulka and Gorbett (2014) found that early attempts at communication by L2 learners emerged mainly in interaction with their teachers and mostly in highly structured contexts. This is not completely in line with our findings since our participants at a DL2-school had an almost equal proportion of time interaction with their peers as with their teachers. The pupils at Mainstream school even had more than twice as much interaction with their peers than with their teachers. Thus, even though they were emerging second language learners the observed pupils dared to interact with interlocutors other than their teacher. The finding of overall more didactic interaction with teachers was similar to Blum-Kulka and Gorbett (2014).

During the above-mentioned interactions there could be different types of language involved. The first question concerning the language use in class was: what kind of language situation is there? If the focal pupils were engaged in verbal interactions, this was most of the time with a peer (on average for the two school types respectively 26% for pupils at a DL2-school and 42% of the time for pupils at a Mainstream school). Overall, one fifth of the time teachers and pupils were engaged in a more balanced way in dialogues in a big group or whole class situation. Another fifth of the day there was no language directed at, actively listened to, or produced by the focal pupils. One-to-one and small group dialogues between teachers and pupils were considerably less frequent.

There were two significant differences between school types for language situations. Pupils at Mainstream schools were engaged in language situations among only peers significantly more often compared to pupils at DL2-schools. On the other hand, pupils at DL2-schools were significantly more engaged in Balanced Language Situations in which teachers and pupils had an equal amount of input, compared to pupils at Mainstream schools.

Continuing with the investigation of the language use in the classroom the question what sort of language is it? was answered. The language that was observed in class was more than three quarters of the time Dutch, with almost 10% non-verbal language (e.g., pointing and gesturing).

To further analyze the language that was used we asked the next questions: 'what kind of language do the focal pupil, the peer, and the teacher use?' Overall, most of the time the nature of the language use was qualified as Simple Language. Complex Language was hardly observed. There were no significant differences between the two school types in the language use of the focal pupil, the peers, and the teacher. The difference in amount of time speaking or listening to another language than Dutch did not seem to be significant because the data concerning this had a high variability. We additionally like to mention that

other languages than Dutch were only used between peers who knew that other language, which could be considered an indication of sociocultural awareness.

The question to answer next would be: “what is more helpful for learning? Peer-interaction or Teacher-Focal Pupil Interaction?” Based on the analysis of the language, both teachers and peers use simple language when talking to the focal pupils, thus what difference would one type of interaction over the other make? Blum-Kulka and Gorbatt (2014) found that “the facilitating role of peer interaction has an important time constraint: it is not available to the L2 children as long as they have not mastered at least rudimentary modes of communication in the new language and do not have enough confidence to use them” (Blum-Kulka & Gorbatt, 2014, p. 192). The timing of our data collection could have been a confounding factor: The pupils were observed within the first six months of their stay in the school and thus might still be in the “silent” period or in the period in which they play along and only simple language would be sufficient. Unfortunately, by using only a snapshot procedure without recording the conversations it is difficult to reanalyze the specific conversations and compare the simple language used by teachers and used by peers.

With regard to our field notes, there are both indications of differences and similarities between and within the different school types. Even though all DL2-schools use comparable approaches to teach vocabulary, there are differences regarding the number of words they taught per day. Some Mainstream schools use the same approach to teach vocabulary as the DL2-schools. Whereas some Mainstream schools focus explicitly on vocabulary learning, which can be seen on the walls in the classroom with many word clusters decorating the walls, others lack all these methods.

A final remark should be given to the use of languages other than Dutch. The amount of non-Dutch was on average not higher than 8%, although one school had a percentage of 21.8 %. Even though numerous authors described the benefits of maintaining home languages and supporting their development in education (e.g. Baker 2011; Cummins 2001; Skutnabb-Kangas, 2000), the role of the home languages in the different schools was very limited across the two school types. “Inclusive curricula integrate the language dimension comprehensively and go beyond a simple opposition between monolingual and bilingual educational models or mother tongue versus foreign language” (Herzog-Punzenberger, Le Pichon-Vorstman, & Siarova, 2017, p. 9). According to Herzog-Punzenberger et al. multilingual pupils need inclusive curricula, but this was hardly reflected in the classrooms during our observations: only one of the teachers in our study utilized the multilingual repertoire of the pupils, however no teacher made reference to one of the home language for comparisons with Dutch. Furthermore, no positive attention was given by the teachers on the use of the home language by the pupils, for example when the teacher saw that the pupils were engaged in a discussion about a book they were reading. All in all, the teachers seem to be unaware about practices how to include other languages

than Dutch, especially when the teacher herself/himself cannot understand or speak those languages.

With our data we showed that there were differences in the percentage of use of languages other than Dutch: languages other than Dutch seemed to be used more often by pupils at Mainstream schools compared to pupils at DL2-schools. These differences were only on the level of language during Peer Interaction. However, these differences were not significant because there was mainly one school which had an extreme score in this category. At the time of the observations, deliberately including home languages in the daily activities was not part of the official school policy of any of the schools in the study. Additionally, in the field notes it was noted that teachers from different schools, including the school with a high number of pupils using languages other than Dutch, reacted with comments like, “Shh, at school we only speak Dutch” when pupils used their home languages amongst each other. It is recommended in the literature that home languages are used constructively to let pupils show their capabilities before they have mastered Dutch vocabulary fully. With this they would not need to go through a long “silent” period. Furthermore, paying positive attention to the home languages of the pupils has socio-emotional benefits (Cummins, 2007).

We would like to add the following to the section on rater reliability (in Section 7.3.4): The observations with the snapshot procedure were collected by one observer, thus no interrater reliability could have been calculated. We choose to have one trained observer instead of two, since an additional observer would require an extra researcher with a laptop be present in the classroom, or that everything that happened in the classroom would be recorded. An additional researcher in the classroom is not favorable since it is already crowded in kindergarten classrooms and it might have disturbed the class routines. To make video recordings of everything that happened in the classroom was also not possible since this would have required multiple sophisticated recording devices which were not available. Even if they had been available, there is no guarantee that permission would have been obtained from all parents of all children in each class.

This dissertation will now continue with Chapter 8 in which the data about the characteristics of the school learning environment, reported in Chapter 6 and 7, are used to analyze the receptive vocabulary development (from Chapter 4) and narrative ability development (from Chapter 5). The analysis in Chapter 8 will show whether the school learning environment characteristics that differed significantly between the two school types, – the score on the dimensions Positive Climate and Regard for Student perspectives and the domain Emotional Support (from Chapter 6) and the amount of time engaged in Language and Mathematical Activities, the proportion of Language Situations with Peers, Balanced Language Situations, Peer Interaction, and Teacher-Focal Pupil interaction (from Chapter 7) – have had an effect on the receptive vocabulary development or narrative ability development of the young newly arrived migrant pupils in focus.



## CHAPTER 8

---

**The influence of the school learning environment on the second language development of newly arrived migrant kindergarteners**

---

## 8.1 Introduction

So far, the results of the language development of newly arrived migrant kindergarteners (Chapters 4 and 5) and the pedagogical practices in the school learning environment (Chapter 6 and 7) have been interpreted separately. The finding that differences in Age and Exposure to Dutch at School cause variation in the rate of second language development has been shown in Chapter 4 and 5 by introducing these variables into the different models of language development. In Chapter 4 and 5 the inclusion of a variable concerning the organizational structure of the school (being a separate language school or a mainstream school) did not result in significant different scores between the groups. In other words this broad distinction between educational facilities needed more refinement. Therefore the results of Chapter 6 and Chapter 7 will now be included in the developmental models of Chapter 4 and 5.

In Chapters 6 and 7 the pedagogical practices in the school learning environment of the newly arrived migrant pupils in their first year in the Netherlands were investigated. The school learning environment was analyzed using observations of the pedagogical practices from two perspectives: one focusing on teacher behavior and one focusing on the experiences of the focal pupils. To analyze teacher behavior the teacher was observed using the Classroom Assessment Scoring System (CLASS; Pianta, La Paro, & Hamre et al., 2008). From the observations of the teachers with the CLASS, two variables seemed to differ significantly between the two school types: (1) the specific dimensions of Positive Climate and (2) Regard for Student Perspectives and the general domain of Emotional Support. The Mainstream school teachers seemed to score higher on these three aspects compared to DL2-school teachers.

The experiences of the focal pupils were investigated using a Snapshot method which revealed six variables which differed significantly between the two school types: At Mainstream schools pupils were more engaged in (1) Language Situations with Peers and (2) in Peer Interactions, while pupils at DL2-schools were more engaged in (3) Mathematical Activities, (4) Language Activities, (5) Balanced Language Situations, and (6) Teacher-Focal Pupil Interactions.

In this present chapter the results from Chapter 6 and 7 are related to the outcomes in Chapter 4 and 5 in order to see whether differences in variables in the school learning environment also cause variation in second language development. We therefore will answer in Chapter 8 the last research question:

To what extent do *differences in the characteristics of the school learning environment* during the first year after arrival relate to receptive vocabulary development and to narrative ability development of newly arrived migrant kindergarteners during the first two-and-a-half years of schooling in the Netherlands?<sup>20</sup>

In other words: do the differences in the learning environment between the two school types (DL2-schools and Mainstream schools) result in differences in pupils' scores on different language aspects?

The pupils' language outcomes from Chapter 4 and 5 will thus be related to the two teacher behavior aspects from Chapter 6 and the six aspects of focal pupils' experiences from Chapter 7. By doing so the current research looked at the second language development of newly arrived migrant kindergarteners from an ecological and interdisciplinary perspective. Both linguistic and sociolinguistic factors were considered, as well as classroom and second language educational perspectives.

In the next two sections the significant characteristics of the school learning environment found in Chapter 6 and 7 will be added to the general models of development of different aspects of the second language from Chapter 4 and 5. In Section 8.2 the General Developmental Model of receptive vocabulary from Chapter 4 will be expanded with the variables based on the characteristics of the school learning environment. In Section 8.3 the General Developmental Models of narrative ability from Chapter 5 will be expanded with the variables based on the characteristics of the school learning environment. Given the small sample size and the increasing uncertainty of the models the more variables are included, the variables in Section 8.2 and 8.3 were added separately, one-by-one, to the General Development Models and not cumulative. Chapter 8 ends with a summary of the results and a discussion in Section 8.4, which will be extended in Chapter 9.

## **8.2 The influence of the school learning environment on receptive vocabulary development**

In this section the significant variables from the school learning environment, observed in Chapter 6 and 7, will be related to the individual development of the receptive vocabulary (using the Peabody Picture Vocabulary Task, abbreviated as PPVT; Dunn & Dunn, 2005) answering the first part of the research question from this chapter: To what extent do differences in the characteristics of the school learning environment during the first year after arrival relate to *receptive vocabulary development* of newly arrived migrant kindergarteners during the first two-and-a-half years of schooling in the Netherlands?

---

<sup>20</sup> This research question was written in Chapter 3 as sub-question (e).

The characteristics of the school learning environment were identified using observations of the pedagogical practices in the classrooms of the kindergarteners. First the focus will be on the pedagogical practices based on teacher behavior, followed by the pedagogical practices based on the experiences of the focal pupils.

The central question in this chapter concerns the relation between the observations of pedagogical practices in the classroom based on teacher behavior and pupil's language scores. These pedagogical practices will be introduced in the Multi-Level Models to explain differences in learning growth between pupils. Therefore, we build growth models in the same way as reported in Chapter 4. Unfortunately, for some teachers the observations of their behavior were made after the pupils took the tests. Therefore, these observations cannot be used for our analysis of the relation between teacher behavior and language outcomes. Consequently, these observations are not taken into account in the present chapter.

The pedagogical variables were added to the model in two ways: as a main effect and as an interaction effect with Age. The former indicates that the pupils' PPVT scores increase (or decrease) due to differences in scores by their teachers on a pedagogical variable. The latter effect indicates that the influence of a pedagogical variable depends on the age of the pupils.

### **8.2.1 Pedagogical practices: focusing on teacher behavior**

The pedagogical practices at the classroom level were investigated by observing the teacher with the CLASS (Pianta, La Paro, & Hamre, 2008); see Chapters 3 and 6 for the description of the participants, method, procedure, and results. Due to logistical issues, some observations of the teachers took place after the assessment of the first (or first two) receptive vocabulary assessments of pupils with the PPVT. In total, fifty PPVT scores mainly from Session 0 were taken out of the analyses. Most of these language measures were taken out of the analysis because they were collected "too early" and had to be removed. However, there were also two teachers (each with two focal pupils in the classroom) who could not be observed, one due to illness, the other because of conflicting interests<sup>21</sup>, and thus the collected language data of their pupils could also not be taken into account in the analysis. Thus, in total 118 cases from 38 different pupils could be taken into account (compared to 168 in the first General Development Model of PPVT), and therefore a new developmental model had to be modelled with these 118 cases.

---

<sup>21</sup> This second class was the class of the researcher herself and it seemed inappropriate to make observations of her direct colleague in her own class. The pupils in the classroom which the researcher herself taught twice a week might behave differently while she was in the classroom compared to regular days with only the other teacher. Also it could be more difficult for the researcher to observe in the moment without taking into account the behavior she might have witnessed during the days she teaches these pupils herself.

A development model was build based on the 118 cases of PPVT scores, starting with only an intercept. From the comparison between the consecutive models (see Table 8.1) it is apparent that a model with a fixed linear component – allowing for differences in Age – fit the data better than a model with only an intercept ( $\Delta\chi^2$  (PPVT<sub>b</sub>) = 82.84;  $df = 1$ ;  $p < .001$ ). The variance within individuals depends on the age of participants ( $\Delta\chi^2$  (PPVT<sub>c</sub>) = 19.30;  $df = 1$ ;  $p < .001$ ). However, the between individuals variation did not depend on Age (PPVT<sub>d</sub>). Furthermore, adding Age<sup>2</sup> to the model did not improve the fit (PPVT<sub>e</sub>). Table 6.1 in Appendix 6 shows the parameter estimates. Hence, in the final model (PPVT<sub>c</sub>) a fixed effect of Age, as well as a variance within pupils component which depends on Age needs to be included, and with this model we continued the analysis.

Table 8.1: Fit of Different Polynomials (-2LL) for Changes in PPVT score (118 cases) as well as the Comparison of Consecutive Models.

Model	-2LL	Comparison			
		Models	$\Delta\chi^2$	$\Delta df$	$p$
PPVT <sub>a</sub> : $\beta_{0ij}cons^a$	1019.91				
PPVT <sub>b</sub> : $PPVT_a + \beta_1 Age^1_{ij}$	937.07	PPVT <sub>a</sub> vs PPVT <sub>b</sub>	82.84	1	<.001
PPVT <sub>c</sub> : $PPVT_b + e_{1ij}Age^1_{ij}$	917.77	PPVT <sub>b</sub> vs PPVT <sub>c</sub>	19.30	2	<.001
PPVT <sub>d</sub> : $PPVT_c + u_{10j}Age^1_{ij}$	912.53	PPVT <sub>c</sub> vs PPVT <sub>d</sub>	4.00	2	.07ns
PPVT <sub>e</sub> : $PPVT_c + \beta_2 Age^2_{ij}$	917.43	PPVT <sub>c</sub> <sup>b</sup> vs PPVT <sub>e</sub>	0.43	1	.51ns

<sup>a</sup> In addition to the intercept, variance components for differences within and between individuals are estimated.

<sup>b</sup> This model could not be estimated when  $u_{10j}$  was included, therefore PPVT<sub>e</sub> is compared to model PPVT<sub>c</sub>.

Based on this General Development Model we constructed Figure 8.1, in which both the average development as well as the differences within and between individuals are represented (see Table 6.1 in Appendix 6 for the parameter estimates). The average receptive vocabulary at an age of 72 months was estimated as 68.46. Each month a child grew older, his receptive vocabulary increased by 1.11.

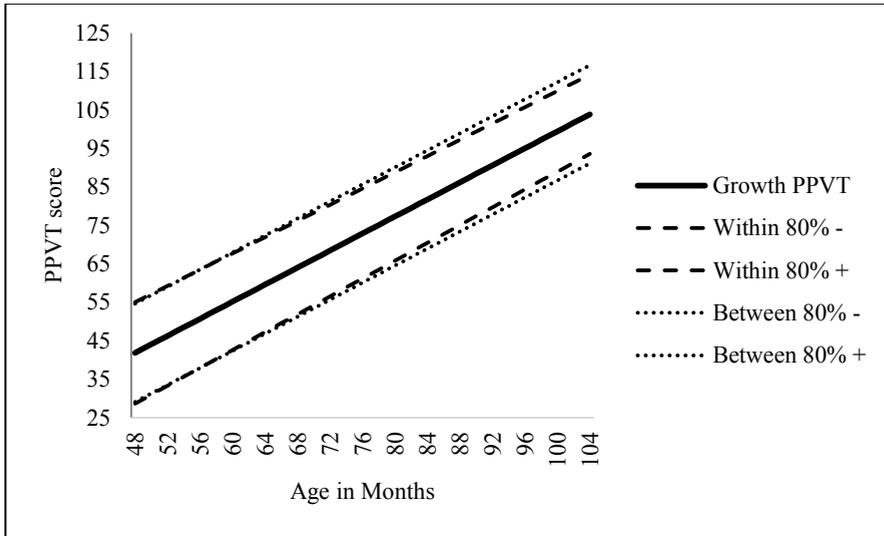


Figure 8.1: Graphical Representation of the General Development Model of PPVT (118 cases) with 80% Reliability.

Figure 8.1 shows that the effect of Age on PPVT scores is significant. The differences within individuals (represented by the striped lines in Figure 8.1) also depend on Age; we can give a more precise estimate of the vocabulary development for older pupils as compared to younger pupils.

The analysis will now continue with the new General Development Model,  $PPVT_c$ . The two significant variables of CLASS namely, Positive Climate and Regard for Student Perspectives, were first added to this new General Development Model. A likelihood ratio test showed that the main effect of Positive Climate did not contribute significantly to the fit of the model of the observed data ( $PPVT_f$ ). However, the model fit increased when the interaction between Age and Positive Climate ( $\Delta\chi^2 (PPVT_g) = 3.86; df = 1; p = .05$ ; See Table 8.2) was included.

Table 8.2: Fit of Different Polynomials (-2LL) for Changes in PPVT score (118 cases) with the Addition of Positive Climate (PosCli) as Explanatory Variable.

Model	-2LL	Comparison		
		Models	$\Delta\chi^2$	$\Delta df$ p
$PPVT_c: \beta_{0ij}cons + \beta_{1i}Age^j_{ij} + \beta_{2i}Age^2_{ij}$	917.77			
$PPVT_f: PPVT_c + \beta_2PosCli_{ij}$	917.61	$PPVT_c$ vs $PPVT_f$	0.15	1 .70ns
$PPVT_g: PPVT_f + \beta_3Age*PosCli_{ij}$	913.76	$PPVT_f$ vs $PPVT_g$	3.86	1 .050

There was an interaction effect of Positive Climate, although the fixed main effect of Positive Climate was not significant. The significant interaction effect between Age and Positive Climate means that the relation between Age and PPVT scores differs per degree of Positive Climate. However, for the interpretation this interaction effect a word of caution is warranted as the standard errors in the last model appeared to be large. Therefore, we are not sure about the specific values for each parameter. Looking at Table 6.1 in Appendix 6 with the estimated parameters we see that when Positive Climate was introduced into the growth model of PPVT this seems to overshadow the effect of Age. Since the fixed main effect of Positive Climate is not significant we can only say that the influence of Positive Climate is larger for older pupils than for younger pupils. We however do refrain from plotting the model in a graph because this would be difficult to interpret. A larger sample is necessary to confirm the effect.

The addition of the variable Regard for Student Perspectives as a fixed main effect did not improve the general model of growth in PPVT. It seems that Regard for Student Perspectives is not a significant predictor of differences in PPVT scores or growth. Therefore, we could not show that there are differences in PPVT scores or growth between pupils in a classroom with a teacher that scored high on Regard for Students Perspectives and pupils in classrooms with a teacher that scored low on Regard for Students Perspectives.

### **8.2.2 Pedagogical practices: focusing on focal pupil experiences**

The pedagogical practices focusing on focal pupils' experiences were obtained by observing the focal pupils with a snapshot method. Due to logistical issues, some observations of the focal pupils took place after the first (or first two) PPVT assessments. The PPVT scores which were obtained before the Snapshot observation were taken out of the analyses. In total 127 cases could be taken into account, and therefore a new developmental model has been modelled with these 127 cases.

From the comparison between the consecutive models (see Table 8.3) it is apparent that a model with a fixed linear component – allowing for differences in Age – fit the data better than a model with only an intercept ( $\Delta\chi^2$  (PPVT<sub>II</sub>) = 85.85;  $df = 1$ ;  $p < .001$ ). However, the variance within individuals did not depend on Age (PPVT<sub>III</sub>), while the variance between individuals was a function of Age ( $\Delta\chi^2$  (PPVT<sub>IV</sub>) = 9.59;  $df = 2$ ;  $p = .01$ ). Finally, adding Age<sup>2</sup> to the model did not improve the fit significantly (PPVT<sub>v</sub>). Hence, in the final model (PPVT<sub>DEF</sub>) a fixed effect of Age, as well as a variance between pupils component which depends on Age needs to be included, and with this model we continued the analysis.

Table 8.3: Fit of Different Polynomials (-2LL) for Changes in PPVT score (127 cases) as well as the Comparison of Consecutive Models.

Model	-2LL	Comparison			
		Models	$\Delta\chi^2$	$\Delta df$	<i>p</i>
PPVT <sub>I</sub> : $\beta_{0ij}cons^a$	1044.28				
PPVT <sub>II</sub> : $PPVT_I + \beta_1 Age^1_{ij}$	958.07	PPVT <sub>I</sub> vs PPVT <sub>II</sub>	85.85	1	<.001
PPVT <sub>III</sub> : $PPVT_{II} + e_{1ij}Age^1_{ij}$	955.84	PPVT <sub>II</sub> vs PPVT <sub>III</sub>	2.59	2	.27ns
PPVT <sub>IV</sub> : $PPVT_{III} + u_{10j}Age^1_{ij}$	946.25	PPVT <sub>III</sub> vs PPVT <sub>IV</sub>	9.59	2	.01
PPVT <sub>V</sub> : $PPVT_{IV} + \beta_2 Age^2_{ij}$	945.18	PPVT <sub>IV</sub> vs PPVT <sub>V</sub>	1.07	1	.30ns
PPVT <sub>DEF</sub> : $PPVT_I + \beta_{1j}Age^1_{ij}$	951.08				

<sup>a</sup> In addition to the intercept, variance components for differences within and between individuals are estimated.

Based on this General Development Model we constructed Figure 8.2, in which both the average development of PPVT as well as the differences within and between individuals are represented (see Table 6.2 in Appendix 6 for the parameter estimates). The average receptive vocabulary at an age of 72 months was estimated as 68.48. Each month a child grew older, his receptive vocabulary increased by 0.99.

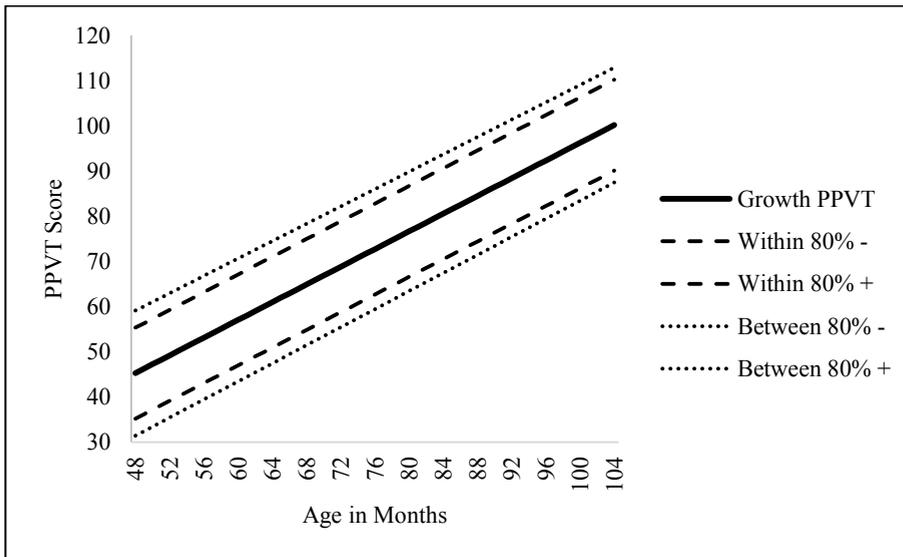


Figure 8.2. Graphical Representation of the General Development Model of PPVT (127 cases) with 80% Reliability.

Figure 8.2 shows that the effect of Age on PPVT scores is significant. The differences within individuals (represented by the striped lines in Figure 8.2) did not depend on Age. The differences between individuals (represented by the dotted lines in Figure 8.2) are a

function of Age; differences between older individuals are smaller than those between younger pupils.

Chapter 7, about the pedagogical practices in the classroom focusing on the focal pupils, revealed that six aspects of the school learning environment, as measured by the snapshot method, showed significant differences between the two school types. The individual pupil's percentages of time spent engaged in these six types of activities, language situations, or interaction settings (Language Activities, Mathematical Activities, Language Situation with Peers, Balanced Language Situation, Peer Interactions, and Teacher-Focal Pupil Interaction) were added one by one to the new General Development Model of PPVT, to see whether these aspects of the school learning environment were significant predictors of differences in pupils' PPVT development. However, the addition of the six variables from the Snapshot did not contribute significantly to the fit of the model to the observed data. This means that we could not show that the fact that there were differences between the school types in time spent by the pupils in these six activities, language situations, or interaction settings influenced their PPVT development.

### **8.3 The influence of the school learning environment on narrative ability development**

In this section the significant variables from the school learning environment, observed in Chapter 6 and 7, will be related to the individual development of the narrative ability of newly arrived migrant kindergarteners (using the Multilingual Assessment Instrument for Narratives, abbreviated as MAIN; Gagarina et al., 2012) answering the second part of the research question: To what extent do differences in the characteristics of the school learning environment during the first year after arrival relate to *narrative ability development* of newly arrived migrant kindergarteners during the first two-and-a-half years of schooling in the Netherlands?

The characteristics of the school learning environment were identified based on the observations of the pedagogical practices in the classrooms of the kindergarteners. First the relation between narrative ability and pedagogical practices concerning teacher behavior will be discussed. Then the relation between the pedagogical practices concerning focal pupils' experiences will be discussed. In the following sections all six aspects of narrative ability obtain by using the MAIN will be investigated. First the three microstructural elements: Number of Different Words, Guiraud Index Score, and Measure of Lexical Richness will be modelled. Then the three macrostructural elements will be modelled: Story Structure, Structural Complexity, and use of Internal State Terms.

As in Chapter 5, developmental models of each of the six narrative aspects will be built and compared to see whether the model fit improved. The pedagogical variables are added to the models in two ways: as main effect and as an interaction effect with Age. The former indicates that the narrative measure scores increase (or decrease) due to differences

in scores on the pedagogical variables. The latter effect indicates that the influence of the pedagogical variables on narrative ability depend on the age of the pupils.

### **8.3.1 Pedagogical practices: focusing on teacher behavior**

Due to logistical issues, some observations of the teachers took place after the assessment of the first (or first two) MAIN assessments on the pupils. Therefore, 54 MAIN scores obtained from pupils prior to the CLASS observations, mainly from Session 0, were taken out of the analyses. In total 114 cases from 38 different pupils could be taken into account, and therefore new developmental models needed to be modelled first with these 114 cases for all aspects of microstructure and macrostructure. After that, the analysis of the relation of pedagogical practices and narrative ability continued. The two significant different outcomes between school types on aspects of teacher behavior measured by CLASS obtained in Chapter 6, namely Positive Climate and Regard for Student Perspectives, will be added to these new General Development Models.

#### **Microstructure**

In the next three sub-sections, the microstructural elements of the narrative ability will be modelled and related to learning environmental aspects, based on observed teacher behavior. First the Number of Different Words (NDW), then the Guiraud Index Score (GIS), and finally the Measure of Lexical Richness (MLR) will be modelled.

#### ***Number of Different Words( NDW)***

From the comparison between the consecutive models (see Table 8.4) it is apparent that a model with a fixed linear component – allowing for differences in Age – fit the data better than a model with only an intercept ( $\Delta\chi^2$  (NDW<sub>b</sub>) = 48.08;  $df = 1$ ;  $p < .001$ ). Neither the variance within individuals (NDW<sub>c</sub>) nor the variance between (NDW<sub>d</sub>) depends on the age of the participants. Further, adding Age<sup>2</sup> to the model did not improved the fit significantly (NDW<sub>e</sub>). Hence, in the final model (NDW<sub>b</sub>) a fixed effect of Age is needed, and with this model we continued the analysis.

Table 8.4: Fit of Different Polynomials (-2LL) for Changes in NDW Score (114 cases) as well as the Comparison of Consecutive Models.

Model	-2LL	Comparison			
		Models	$\Delta\chi^2$	$\Delta df$	<i>p</i>
NDW <sub>a</sub> : $\beta_{0ij}cons^a$	907.83				
NDW <sub>b</sub> : $NDW_a + \beta_1 Age^1_{ij}$	859.75	NDW <sub>a</sub> vs NDW <sub>b</sub>	48.08	1	<.001
NDW <sub>c</sub> : $NDW_b + e_{1ij}Age^1_{ij}$	859.35	NDW <sub>b</sub> vs NDW <sub>c</sub>	0.40	2	.82ns
NDW <sub>d</sub> : $NDW_c + u_{10j}Age^1_{ij}$	857.48	NDW <sub>c</sub> vs NDW <sub>d</sub>	1.87	2	.39ns
NDW <sub>e</sub> : $NDW_b + \beta_2 Age^2_{ij}$	858.71	NDW <sub>b</sub> vs NDW <sub>e</sub> <sup>b</sup>	1.04	1	.31ns

<sup>a</sup> In addition to the intercept, variance components for differences within and between individuals are estimated.

<sup>b</sup> The addition of Age<sup>2</sup> to model NDW<sub>d</sub> did not converge within 250 iterations, hence Age<sup>2</sup> is added to NDW<sub>b</sub> since models <sub>c</sub> and <sub>d</sub> were not significant anyway. Therefore, models NDW<sub>b</sub> and NDW<sub>e</sub> are compared with each other.

Based on this General Development Model we constructed Figure 8.3, in which both the average development as well as the differences within and between individuals are represented (see Table 6.3 in Appendix 6 for the parameter estimates). The average NDW score at the age of 73 months was estimated as 37.37. Each month a child grew older, his NDW score increased by 0.63

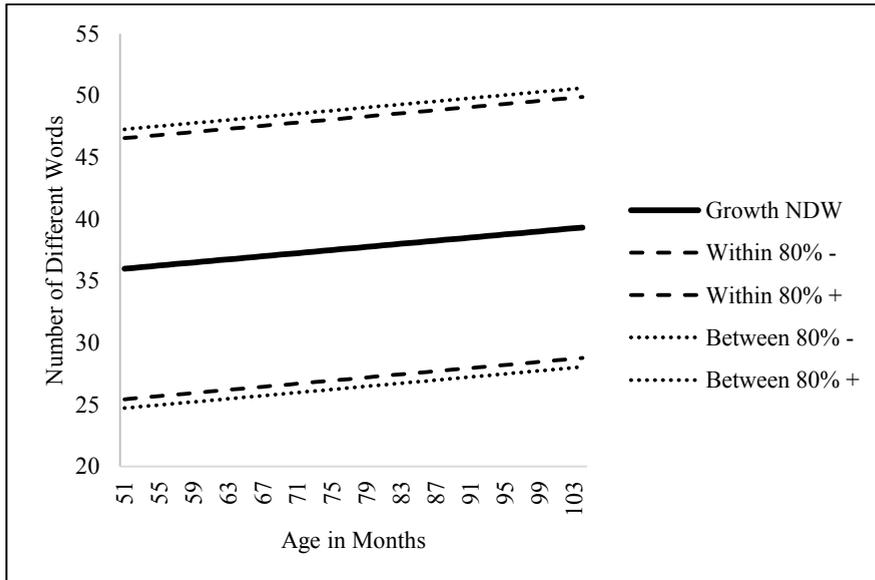


Figure 8.3: Graphical Representation of the General Development Model NDW (114 cases) with 80% Reliability.

Figure 8.3 shows that the effect of Age is significant. The differences within individuals (represented by the striped lines in Figure 8.3) and the differences between individuals (represented by the dotted lines in Figure 8.3) do not depend on Age. We cannot give more precise estimates of the NDW development for older pupils or younger pupils, nor can we say that the variation between older pupils is different from that of younger pupils.

Likelihood ratio tests did not show that main effects of Positive Climate, or Regard for Student Perspectives, contributed significantly to the fit of the model to the observed data, nor did we find an interaction between Age and Positive Climate or Regard for Student Perspectives and NDW. This means that we could not show that the fact that teachers at Mainstream schools were more likely to provide a higher Positive Climate and take into account student perspectives compared to teacher at the DL2-schools had any significant impact on the pupils' development of lexical diversity as measured with the NDW.

### **Guiraud Index Score (GIS)**

From the comparison between the consecutive models (see Table 8.5) it is apparent that a model for GIS with a fixed linear component – allowing for differences in Age – fit the data better than a model with only an intercept ( $\Delta_X^2$  (GIS<sub>b</sub>) = 41.83;  $df = 1$ ;  $p < .001$ ). The variance within individuals did not depend on the age of participants (GIS<sub>c</sub>). However, the variance between individuals is a (linear) function of age (GIS<sub>d</sub>). Adding Age<sup>2</sup> to the model did not improve the fit significantly (GIS<sub>e</sub>). Hence, in the final model (GIS<sub>DEF</sub>) a fixed effect of Age, as well as a variance between pupils component which depends on Age needs to be included, and with this model we continued the analysis.

Table 8.5: Fit of Different Polynomials (-2LL) for Changes in Guiraud Index Score (114 cases) as well as the Comparison of Consecutive Models.

Model	-2LL	Comparison			
		Models	$\Delta_X^2$	$\Delta df$	$p$
GIS <sub>a</sub> : $\beta_{0ij}cons$ <sup>a</sup>	262.07				
GIS <sub>b</sub> : GIS <sub>a</sub> + $\beta_1 Age^1_{ij}$	220.25	GIS <sub>a</sub> vs GIS <sub>b</sub>	41.83	1	<.001
GIS <sub>c</sub> : GIS <sub>b</sub> + $e_{1ij}Age^1_{ij}$	219.00	GIS <sub>b</sub> vs GIS <sub>c</sub>	1.25	1 <sup>b</sup>	.26ns
GIS <sub>d</sub> : GIS <sub>c</sub> + $u_{10j}Age^1_{ij}$	214.75	GIS <sub>c</sub> vs GIS <sub>d</sub>	4,25	1 <sup>b</sup>	.04
GIS <sub>e</sub> : GIS <sub>d</sub> + $\beta_2 Age^2_{ij}$	212.15	GIS <sub>d</sub> vs GIS <sub>e</sub>	2.60	1	.11ns
GIS <sub>DEF</sub> : GIS <sub>a</sub> + $\beta_{1j} * Age^1_{ij}$	214.81				

<sup>a</sup> In addition to the intercept, variance components for differences within and between individuals are estimated.

<sup>b</sup> Only the covariance-coefficient between the intercept- and the age-residuals was estimated.

Based on this General Development Model we constructed Figure 8.4, in which both the average development of Guiraud Index Score as well as the differences within and between individuals are represented (see Table 6.4 in Appendix 6 for the parameter estimates). The average GIS at an age of 73 months was estimated as 3.78. Each month a child grew older, his GIS increased by 0.04

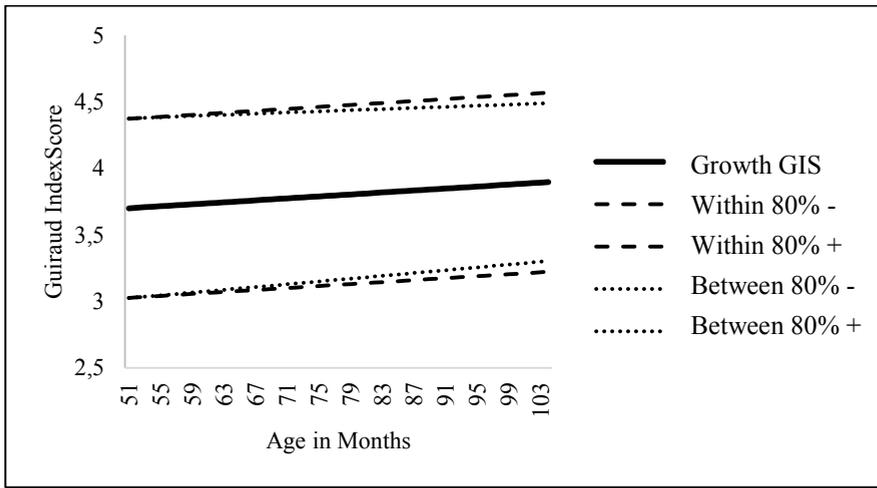


Figure 8.4: Graphical Representation of the General Development Model Guiraud Index Score (114 cases) with 80% Reliability.

Figure 8.4 shows that the effect of Age on GIS is significant. The differences within individuals (represented by the striped lines in Figure 8.4) do not depend on Age, however the differences between individuals (represented by the dotted lines in Figure 8.4) was a function of Age; differences in GIS score between older pupils is smaller than those between younger pupils.

The analysis was continued with the new General Development Model, GIS<sub>DEF</sub>. Positive Climate and Regard for Student Perspectives were added to this new General Development Model. Likelihood ratio tests did not show that main effects of Positive Climate and Regard for Student Perspectives contributed significantly to the fit of the model to the observed data, nor did we find an interaction between Age and Positive Climate or Regard for Student Perspectives for GIS. This means that we could not show that the fact that teachers at Mainstream schools were more likely to provide a higher Positive Climate and take into account student perspectives compared to teacher at the DL2-schools had any significant impact on the pupils' development of lexical diversity, as measured with the GIS.

**Measure of Lexical Richness (MLR)**

From the comparison between the consecutive models (see Table 8.6) it is apparent that a model for MLR with a fixed linear component – allowing for differences in Age – fit the data better than a model with only an intercept ( $\Delta\chi^2$  (MLR<sub>b</sub>) = 19.25;  $df = 1$ ;  $p < .001$ ). Neither the variance within individuals (MLR<sub>c</sub>) nor the variance between individuals (MLR<sub>d</sub>) depends on the age of participants. Further, adding Age<sup>2</sup> to the model did not improve the fit significantly (MLR<sub>e</sub>). Hence, in the final model (MLR<sub>b</sub>) a fixed effect of Age is needed, and with this model we continued the analysis<sup>22</sup>.

Table 8.6: Fit of Different Polynomials (-2LL) for Changes in MLR score (114 cases) as well as the Comparison of Consecutive Models.

Model	-2LL	Comparison			
		Models	$\Delta\chi^2$	$\Delta df$	$p$
MLR <sub>a</sub> : $\beta_{0ij}cons^a$	537.59				
MLR <sub>b</sub> : $MLR_a + \beta_1 Age^1_{ij}$	518.33	MLR <sub>a</sub> vs MLR <sub>b</sub>	19.25	1	<.001
MLR <sub>c</sub> : $MLR_b + e_{1ij}Age^1_{ij}$	516.67	MLR <sub>b</sub> vs MLR <sub>c</sub>	1.67	2	.44ns
MLR <sub>d</sub> : $MLR_c + u_{10i}Age^1_{ij}$	512.74	MLR <sub>c</sub> vs MLR <sub>d</sub>	3.93	2	.14ns
MLR <sub>e</sub> : $MLR_d + \beta_2 Age^2_{ij}$	512.73	MLR <sub>d</sub> vs MLR <sub>e</sub>	0.00	1	.96ns

<sup>a</sup> In addition to the intercept, variance components for differences within and between individuals are estimated.

Based on this General Development Model we constructed Figure 8.5, in which both the average development as well as the differences within and between individuals are represented (see Table 6.5 in Appendix 6 for the parameter estimates). The average MLR score at an age of 73 months was estimated as 2.73. Each month a child grew older, his MLR score increased by 0.09

<sup>22</sup> The reliability of this model is low, .53, this might cause the non-significant improvement of the model when we allowed the variance between and within individuals depend on age, or the non-significant effect of the addition of Age<sup>2</sup> in the model.

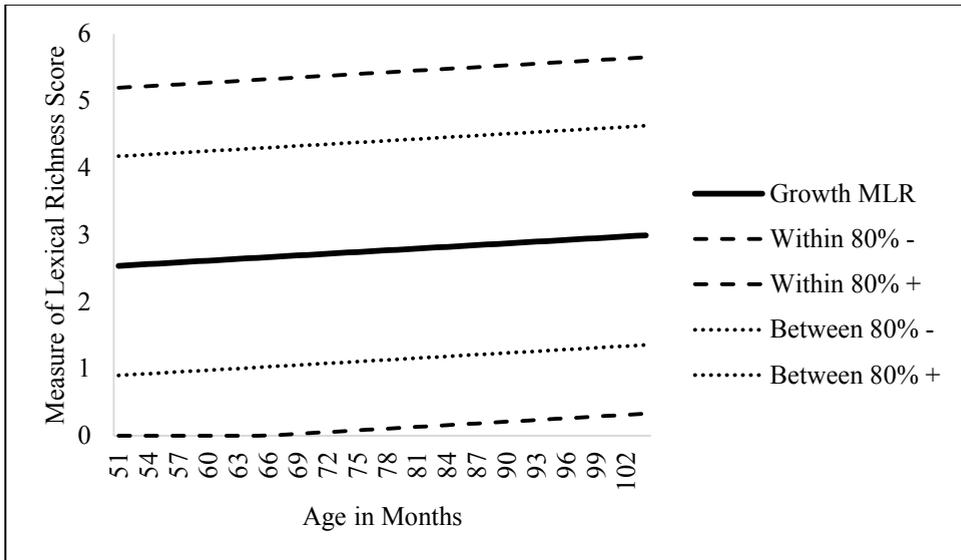


Figure 8.5: Graphical Representation of the General Development Model Measure of Lexical Richness (114 cases) with 80% reliability.

Figure 8.5 shows that the effect of Age on MLR score is significant. A likelihood ratio test showed that the main effect of Positive Climate ( $MLR_f$ ) did not contribute to the fit of the model to the observed data. However, the interaction between Age and Positive Climate ( $\Delta\chi^2 (MLR_g) = 3.90; df = 1; p = .048$ ; see Table 8.7) did improve the model significantly.

Table 8.7: Fit of Different Polynomials (-2LL) for Changes in MLR score (114 cases) with the Addition of Positive Climate (PosCli) as Explanatory Variable.

Model	-2LL	Comparison			
		Models	$\Delta\chi^2$	<i>Adf</i>	<i>p</i>
$MLR_b: \beta_{0ij}cons + \beta_1Age^{1}_{ij}$	518.33				
$MLR_f: MLR_b + \beta_2PosCli_{ij}$	515.65	$MLR_b$ vs $MLR_f$	2.68	1	.10ns
$MLR_g: MLR_f + \beta_3Age*PosCli_{ij}$	511.76	$MLR_f$ vs $MLR_g$	3.90	1	.048

There was an interaction effect of Positive Climate, although the fixed main effect of Positive Climate was not significant. The significant interaction effect between Age and Positive Climate means that the relation between Age and MLR scores differs per score on Positive Climate. However, for the interpretation this interaction effect a word of caution is warranted as the standard errors in the last model appeared to be large. Therefore, we are not sure about the specific values for each parameter (See Table 6.5 in Appendix 6 with the estimated parameters). Since the fixed main effect of Positive Climate is not significant we can only say that the influence of Positive Climate is larger for younger pupils than for

older pupils. We however do refrain from plotting the model in a graph because this would be difficult to interpret. A larger sample is necessary to confirm the effect.

A likelihood ratio test did not show that the main effect of Regard for Student Perspectives contributed significantly to the fit of the model to the observed data, nor did we find an interaction between Age and Regard for Student Perspectives for MLR. This means that we could not show that the fact that teachers at Mainstream schools were more likely to take into account student perspectives compared to teacher at the DL2-schools had any significant impact on the pupils' development of lexical richness.

### Macrostructure

In the next three sub-sections the macrostructural elements of the narrative ability will be modelled and related to learning environmental aspects based on observed teacher behavior. First the growth model of Story Structure (SS), then Structural Complexity (SC), and finally Internal State Terms (IST) is presented.

#### Story Structure (SS)

From the comparison between the consecutive models for SS score (Table 8.8) it is apparent that a model with a fixed linear component – allowing for differences in Age – fit the data better than a model with only an intercept ( $\Delta\chi^2$  (SS<sub>b</sub>) = 55.90;  $df = 1$ ;  $p < .001$ ). The variance within individuals depends on the age of participants ( $\Delta\chi^2$  (SS<sub>c</sub>) = 11.98;  $df = 2$ ;  $p = .002$ ). However, the variance between individuals (SS<sub>d</sub>) did not. Finally, adding Age<sup>2</sup> to the model improved the fit significantly ( $\Delta\chi^2$  (SS<sub>e</sub>) = 10.17;  $df = 1$ ;  $p = .001$ ). Hence, in the final model (SS<sub>DEF</sub>) a fixed effect of Age<sup>1</sup> and Age<sup>2</sup>, as well as a variance within pupils component which depends on Age needs to be included, and with this model we continued the analysis.

Table 8.8: Fit of Different Polynomials (-2LL) for Changes in Story Structure (114 cases) as well as the Comparison of Consecutive Models.

Model	-2LL	Comparison			
		Models	$\Delta\chi^2$	$\Delta df$	$p$
SS <sub>a</sub> : $\beta_{0ij}cons$ <sup>a</sup>	587.10				
SS <sub>b</sub> : $SS_a + \beta_1 Age^1_{ij}$	520.20	SS <sub>a</sub> vs SS <sub>b</sub>	55.90	1	<.001
SS <sub>c</sub> : $SS_b + e_{1ij} Age^1_{ij}$	508.22	SS <sub>b</sub> vs SS <sub>c</sub>	11.98	2	.002
SS <sub>d</sub> : $SS_c + u_{10j} Age^1_{ij}$	506.53	SS <sub>c</sub> vs SS <sub>d</sub>	1.69	2	.43ns
SS <sub>e</sub> : $SS_d + \beta_2 Age^2_{ij}$	496.36	SS <sub>d</sub> vs SS <sub>e</sub>	10.17	1	.001
SS <sub>DEF</sub> : $SS_a + \beta_1 Age^1_{ij} + \beta_2 Age^2_{ij}$	503.37				

<sup>a</sup> In addition to the intercept, variance components for differences within and between individuals are estimated.

Based on this General Development Model we constructed Figure 8.6, in which both the average development of SS as well as the differences within and between individuals are represented (see Table 6.6 in Appendix 6 for the parameter estimates). The average SS score at an age of 73 months was estimated as 7.39. Each month a child grew older, his SS score increased by 0.19, but the quadratic function of Age decreased the SS score by 0.002 (since this is a small effect it is not visible in Figure 8.6).

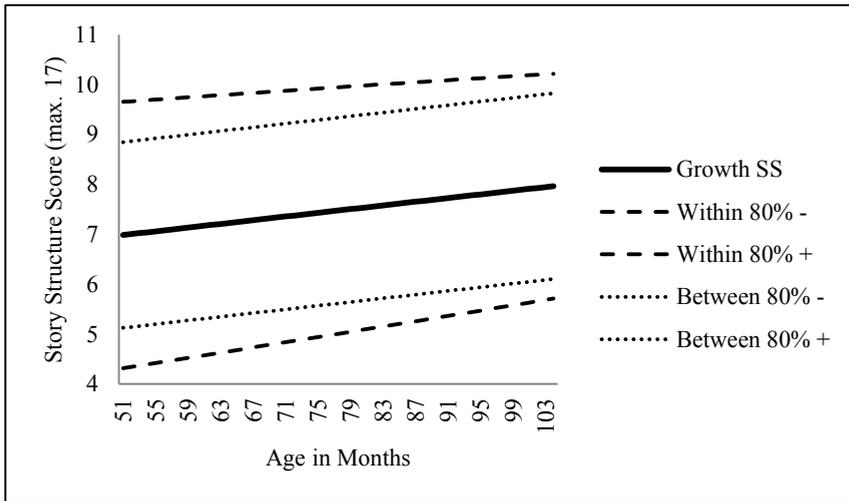


Figure 8.6: Graphical Representation of the General Development Model Story Structure (114 cases) with 80% Reliability.

Figure 8.6 shows that the effect of Age on SS is significant. This was a quadratic relation: the increase in SS development is larger for younger pupils. The differences within individuals (represented by the striped lines in Figure 8.6) also depend on Age; we can give a more precise estimate of the SS development for older pupils as compared to younger pupils. The difference between individuals (represented by the dotted lines in Figure 8.6) is not a function of Age; differences between older pupils are similar to that of younger pupils.

The analysis was continued with the new General Development Model,  $SS_{DEF}$ . The two significant variables of CLASS namely, Positive Climate and Regard for Student Perspectives were added to this new General Development Model. Likelihood ratio tests did not show that main effects of Positive Climate or Regard for Student Perspectives contributed significantly to the fit of the model to the observed data, nor that there was an interaction between Age and Positive Climate or Regard for Student Perspectives for SS. This means that we could not show that the fact that teachers at Mainstream schools were more likely to provide a higher Positive Climate and take into account student perspectives

compared to teacher at the DL2-schools had any significant impact on the pupils' development of Story Structure.

### ***Structural Complexity (SC)***

The reliability of the new growth model for SC, with only 114 cases, has a reliability of .22 which is too low. Therefore, we refrain from further analysis with this model for the relation between CLASS and Structural Complexity scores. Since the reliability in Chapter 5 already was low, .39, we expected that it would not be reliable in this chapter as well.

### ***Internal State Terms (IST)***

Even though the growth model for IST was reliable in Chapter 5, in this chapter, with only 114 cases instead of 168, we were not able to build a reliable model. The reliability of the General Development Model for the dependent variable Internal State Terms was .39, which is too low. Therefore, we also refrain from further analysis of this model and will not investigate the relation between CLASS and Internal State Terms.

## **8.3.2 Pedagogical practices: focusing on focal pupils' experiences**

The pedagogical practices on the individual level were obtained by observing the focal pupils with a snapshot method. Due to logistical issues, some observations of the focal pupils took place after the administration of the first MAIN assessments. Therefore, six MAIN assessments were taken out of the analyses. In total 162 cases, from 40 different pupils, could be taken into account, and therefore new developmental models have been modelled with these 162 cases.

Chapter 7, about the pedagogical practices in the classroom focusing on the focal pupils, revealed that six aspects of the school learning environment, as measured by the snapshot method, showed significant differences between the two school types. The individual pupil's percentages of time spent engaged in these six types of activities, language situations, or interaction settings (Language Activities, Mathematical Activities, Language Situation with Peers, Balanced Language Situation, Peer Interactions, and Teacher-Focal Pupil Interaction) were added one by one to the new General Development Model of the different aspects of narrative ability, to see whether these aspects of the school learning environment were significant predictors of differences in pupils' narrative ability development.

### **Microstructure**

In the next three sub-sections the new growth models for microstructural elements of the narrative ability will be modelled and related to learning environmental aspects focusing on focal pupils' experiences. First the Number of Different Words, then the Guiraud Index Score, and finally the Measure of Lexical Richness will be modelled. After the building of

the new growth models, we will present the significant explanatory variables from the school learning environment first, followed by the discussion of the non-significant variables.

**Number of Different Words (NDW)**

From the comparison between the models for NDW (Table 8.9) it is apparent that a model with a fixed linear component – allowing for differences in Age – fit the data better than a model with only an intercept ( $\Delta\chi^2$  (NDW<sub>II</sub>) = 71.72;  $df = 1$ ;  $p < .001$ ). The variance within, or between individuals depends on Age (NDW<sub>III</sub> and NDW<sub>IV</sub>). Adding Age<sup>2</sup> to the model did not improve the fit significantly (NDW<sub>V</sub>). Hence, in the final model (NDW<sub>II</sub>) a fixed effect of Age is needed, and with this model we continued the analysis.

Table 8.9: Fit of Different Polynomials (-2LL) for Changes in NDW (162 cases) as well as the Comparison of Consecutive Models.

Model	-2LL	Comparison			
		Models	$\Delta\chi^2$	$\Delta df$	$p$
NDW <sub>I</sub> : $\beta_{0ij}cons^a$	1314.56				
NDW <sub>II</sub> : $NDW_I + \beta_1 Age^1_{ij}$	1242.85	NDW <sub>I</sub> vs NDW <sub>II</sub>	71.72	1	<.001
NDW <sub>III</sub> : $NDW_{II} + \beta_{1ij} Age^1_{ij}$	1242.59	NDW <sub>II</sub> vs NDW <sub>III</sub>	0.26	2	.88ns
NDW <sub>IV</sub> : $NDW_{III} + \beta_{10j} Age^1_i$	1238.56	NDW <sub>III</sub> vs NDW <sub>IV</sub>	4.03	2	.13ns
NDW <sub>V</sub> : $NDW_{IV} + \beta_2 Age^2_{ij}$	1236.74	NDW <sub>IV</sub> vs NDW <sub>V</sub>	1.82	1	.18ns

<sup>a</sup> In addition to the intercept, variance components for differences within and between individuals are estimated.

Based on this General Development Model we constructed Figure 8.7, in which both the average development of NDW as well as the differences within and between individuals are represented (see Table 6.7 in Appendix 6 for the parameter estimates). The average NDW score at an age of 73 months was estimated as 35.56. Each month a child grew older, his NDW score increased by 0.71.

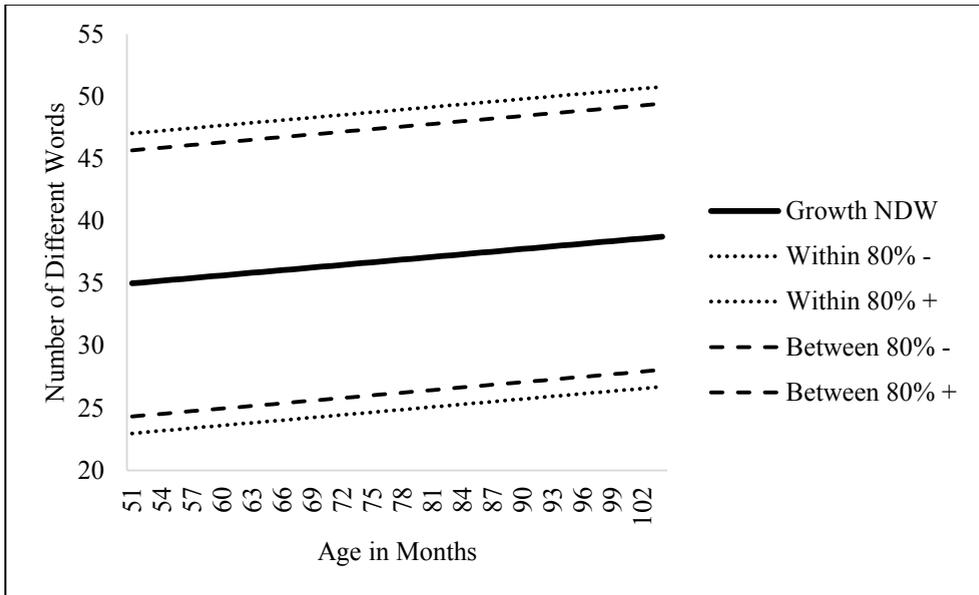


Figure 8.7: Graphical Representation of the General Development Model of NDW (162 cases) with 80% Reliability.

Figure 8.7 shows that of the effect of Age on NDW score is significant. With a constant variance within and between variance we cannot say that there are differences in the precision of our estimates between older and younger pupils, nor can we say that the variance between older pupils is different from that between younger pupils.

A likelihood ratio test showed that the main effect of Language Activities did not contribute significantly to the fit of the model to the observed data ( $NDW_{VI}$ ). However, the interaction between Age and Language Activities ( $\Delta\chi^2 (NDW_{VII}) = 4.90; df = 1; p = .03$ ; see Table 8.10) did improve the fit of the model significantly.

Table 8.10: Fit of Different Polynomials (-2LL) for Changes in NDW score (162 cases) with the Addition of Language Activities (LanAct) as Explanatory Variable.

Model	-2LL	Comparison			
		Models	$\Delta\chi^2$	$\Delta df$	$p$
$NDW_{II}: \beta_{0ij}cons + \beta_1Age^l_{ij}$	1242.85				
$NDW_{VI}: NDW_{II} + \beta_2LanAct_{ij}$	1242.60	$NDW_{II}$ vs $NDW_{VI}$	0.24	1	.62ns
$NDW_{VII}: NDW_{VI} + \beta_3Age * LanAct_{ij}$	1237.71	$NDW_{VI}$ vs $NDW_{VII}$	4.90	1	.03

There was an interaction effect of percentages of time spent in Language Activities, although the fixed main effect of Language Activities was not significant. The significant interaction effect between Age and Language Activities means that the relation between

Age and NDW scores differs per percentage of time spent in Language Activities. However, for the interpretation this interaction effect a word of caution is warranted as the standard errors in the last model appeared to be large. Therefore, we are not sure about the specific values for each parameter (See Table 6.7 in Appendix 6 with the estimated parameters). Since the fixed main effect of Language Activities is not significant we can only say that the influence of Language Activities is larger for older pupils than for younger pupils. We do however refrain from plotting the model in a graph because this would be difficult to interpret. A larger sample is necessary to confirm the effect. A likelihood ratio test showed that the main effect of Language Situations with Peers did not contribute significantly to the fit of the model to the observed data (NDW<sub>VIII</sub>). However, the interaction between Age and Language Situations with Peers ( $\Delta_X^2$  (NDW<sub>IX</sub>) = 4.11;  $df = 1$ ;  $p = .04$ ; see Table 8.11) did improve the model significantly.

Table 8.11: Fit of Different Polynomials (-2LL) for Changes in NDW score (162 cases) with the Addition of Language Situation with Peers (LanSitPeer) as Explanatory Variable.

Model	-2LL	Comparison			
		Models	$\Delta_X^2$	$\Delta df$	$p$
NDW <sub>II</sub> : $\beta_{0ij}cons + \beta_1Age^l_{ij}$	1242.85				
NDW <sub>VIII</sub> : NDW <sub>II</sub> + $\beta_2LanSitPeer_{ij}$	1240.38	NDW <sub>II</sub> vs NDW <sub>VIII</sub>	2.46	1	.12ns
NDW <sub>IX</sub> : NDW <sub>VIII</sub> + $\beta_3Age * LanSitPeer_{ij}$	1236.28	NDW <sub>VIII</sub> vs NDW <sub>IX</sub>	4.11	1	.04

There was an interaction effect of Language Situations with Peers, although the fixed main effect of Language Situations with Peers was not significant. This means that the relation between Age and NDW scores differs due to the percentages of Language Situations with Peers (see Table 6.7 in Appendix 6 for the estimated parameters). However, for the interpretation this interaction effect a word of caution is warranted as the standard errors in the last model appeared to be large. Therefore, we are not sure about the specific values for each parameter (See Table 6.7 in Appendix 6 with the estimated parameters). Since the fixed main effect of Language Situations with Peers is not significant we can only say that the influence of Language Situations with Peers is larger for older pupils than for younger pupils. We do however refrain from plotting the model in a graph because this would be difficult to interpret. A larger sample is necessary to confirm the effect.

The addition of the percentage of time spent in Peer Interaction improved the growth model of the NDW ( $\Delta_X^2$  (NDW<sub>XIV</sub>) = 3.87;  $df = 1$ ;  $p = .05$ ; see Table 8.12): the higher the percentage of time spent in Peer Interaction, the higher the NDW. An interaction between Age and percentage of time spent in Peer Interaction was not found (NDW<sub>XV</sub>).

Table 8.12: Fit of Different Polynomials (-2LL) for Changes in NDW (162 cases) with the Addition of Peer Interaction (PeerInt) as Explanatory Variable.

Model	-2LL	Comparison			
		Models	$\Delta\chi^2$	$\Delta df$	$p$
$NDW_{II}: \beta_{0ij}cons + \beta_1 Age^l_{ij}$	1242.85				
$NDW_X: NDW_{II} + \beta_2 PeerInt_{ij}$	1238.98	$NDW_{II}$ vs $NDW_X$	3.87	1	.05
$NDW_{Xi}: NDW_X + \beta_3 Age * PeerInt_{ij}$	1236.07	$NDW_X$ vs $NDW_{XI}$	2.91	1	.09ns

In Figure 8.8 the general development for pupils with different percentages of Peer Interaction is presented. There was a fixed main effect of the percentage of time spent in Peer Interaction, but there was no interaction effect between Age and percentage of Peer Interaction (see Table 6.7 in Appendix 6 with the estimated parameters). Pupils with a higher percentage of time spent in Peer Interaction had a higher NDW than pupils who spent a lower percentage of time in Peer Interaction.

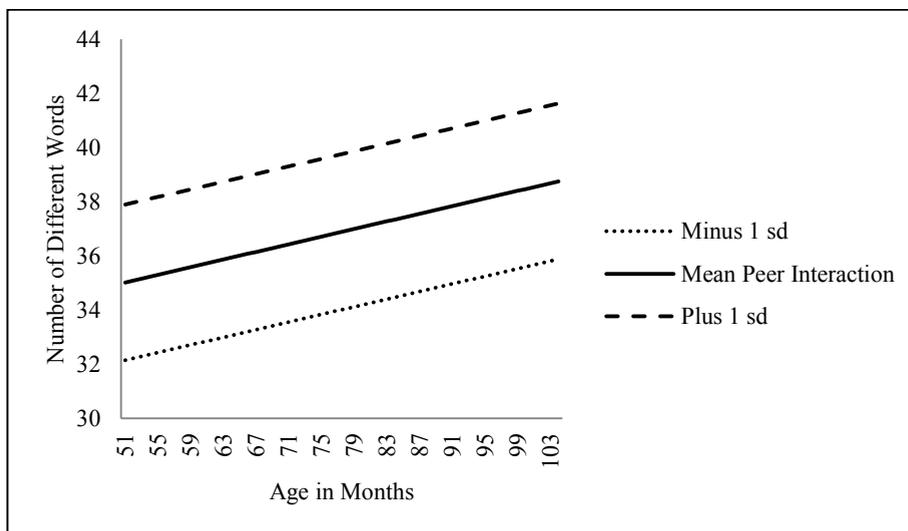


Figure 8.8: Development of NDW with Peer Interaction as Explanatory Variable.

The addition of the variables percentage of time spent on Mathematical Activities, Balanced Language Situations, or Teacher-Focal Pupil Interaction did not improve the general model of growth in NDW. This means that we could not show that the fact that pupils at DL2-schools were more likely to be engaged in Mathematical Activities, Balanced Language Situations, or Teacher-Focal Pupil Interactions compared to pupils at Mainstream schools had any significant impact on the pupils' development of lexical diversity as measured with the NDW.

**Guiraud Index Score (GIS)**

From the comparison between the consecutive models (see Table 8.13) it is apparent that a model for GIS with a fixed linear component – allowing for differences in Age – fit the data better than a model with only an intercept ( $\Delta_X^2$  (GIS<sub>II</sub>) = 57.07;  $df = 1$ ;  $p < .001$ ). The variance within individuals did not depend on the age of participants (GIS<sub>III</sub>). However, the variance between individuals is a function of Age (GIS<sub>IV</sub>). Adding Age<sup>2</sup> to the model did not improve the fit significantly (GIS<sub>IV</sub>). Hence, in the final model (GIS<sub>DEF</sub>) a fixed effect of Age, as well as a variance between pupils component which depends on Age needs to be included, and with this model we continued the analysis.

Table 8.13: Fit of Different Polynomials (-2LL) for Changes in Guiraud Index Score (162 cases) as well as the Comparison of Consecutive Models.

Model	-2LL	Comparison			
		Models	$\Delta_X^2$	$\Delta df$	$p$
GIS <sub>I</sub> : $\beta_{0ij}cons^a$	414.70				
GIS <sub>II</sub> : $GIS_I + \beta_1 Age^1_{ij}$	357.63	GIS <sub>I</sub> vs GIS <sub>II</sub>	57.07	1	<.001
GIS <sub>III</sub> : $GIS_{II} + e_{1i} Age^1_{ij}$	355.47	GIS <sub>II</sub> vs GIS <sub>III</sub>	2.16	2	.34ns
GIS <sub>IV</sub> : $GIS_{III} + u_{10j} Age^1_{ij}$	349.22	GIS <sub>III</sub> vs GIS <sub>IV</sub>	6.25	2	.04
GIS <sub>V</sub> : $GIS_{IV} + \beta_2 Age^2_{ij}$	348.05	GIS <sub>IV</sub> vs GIS <sub>V</sub>	1.17	1	.28ns
GIS <sub>DEF</sub> : $GIS_I + \beta_{1j} Age^1_{ij}$	349.52				

<sup>a</sup> In addition to the intercept, variance components for differences within and between individuals are estimated.

Based on this General Development Model we constructed Figure 8.9, in which both the average development of Guiraud Index Score as well as the differences within and between individuals are represented (see Table 6.8 in Appendix 6 for the parameter estimates). The average GIS at an age of 73 months was estimated as 3.66. Each month a child grew older, his GIS increased by 0.04.

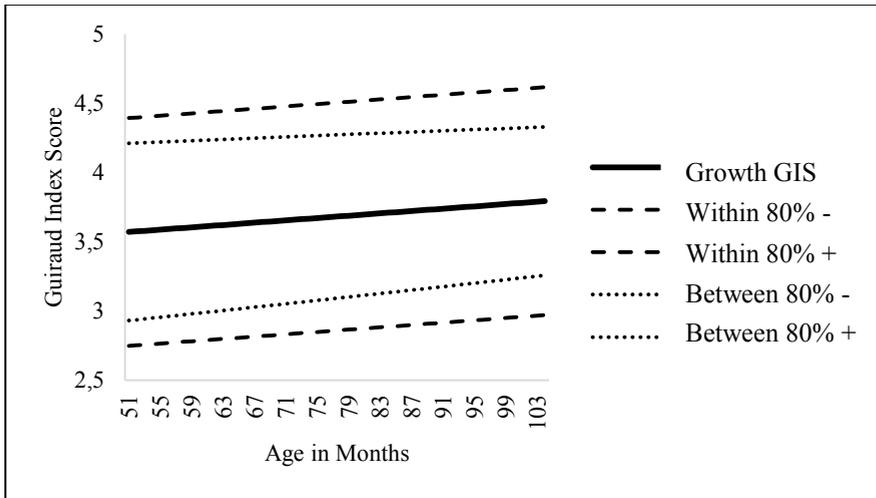


Figure 8.9: Graphical Representation of the General Development Model of Guiraud Index Score (162 cases) with 80% Reliability

Figure 8.9 shows that of the effect of Age on GIS is significant. The variance within individuals (represented by the striped lines in Figure 8.9) does not depend on Age, however the variance between individuals (represented by the dotted lines in Figure 8.9) was a function of Age; variance in GIS score between older individuals is smaller than those between younger pupils.

The analysis was continued with the new General Development Model,  $GIS_{DEF}$ . The inclusion of the percentage of time spent on Mathematical Activities ( $\Delta\chi^2 GIS_{VI} = 6.61$ ;  $df = 1$ ;  $p = .01$ ; see Table 8.14) as a fixed main effect improved the fit of the model for GIS significantly: the higher the percentage of time spent on Mathematical Activities, the lower the GIS. There was no interaction effect between Age and percentage of time spent in Mathematical Activities ( $GIS_{VII}$ ).

Table 8.14: Fit of Different Polynomials (-2LL) for Changes in Guiraud Index Score (162 cases) with the Addition of Mathematic Activities (MathAct) as Explanatory Variable.

Model	-2LL	Comparison			
		Models	$\Delta\chi^2$	$\Delta df$	$p$
$GIS_{IV}: \beta_{0ij}cons + \beta_{1j}Age^l_{ij}$	349.52				
$GIS_{VI}: GIS_{IV} + \beta_2 MathAct_{ij}$	342.90	$GIS_{IV}$ vs $GIS_{VI}$	6.61	1	.01
$GIS_{VII}: GIS_{VI} + \beta_3 Age * MathAct_{ij}$	341.66	$GIS_{VI}$ vs $GIS_{VII}$	1.24	1	.27ns

The difference in GIS score when the percentage of time spent on Mathematical Activities was taken into account is graphically presented in Figure 8.10. There was a negative fixed main effect of the percentage of time spent on Mathematical Activities, but there was no

interaction effect between Age and percentage of time spent on Mathematical Activities (see Table 6.8 in Appendix 6 with the estimated parameters). Pupils with a lower percentage of time spent on Mathematical Activities score higher on the GIS than pupils who spent a higher percentage of time on Mathematical Activities.

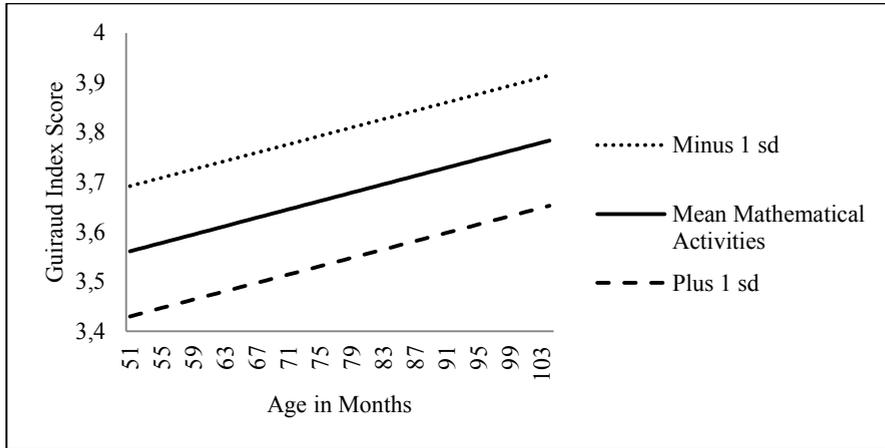


Figure 8.10: Development of Guiraud Index Score with Mathematical Activities as Explanatory Variable.

The addition of the percentage of time spent in Balanced Language Situations improved the General Development Model of the GIS ( $\Delta\chi^2$  (GIS<sub>VIII</sub>) = 3.83;  $df = 1$ ;  $p = .05$ ; see Table 8.15): the higher the percentage of time spent in Balanced Language Situations, the higher the GIS. No interaction between Age and percentage of time spent in Balanced Language Situations was found (GIS<sub>IX</sub>).

Table 8.15: Fit of Different Polynomials (-2LL) for Changes in Guiraud Index Score (162 cases) with the Addition of Balanced Language Situations (BalLanSit) as Explanatory Variable.

Model	-2LL	Comparison			
		Models	$\Delta\chi^2$	$\Delta df$	$p$
GIS <sub>IV</sub> : $\beta_{0ij}cons + \beta_{1j}Age^{l_{ij}}$	349.52				
GIS <sub>VIII</sub> : GIS <sub>IV</sub> + $\beta_2BalLanSit_{ij}$	345.69	GIS <sub>IV</sub> vs GIS <sub>VIII</sub>	3.83	1	.05
GIS <sub>IX</sub> : GIS <sub>VIII</sub> + $\beta_3Age*BalLanSit_{ij}$	344.61	GIS <sub>VIII</sub> vs GIS <sub>IX</sub>	1.08	1	.30ns

The difference in GIS when the percentage of time spent in Balanced Language Situations was taken into account is graphically presented in Figure 8.11. There was a fixed main effect of the percentage of time spent in Balanced Language Situations, but there was no interaction effect between Age and percentage of time spent in Balanced Language

Situations (see Table 6.8 in Appendix 6 for the estimated parameters). Pupils with higher percentages of time spent in Balanced Language Situations, had a higher GIS compared to pupils who spent a lower percentage of time in Balanced Language Situations.

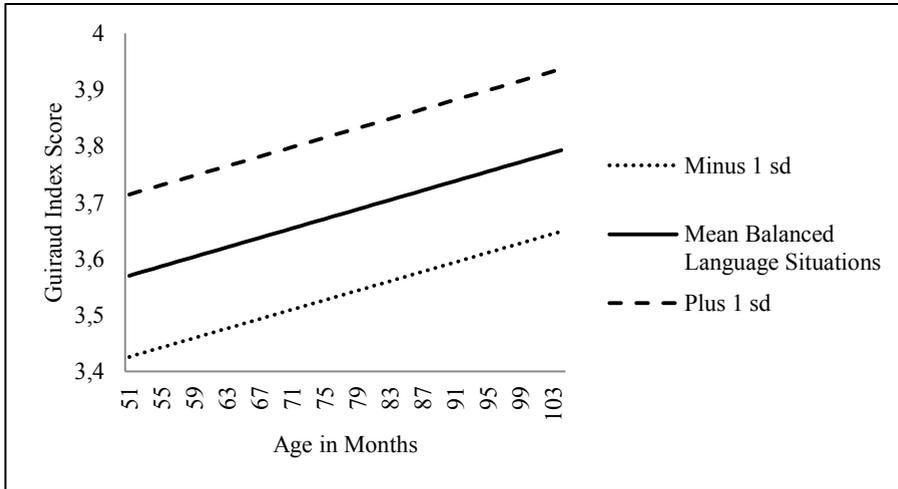


Figure 8.11: Development of Guiraud Index Score with Balanced Language Situations as Explanatory Variable.

The addition of the percentage of time spent in Teacher-Focal Pupil Interaction improved the General Development Model of the GIS ( $\Delta\chi^2$  (GIS<sub>X</sub>) = 8.77;  $df = 1$ ;  $p = .003$ ; see Table 8.16): the higher the percentage of time spent in Teacher-Focal Pupil Interaction, the higher the GIS. No interaction between Age and percentage of time spent in Teacher-Focal Pupil Interaction was found (GIS<sub>IX</sub>).

Table 8.16: Fit of Different Polynomials (-2LL) for Changes in Guiraud Index Score (162 cases) with the Addition of Teacher-Focal Pupil Interaction (TeFPInt) as Explanatory Variable.

Model	-2LL	Comparison			
		Models	$\Delta\chi^2$	$\Delta df$	$p$
GIS <sub>IV</sub> : $\beta_{0ij}cons + \beta_{1j}Age^l_{ij}$	349.52				
GIS <sub>X</sub> : GIS <sub>IV</sub> + $\beta_2TeFPInt_{ij}$	340.75	GIS <sub>IV</sub> vs GIS <sub>X</sub>	8.77	1	.003
GIS <sub>XI</sub> : GIS <sub>X</sub> + $\beta_3Age*TeFPInt_{ij}$	338.85	GIS <sub>X</sub> vs GIS <sub>XI</sub>	1.90	1	.17ns

The difference in GIS when the percentage of time spent in Teacher-Focal Pupil Interaction was taken into account is graphically presented in Figure 8.12. There was a fixed main effect of the percentage of time spent in Teacher-Focal Pupil Interaction, but there was no interaction effect between Age and percentage of time spent in Teacher-Focal Pupil

Interaction (see Table 6.8 in Appendix 6 with the estimated parameters). Pupils with higher percentages of time spent in Teacher-Focal Pupil Interaction, had a higher GIS compared to pupils who spent lower percentages of time in Pupil Interaction.

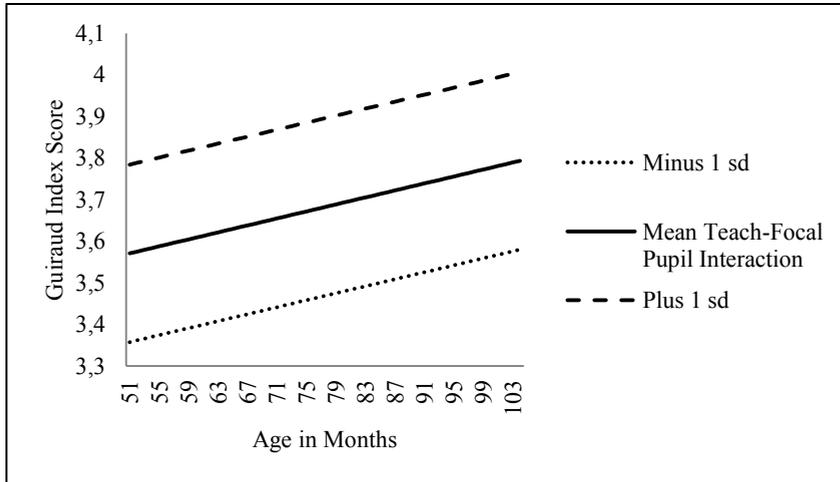


Figure 8.12: Development of Guiraud Index Score with Teacher-Focal Pupil Interaction as Explanatory Variable.

The addition of the variables percentage of time spent on Language Activities or percentage of time spent in Language Situations with Peers or Peer Interaction did not improve the general model of growth in GIS. This means that we could not show that the fact that pupils at DL2-schools were more likely to be engaged in Language Activities compared to pupils at Mainstream schools had any significant impact on the pupils' development of lexical diversity as measured with the GIS. Additionally, we could not show that the fact that pupils at Mainstream schools were more likely to be engaged in Language Situations with Peers and Peer Interaction compared to pupils at the DL2-schools had any significant impact on the pupils' development of lexical diversity as measured with the GIS.

### Measure of Lexical Richness (MLR)

From the comparison between the consecutive models for the MLR score (Table 8.17) it is apparent that a model with a fixed linear component – allowing for differences in Age – fit the data better than a model with only an intercept ( $\Delta_X^2$  (MLR<sub>II</sub>) = 19.73;  $df = 1$ ;  $p < .001$ ). The variance within individuals depends on the age of the participants ( $\Delta_X^2$  (MLR<sub>III</sub>) = 54.56;  $df = 2$ ;  $p < .001$ ). The variance between individuals is a (linear) function of age as well ( $\Delta_X^2$  (MLR<sub>IV</sub>) = 13.00;  $df = 1$ ;  $p < .001$ ). Adding Age<sup>2</sup> to the model improved the fit significantly ( $\Delta_X^2$  (MLR<sub>V</sub>) = 6.43;  $df = 1$ ;  $p = .01$ ). Hence, in the final model (MLR<sub>V</sub>) a fixed

effect of Age<sup>1</sup> and Age<sup>2</sup>, as well as variance within and between pupils components which depends on Age needs to be included, and with this model we continued the analysis.

Table 8.17: Fit of Different Polynomials (-2LL) for Changes in MLR score (162 cases) as well as the Comparison of Consecutive Models.

Model	-2LL	Comparison			
		Models	$\Delta\chi^2$	$\Delta df$	$p$
MLR <sub>I</sub> : $\beta_{0ij}cons$ <sup>a</sup>	736.53				
MLR <sub>II</sub> : $MLR_I + \beta_1 Age^1_{ij}$	716.81	MLR <sub>I</sub> vs MLR <sub>II</sub>	19.73	1	<.001
MLR <sub>III</sub> : $MLR_{II} + e_{1ij}Age^1_{ij}$	662.24	MLR <sub>II</sub> vs MLR <sub>III</sub>	54.56	2	<.001
MLR <sub>IV</sub> : $MLR_{III} + u_{10j}Age^1_{ij}$	649.25	MLR <sub>III</sub> vs MLR <sub>IV</sub>	13.00	1	<.001 <sup>b</sup>
MLR <sub>V</sub> : $MLR_{IV} + \beta_2 Age^2_{ij}$	642.81	MLR <sub>IV</sub> vs MLR <sub>V</sub>	6.43	1	.01
MLR <sub>VI</sub> : $MLR_V + e_{2ij}Age^2_{ij}$	636.30	MLR <sub>V</sub> vs MLR <sub>VI</sub>	6.51	3	.09ns
MLR <sub>VII</sub> : $MLR_{VI} + u_{2ij}Age^2_{ij}$ <sup>c</sup>					

<sup>a</sup> In addition to the intercept, variance components for differences within and between individuals are estimated.

<sup>b</sup> Only the covariance-coefficient between the intercept- and the age-residuals was estimated.

<sup>c</sup> No convergence within 250 iterations.

Based on this General Development Model we constructed Figure 8.13, in which both the average development as well as the differences within and between individuals are represented (see Table 6.9 in Appendix 6 for the parameter estimates). The average MLR at an age of 73 months was estimated as 2.26. Each month a child grew older, his MLR score increased by 0.07, but the quadratic function of Age decreased the MLR score with 0.003 (since this is a small effect it is not visible in Figure 8.13).

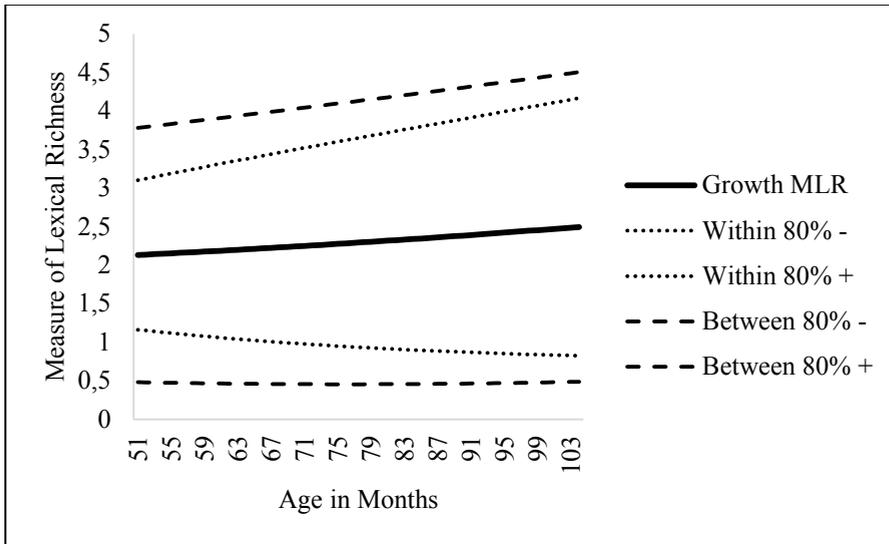


Figure 8.13: Graphical Representation of the General Development Model Growth MLR score (162 cases) with 80% Reliability.

Figure 8.13 shows that the effect of Age on MLR was significant. This is a quadratic relation; the increase in vocabulary development is larger for younger pupils. The differences within individuals (represented by the striped lines in Figure 8.13) also depend on Age; we can give a more precise estimate of the vocabulary development for younger pupils as compared to older pupils. The differences between individuals (represented by the dotted lines in Figure 8.13) are a function of Age as well; differences between younger individuals are smaller than those between older pupils.

The addition of the percentage of time spent in Peer Interaction did not improve the General Development Model of MLR (MLR<sub>VIII</sub>). However, an interaction between Age and percentage of time spent in Peer Interaction was found ( $\Delta\chi^2$  (MLR<sub>IX</sub>) = 5.21;  $df = 1$ ;  $p = .02$ ; see Table 8.18).

Table 8.18: Fit of Different Polynomials (-2LL) for Changes in MLR score (162 cases) with the Addition of Peer Interaction (PeerInt) as Explanatory Variable.

Model	-2LL	Comparison			
		Models	$\Delta\chi^2$	$\Delta df$	$p$
MLR <sub>V</sub> : $\beta_{0ij}cons + \beta_{1ij}Age^1_{ij} + \beta_{2ij}Age^2_{ij}$	642.81				
MLR <sub>VIII</sub> : MLR <sub>V</sub> + $\beta_3PeerInt_{ij}$	642.39	MLR <sub>V</sub> vs MLR <sub>VIII</sub>	0.42	1	.52ns
MLR <sub>IX</sub> : MLR <sub>VIII</sub> + $\beta_4Age*PeerInt_{ij}$	637.18	MLR <sub>VIII</sub> vs MLR <sub>IX</sub>	5.21	1	.02

There was an interaction effect of Peer Interaction, although the fixed main effect of Peer Interaction was not significant. The significant interaction effect between Age and Peer Interaction means that the relation between Age and MLR scores differs per percentage of time spent in Peer Interaction. However, for the interpretation this interaction effect a word of caution is warranted as the standard errors in the last model appeared to be large. Therefore, we are not sure about the specific values for each parameter (See Table 6.9 in Appendix 6 with the estimated parameters). Since the fixed main effect of Peer Interaction is not significant we can only say that the influence of Peer Interaction is larger for younger pupils than for older pupils. We do however refrain from plotting the model in a graph because this would be difficult to interpret. A larger sample is necessary to confirm the effect.

The addition of the other five variables, percentage of time spent on Language Activities and Mathematical Activities or time spent in Language Situations with Peers, Balanced Language Situations, or Teacher-Focal Pupil Interaction did not improve the general model of growth in MLR. This means that we could not show that the fact that pupils at DL2-schools were more likely to be engaged in Language Activities, Mathematical Activities, Balanced Language Situations, or Teacher-Focal Pupil Interactions compared to pupils at Mainstream schools had any significant impact on the pupils' development of lexical richness. Additionally, we could not show that the fact that pupils at Mainstream schools were more likely to be engaged in Language Situations with Peers compared to pupils at the DL2-schools had any significant impact on the pupils' development of lexical richness.

### **Macrostructure**

In the next three sub-sections the new growth models for macrostructural elements of the narrative ability will be modelled and related to learning environmental aspects focusing on focal pupils' experiences. First Story Structure (SS), then Structural Complexity (SC), and finally Internal State Term (IST) will be modelled. After the building of the new growth models, we will present the significant explanatory variables from the school learning environment first, followed by the discussion of the non-significant variables.

#### ***Story Structure (SS)***

From the comparison between the consecutive models for the SS score (Table 8.19) it is apparent that a model with a fixed linear component – allowing for differences in Age – fit the data better than a model with only an intercept ( $\Delta_X^2$  (SS<sub>II</sub>) = 93.14;  $df = 1$ ;  $p < .001$ ). The variance within (SS<sub>III</sub>) or between (SS<sub>IV</sub>) individuals did not depend on the age of the participants. Adding Age<sup>2</sup> to the model improved the fit significantly ( $\Delta_X^2$  (SS<sub>V</sub>) = 5.15;  $df = 1$ ;  $p = .02$ ). Hence, in the final model (SS<sub>DEF</sub>) a fixed effect of Age<sup>1</sup> and Age<sup>2</sup> is needed, and with this model we continued the analysis.

Table 8.19: Fit of Different Polynomials (-2LL) for Changes in Story Structure (162 cases).

Model	-2LL	Comparison			
		Models	$\Delta\chi^2$	$\Delta df$	<i>p</i>
SS <sub>I</sub> : $\beta_{0ij}cons^a$	840.33				
SS <sub>II</sub> : $SS_I + \beta_1 Age^1_{ij}$	747.19	SS <sub>I</sub> vs SS <sub>II</sub>	93.14	1	<.001
SS <sub>III</sub> : $SS_{II} + e_{1ij}Age^1_{ij}$	746.58	SS <sub>II</sub> vs SS <sub>III</sub>	0.61	1	.44ns
SS <sub>IV</sub> : $SS_{III} + u_{10j}Age^1_{ij}$	746.04	SS <sub>III</sub> vs SS <sub>IV</sub>	0.54	2	.76ns
SS <sub>V</sub> : $SS_{IV} + \beta_2 Age^2_{ij}$	740.87	SS <sub>IV</sub> vs SS <sub>V</sub>	5.15	1	.02
SS <sub>VI</sub> : $SS_V + e_{2ij}Age^2_{ij}$	737.26	SS <sub>V</sub> vs SS <sub>VI</sub>	3.63	3	.31ns
SS <sub>VII</sub> : $SS_{VI} + u_{20j}Age^2_{ij}$					
SS <sub>DEF</sub> : $SS_I + \beta_1 Age^1_{ij} + \beta_2 Age^2_{ij}$	743.78				

<sup>a</sup> In addition to the intercept, variance components for differences within and between individuals are estimated.

<sup>b</sup> No convergence within 250 iterations.

Based on this General Development Model we constructed Figure 8.14, in which both the average development as well as the differences within and between individuals are represented (see Table 6.10 in Appendix 6 for the parameter estimates). The average SS score at an age of 73 months was estimated as 6.91. Each month a child grew older, his SS score increased by 0.21, but the quadratic function of Age decreased the score with 0.002 (since this is a small effect it is not visible in Figure 8.14).

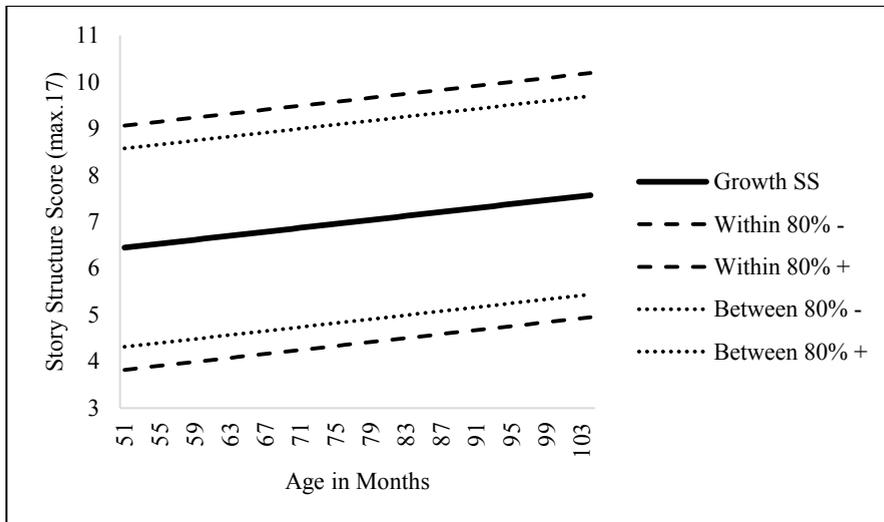


Figure 8.14: Graphical Representation of the General Development Model of Story Structure (162 cases) with 80% Reliability.

Figure 8.14 shows that the effect of Age on SS is significant. This is a quadratic relation; the increase in SS development is larger for younger pupils. With a constant variance within and between variance we cannot say that there are differences in the precision of our estimates between older and younger pupils, nor can we say that the variance between older pupils is different than the variance between younger pupils.

The analysis was continued with the new General Development Model,  $SS_{DEF}$ . The addition of the percentage of time spent in Balanced Language Situations improved the General Development Model of the SS score ( $\Delta_X^2(SS_{VIII}) = 4.37$ ;  $df = 1$ ;  $p = .04$ ; see Table 8.20) the higher the percentage of time spent in Balanced Language Situations, the higher the SS.. No interaction between Age and percentage of time spent in Balanced Language Situations was found ( $SS_{IX}$ ).

Table 8.20: Fit of Different Polynomials (-2LL) for Changes in Story Structure (162 cases) with the Addition of Balanced Language Situations (BalLanSit) as Explanatory Variable.

Model	-2LL	Comparison		
		Models	$\Delta_X^2$	$\Delta df$ $p$
$SS_{DEF}: \beta_{0ij}cons + \beta_1Age^1_{ij} + \beta_2Age^2_{ij}$	743.78			
$SS_{VIII}: SS_{DEF} + \beta_3BalLanSit_{ij}$	739.41	$SS_{DEF}$ vs $SS_{VIII}$	4.37	1 .04
$SS_{IX}: SS_{VIII} + \beta_4Age * BalLanSit_{ij}$	736.16	$SS_{VIII}$ vs $SS_{IX}$	3.24	1 .07ns

The difference in SS when the percentage of time spent in Balanced Language Situations was taken into account is graphically presented in Figure 8.15. There was a fixed main effect of the percentage of time spent in Balanced Language Situations, but there was no interaction effect between Age and percentage of time spent in Balanced Language Situations (see Table 6.10 in Appendix 6 for the estimated parameters). Pupils with higher percentages of time spent in Balanced Language Situations had higher SS scores.

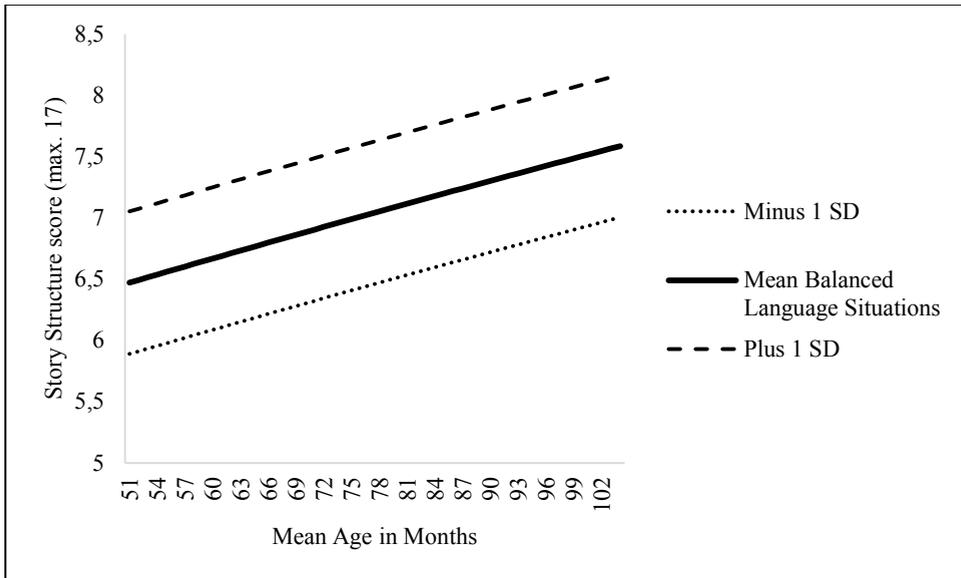


Figure 8.15: Development of Story Structure with Balanced Language Situation as Explanatory Variable.

A likelihood ratio test showed that the main effect of the percentage of time spent in Teacher-Focal Pupil Interaction contributed significantly to the fit of the model to the observed data ( $\Delta\chi^2 (SS_X) = 8.45; df = 1; p = .004$ ). The same holds for the interaction between Age and percentage of time spent in Teacher-Focal Pupil Interaction ( $\Delta\chi^2 (SS_{XI}) = 6.02; df = 1; p = .01$ ; see Table 8.21).

Table 8.21: Fit of Different Polynomials (-2LL) for Changes in Story Structure (162 cases) with the Addition of Teacher-Focal Pupil Interaction (TeFPInt) as Explanatory Variable.

Model	-2LL	Comparison			
		Models	$\Delta\chi^2$	$\Delta df$	$p$
$SS_{DEF}: \beta_{0ij}cons + \beta_1Age^1_{ij} + \beta_2Age^2_{ij}$	743.78				
$SS_X: SS_{DEF} + \beta_3TeFPInt_{ij}$	735.33	$SS_{DEF}$ vs $SS_X$	8.45	1	.004
$SS_{XI}: SS_X + \beta_4Age*TeFPInt_{ij}$	729.32	$SS_X$ vs $SS_{XI}$	6.02	1	.01

In Figure 8.16 the general development for pupils with different percentages of Teacher-Focal Pupil Interaction is presented. There was a fixed main effect of the percentage of time spent in Teacher-Focal Pupil Interaction and an interaction effect between Age and percentage of time spent in Teacher-Focal Pupil Interaction, which means that the relationship between Age and SS scores differs per percentage of time spent in Teacher-Focal Pupil Interaction (see Table 6.10 in Appendix 6 with the estimated parameters).

Pupils who are more engaged in Teacher-Focal Pupil Interaction tell more complete stories than pupils who are less engaged in such interactions, however, the influence of the percentage of time spent in Teacher Focal Pupil Interaction on SS score is stronger for younger pupils than for older pupils. Additionally, as the lines show in Figure 8.16, all Teacher-Focal Pupil interaction results in higher Story Structure scores.

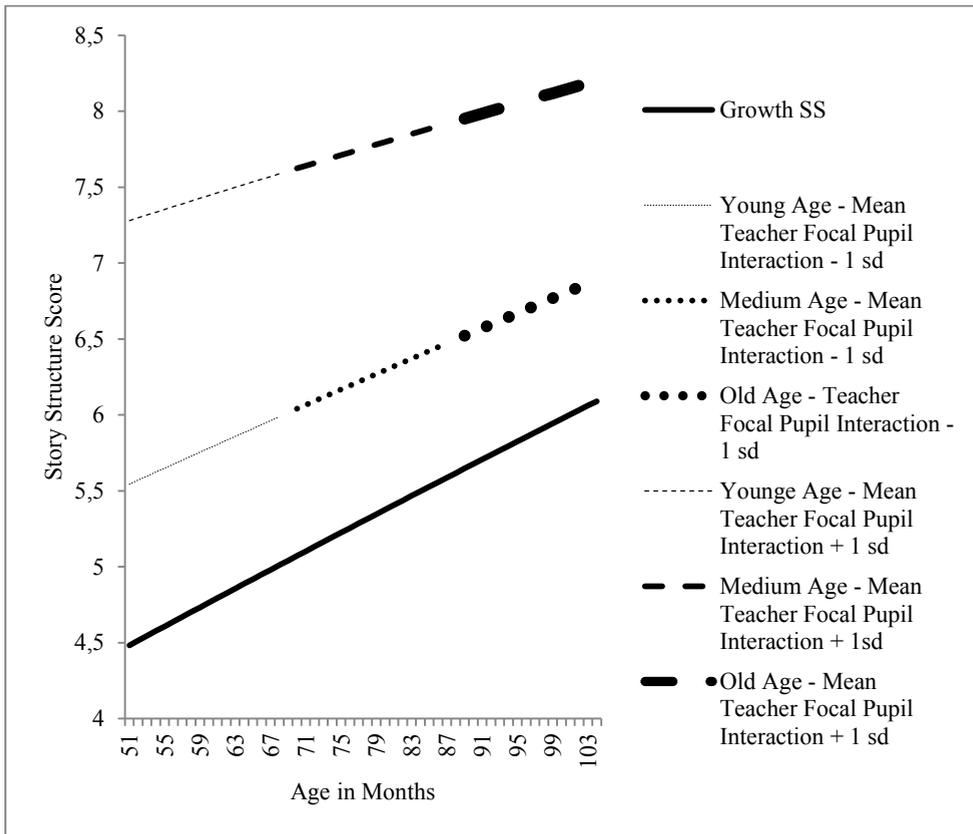


Figure 8.16: Development of Story Structure with Teacher-Focal Pupil Interaction as Explanatory Variable.

The addition of the other four variables, percentage of time spent on Language Activities and Mathematical Activities or time spent in Language Situations with Peers or Peer Interaction did not improve the general model of growth in SS score. This means that we could not show that the fact that pupils at DL2-schools were more likely to be engaged in Language Activities and Mathematical Activities compared to pupils at Mainstream schools had any significant impact on the pupils' development of Story Structure. Additionally, we

could not show that the fact that pupils at Mainstream schools were more likely to be engaged in Language Situations with Peers and Peer Interaction compared to pupils at the DL2-schools had any significant impact on the pupils' development of Story Structure either.

### ***Structural Complexity (SC)***

Like with SC and CLASS, the reliability for the General Development Model of SC and Snapshot (with 162 cases) was too low: .35. Hence, at least 65% of the observed differences do not relate to SC but can be seen as random noise (error of measurement) instead. Therefore, we refrained from building a General Development Model and did not investigate what the effects of the school learning environment are in this General Development Model of SC.

### ***Internal State Terms (IST)***

Like with IST and CLASS, the reliability for the General Development Model of IST and Snapshot (with 162 cases) was too low: .48. Hence, at least 52% of the observed differences do not relate to IST but can be seen as random noise (error of measurement) instead. Therefore, we refrained from building a General Development Model and did not investigate what the effects of the school learning environment are in this General Development Model of IST.

## **8.4 Summary and conclusion**

The goal of this chapter was to answer the final research question: To what extent do differences in the characteristics of the school learning environment during the first year after arrival relate to receptive vocabulary development and to narrative ability development of newly arrived migrant kindergarteners during the first two-and-a-half years of schooling in the Netherlands, in school types 1 and 2?

The pedagogical practices in the school learning environment have been studied on two levels: with the focus on teacher behavior, see Chapter 6, and the experiences of the focal pupil, see Chapter 7. As reported in Chapter 6, the two variables that differed significantly between the school types with respect to teacher behavior were the specific dimensions of Positive Climate and Regard for Student Perspectives. Pupils at Mainstream schools were more likely to have a teacher that provided a more positive climate and took into account students perspectives than pupils in DL2-schools.

As reported in Chapter 7, when we looked at how the focal pupils' experiences differed based on whether they were enrolled in a mainstream Dutch school after arrival to the Netherlands versus whether they were enrolled in a program with intensive support in the Dutch language, we found that there was a significant difference in the amount of time they spent in six types of activities, interactions, and language situations. Pupils enrolled in

a DL2-school spent more time doing mathematical and language activities, found themselves more often in balanced language situations, and more often interacted directly with the teacher, whereas pupils enrolled in a Mainstream school more often found themselves using language and interacting with peers.

In the present chapter we wanted to determine whether these eight differences that were found between DL2-schools and Mainstream schools had any effect on the language development of the pupils. Since we focused in this dissertation on the differences between the two school types, we only considered the variables that differed between the school types. Table 8.22 gives an overview of all the outcomes. Due to the low reliability of the growth models for SC and IST we refrained from analysis of the influence of pedagogical practices on these two narrative measures. Of the 50 remaining possible relations, 11 were significant.

Table 8.22: Summary Overview of Significant Learning Environment Characteristics on the Receptive Vocabulary and Narrative Ability development.

Explaining variable	Receptive vocabulary		Narrative Ability			
	PPVT		Micro-structure		Macro-structure	
			NDW	GIS	MLR	SS
Teacher Behavior	Positive Climate	+ <sup>b</sup>				- <sup>b</sup>
	Regard for Student Perspectives					
Focal Pupils' Experiences	Math Activities			- <sup>a</sup>		
	Language Activities		+ <sup>b</sup>			
	Language Situations with Peers		- <sup>b</sup>			
	Balanced Language Situations			+ <sup>a</sup>		+ <sup>a</sup>
	Peer Interaction		+ <sup>a</sup>		- <sup>b</sup>	
	Teacher- Focal Pupil Interaction			+ <sup>a</sup>		+/-

+ indicates a positive effect; - indicates a negative effect; +/- indicates a positive main effect but a negative interaction effect.

<sup>a</sup> Only the inclusion of the variable as a main effect was significant.

<sup>b</sup> Only an interaction effect with Age was found.

#### **8.4.1 The influence of the school learning environment: focusing on teacher behavior**

With our analysis we could show that the result found in Chapter 6 about Positive Climate – that Mainstream schools teachers score higher on the dimension Positive Climate compared to DL2-school teachers – influenced the PPVT score and the MLR score. A positive interaction between Age and Positive Climate was found for the receptive vocabulary scores, but a negative interaction between Age and Positive Climate was found for the MLR score. Meaning that a positive classroom climate provided by the teacher has a stronger effect on older pupils' receptive vocabulary scores than on younger pupils' scores, while a positive classroom climate provided by the teacher has a stronger effect on younger pupils compared to older pupils regarding the Measure of Lexical Richness scores. However, since the main effects of Positive Climate were not significant (due to large standard errors) it is difficult to interpret these effects.

In Chapter 6 on teacher behavior we found that that Mainstream schools teachers scored higher on the dimension Regard for Student Perspectives compared to DL2-school teachers. However in our analysis in Chapter 8 we did not find that Regard for Student Perspectives was a significant predictor of the difference in any of the language measures.

We only found two out of the ten differences in teacher behavior in the different school types to have a significant effect on our language measures. This might be explained by the fact that Positive Climate and Regard for Student Perspectives are not the main aspects of teacher behavior that influence language learning. Dimensions in the domain of Instructional Support are more often predictors of differences in development. However, since we focused in this dissertation on differences in school type we only considered these two variables of teacher behavior in our analysis. For more information on the influence of teacher behavior on second language development in general the data could be analyzed again including all the other dimensions.

#### **8.4.2 The influence of the school learning environment: focusing on focal pupils' experiences**

In Chapter 7 we found that pupils at DL2-school were more engaged in Mathematical Activities compared to pupils at Mainstream schools. In our analysis in Chapter 8 we found that the percentage of time spent in Mathematical Activities had a significant negative effect on the Guiraud Index Score. Pupils who spent more time on Mathematical Activities used less diverse language when telling stories.

Furthermore, in Chapter 7 we found that pupils at DL2-school were more often engaged in Language Activities compared to pupils at Mainstream schools. The percentage of time spent in Language Activities was found to influence the Number of Different Words, but it was an interaction between Age and the percentages of Language Activities. The effect of the percentage of time spent in Language Activities on NDW was larger for older pupils compared to younger pupils. However, since the main effect of Language

Activities was not significant (due to large standard errors) it is difficult to interpret this effect.

Additionally, in Chapter 7 we found that pupils at Mainstream schools were engaged in more Language Situations with Peers than pupils at Mainstream schools. The percentage of Language Situation with Peers was found to influence the NDW score, but it was an interaction between Age and the percentage of time spent in Language Situations with Peers. The effect of Language Situations with Peers on NDW was larger for younger pupils than for older pupils. However, since the main effect of Language Situations with Peers was not significant (due to large standard errors) it is difficult to interpret this effect.

In Chapter 7 we found that pupils at DL2-schools were engaged in more Balanced Language Situations. In Chapter 8 we found that the percentages of Balanced Language Situations influenced the GIS and the Story Structure. Pupils who spent more time in Balanced Language Situations use more diverse language when telling stories compared to pupils who spent less time in Balanced Language Situations. Furthermore, a higher percentage of time spent in Balanced Language Situations resulted in higher scores of SS.

In Chapter 7 we found that at Mainstream schools there are higher percentages of Peer Interaction. The percentages of time spent in Peer Interaction seem to influence the NDW as well as the MLR. Pupils who have more Peer Interaction have a higher NDW when telling stories compared to pupils engaged in less Peer Interactions. The effect of Peer Interaction on MLR seemed to be a negative interaction between Age and the percentage of time spent in Peer Interactions: the effect of Peer Interaction was larger for older pupils compared to younger pupils regarding the MLR. However, since the main effect of Peer Interactions was not significant (due to large standard errors) it is difficult to interpret this effect.

Finally, in Chapter 7 we found that at DL2-schools there are higher percentages of time spent in Teacher-Focal Pupil Interaction. These different percentages of time spent in Teacher-Focal Pupil Interaction influence the GIS and the SS, just like Balanced Language Situations. Pupils who spent more time in Teacher-Focal Pupil Interaction used more diverse language when telling stories compared to pupils who spent less time in Teacher-Focal Pupil Interaction. Furthermore, a higher percentage of time spent in Teacher-Focal Pupil Interaction also seemed to result in higher scores of SS, but the effect of the percentage of time spent in Teacher-Focal Pupil Interaction is larger for younger pupils than for older pupils regarding SS scores.

Balanced Language Situations and Teacher-Focal Pupil Interaction are quite similar, since they both involve teachers and in many situations with Teacher-Focal Pupil Interaction there was a Balanced Language Situation. However, the fact that they have an influence on microstructure as well as on macrostructure might indicate that interaction with the teacher has more influence on the second language of a newly arrived migrant kindergartener than interaction with his or her peers. The variables Language Situations

with Peers and Peer Interaction did not have similar effects on any of the language measures.

The fact that there were many non-significant variables and that two language measures could not be taken into the analysis of the school learning environment at all is challenging, however it makes the significant variables stand out. The significant variables however are not straightforward and do not justify saying that it is better for the second language development of young newly arrived migrant pupils to start in a DL2-school or in a Mainstream school. We cannot justify this since for example the variable Language Activities, more present in a DL2-school, had an effect on the language measure NDW, while another variable, Positive Climate for instance, more present in a Mainstream school, had an effect on the languages measure for receptive vocabulary. Further research should, first of all, focus on these effects in the longer term in order to see which variable is more important for second language development. Second, researchers should focus on how to create a learning environment with all these important variables present for newly arrived migrant pupils to have a supportive school learning environment irrespective of school type.



## CHAPTER 9

---

### **General discussion and conclusions**

---

## **9.1 Introduction**

In this final chapter, we start (Section 9.1) with an overview of the most important results for the two areas under investigation: the second language development of newly arrived migrant kindergarteners and the quality of their learning environment. Following this, the link between the second language development and the school learning environment and is summarized (Section 9.2.). We then continue with the discussion of our results and how they help answer our main research question and research sub-questions (Section 9.3). We also dedicate a section (Section 9.4) to the implications our findings have for the education of newly arrived migrant pupils. Subsequently, we discuss the limitations of our study and we give directions for future research (Section 9.5). We end this final chapter with some remarks and a final conclusion (Section 9.6).

## **9.2 Overview of the results**

### **9.2.1 Receptive vocabulary development**

In Chapter 4 the focus was about receptive vocabulary development. The receptive vocabulary of 42 newly arrived migrant kindergarteners was measured four times over two-and-a-half years with the Peabody Picture Vocabulary Test for Dutch (PPVT; Schlichting, 2005). The main findings were as follows:

- Over the two-and-a-half years there was an increase of receptive vocabulary; however, this development was quadratic: it increased more quickly in the beginning but then it leveled off.
- The more Exposure to Dutch at School, the higher the scores on the receptive vocabulary test. However, the amount of time spent attending a Dutch School had a greater effect on the receptive vocabulary size of younger pupils than that of older pupils.
- When the receptive vocabulary scores of pupils who attended DL2-schools were compared with the receptive vocabulary scores of pupils who attended Mainstream schools, no difference was found.

### **9.2.2 Narrative ability development.**

In Chapter 5 the focus was on narrative ability development. The narrative ability of 42 newly arrived migrant kindergarteners was measured four times over two-and-a-half years with the Multilingual Assessment Instrument for Narratives (MAIN; Gagarina et al., 2012). We looked at narrative ability on two levels: the microstructural level and the macrostructural level. For microstructure, we measured lexical diversity (Number of Different Words and Guiraud Index Score) and lexical richness (Measure of Lexical

Richness). For macrostructure, we looked at Story Structure, Structural Complexity, and Internal State Terms. The main findings were as follows:

- Age was significant for five of the six measurements of narrative development.
  - For Number of Different Words, the Guiraud Index Score, and Internal State Terms the development was linear; each month a child grew older his or her stories contained more diverse words and also contained more terms about mental states and feelings of the story characters.
  - For the Measure of Lexical Richness and Story Structure there was growth over time, but the rate of growth is greater for younger pupils than for older pupils.
- Exposure to Dutch at School had an effect on all language measures except for Internal State Terms.
  - For Number of Different Words, Guiraud Index Score, Measure of Lexical Richness, and Story Structure there was a fixed effect of exposure to Dutch at School: the longer a pupil attended a Dutch school, the more lexically diverse, lexically rich, and structured his or her storytelling.
  - For Story Structure there was an interaction between age and exposure to Dutch: the influence of exposure to Dutch at school was larger for younger pupils than for older pupils.
- Comparisons of scores on measures of narrative ability of pupils who attended DL2-schools with the scores of pupils who attended mainstream schools showed no difference, except for an interaction effect between Age and Number of Different Words: for younger pupils the effect of educational facility on lexical diversity was larger than for older pupils.

### **9.2.3 The school learning environment**

#### **Focusing on teacher behavior**

In Chapter 6 the aim was to investigate the pedagogical practices in the classroom focusing on teacher behavior. To investigate this, the Classroom Assessment Scoring System (CLASS; Pianta, La Paro, & Hamre, 2008) was used to measure the quality of several dimensions of teacher-child interactions in the different classrooms. The main findings were:

- Overall, the teachers from both school types in this study showed a pattern in their scores over the ten dimensions of CLASS which was in accordance with results of previous national and international studies<sup>23</sup>: higher scores on the Emotional Support

---

<sup>23</sup> cf. Mashburn, Justice, Downer, & Pianta, 2009; Pakarinen et al., 2010; Leseman & Veen, 2016; Henrichs & Leseman, 2016; Veen et al., 2017.

domain and lowest score on the Instructional Support domain. On the one hand, teachers created a positive environment for pupils, but on the other hand, teachers did not excel in promoting pupils' higher order thinking or providing feedback and rich language input. Despite this comparable pattern, the scores on the CLASS in our sample were lower overall in comparison the other national and international studies on kindergarten teachers.

- Mainstream school teachers scored on average more highly on the dimension Positive Climate than teachers at DL2-schools. The Mainstream school teachers were able to build better relationships with their pupils: there was more shared enthusiasm, affection and respect in the classroom.
- Mainstream school teachers scored on average more highly on the dimension Regard for Student Perspectives than teachers at DL2-schools. They provided more room for the pupils' initiatives and leadership. They also gave them more responsibility compared to teachers in DL2-schools.

### **Focusing on focal pupils' experiences**

In Chapter 7, the focus was on pedagogical practices in the classroom from the point of view of the focal pupils' experience. A snapshot method was used to investigate the Activities, Interactions, and Language Use of and around the focal pupils. The main findings were:

- *Activities*: Taken together routines, language, literacy, and mathematic activities, cover 69% and 60% of a day in DL2-schools and Mainstream schools, respectively.
  - Focal pupils at DL2-schools were engaged in more Language Activities compared to pupils at Mainstream schools.
  - Focal pupils at DL2-schools were engaged in more Mathematic Activities compared to pupils at Mainstream schools.
- *Interaction*: on average, about one-third of the time the pupils had no interaction with a peer or a teacher. Another third of the time, they interacted with their peers.
  - Focal pupils at Mainstream schools had more interactions with their peers than focal pupils at DL2-schools.
  - Focal pupils at DL2-schools had more interaction with their teachers than focal pupils at a Mainstream school.
  - Teacher interactions were mainly defined as “didactical” instead of “scaffolding.”

*Language use*: If the focal pupils were engaged in language situations, these verbal interactions were usually with a peer (on average 26% of the time for pupils at DL2-schools and 42% for pupils at Mainstream schools) and usually in Dutch.

- Focal pupils at Mainstream schools were engaged significantly more often in language situations among only peers compared to pupils at DL2-schools.
- Focal pupils at DL2-schools were significantly more engaged in balanced situations between teacher and focal pupils compared to pupils at Mainstream schools.
- No differences in language use between pupils at DL2-schools and Mainstream schools were found in this study. Teachers, peers, and focal pupils most of the time used simple language, mostly one-word “sentences.”
- Overall, most of the time the nature of the language use was qualified as simple language. Complex language use, including long sentences, precise referencing, and infrequent words with explanations, was hardly observed.

#### **9.2.4 Linking characteristics of the learning environment to language development**

In Chapter 8, the outcomes on the pedagogical practices in the classroom from Chapter 6 and 7 were related to the language measures from Chapter 4 and 5. The main results were as follows:

##### **Focusing on teacher behavior**

- The higher scores for Positive Climate of Mainstream school teachers compared to DL2-school teachers had an effect on two language measures:
  - A positive interaction between Age and Positive Climate was found for the PPVT score: when teachers created a more positive atmosphere in the classroom, it affected receptive vocabulary size of older pupils more than that of younger pupils.
  - A negative interaction between Age and Positive Climate was found for the Measure of Lexical Richness score: when teachers created a more positive atmosphere in the classroom this had a larger effect on the Measure of Lexical Richness of younger pupils compared to older pupils.
  - Large standard errors refrain us from interpreting these two interaction effects.

##### **Focusing on focal pupils’ experiences**

- The higher percentage of time spent in Mathematical Activities at DL2-school compared to Mainstream schools seems to only have a negative effect on the Guiraud Index Score. Pupils in classes with higher percentages of Mathematical Activities used less diverse language when telling stories.
- The higher percentages of Language Activities at DL2-schools compared to Mainstream schools only seem to have an effect on the Number of Different Words, but it is an interaction between Age and percentages of Language Activities. The effect

of the percentage of time spent in Language Activities on the use of more words per story was larger for older pupils compared to younger pupils.

- The higher percentage of time spent in Language Situations with Peers at Mainstream schools compared to DL2-schools has an effect on the Number of Different Words score, but it is an interaction between Age and percentage of time spent in Language Situations with Peers. The effect of percentage of time spent in Language Situations with Peers on the use of more words per story was larger for younger pupils than for older pupils.
- The higher percentage of time spent in Balanced Language Situations at DL2-schools compared to Mainstream schools seems to have an effect on the Guiraud Index Score and on the Story Structure. Pupils who spent more time in Balanced Language Situations used more diverse language when telling stories compared to pupils who spent less time in Balanced Language Situations. Furthermore, a higher percentage of time spent in Balanced Language Situations also results in more complete stories.
- The higher percentages of time spent in Peer Interaction at Mainstream schools compared to DL2-schools has an effect on the Number of Different Words as well as on the Measure of Lexical Richness.
  - Pupils who have more Peer Interaction used more different words when telling stories compared to pupils engaged in less Peer Interactions.
  - The effect of Peer Interaction was larger for older pupils compared to younger pupils regarding the richness of the words.
- Large standard errors refrain us from interpreting the above interaction effects when there was no main effect.
- The higher percentages of Teacher-Focal Pupil Interaction at DL2-schools has an effect on the Guiraud Index Score and on the Story Structure.
  - Pupils who spent more time in Teacher-Focal Pupil Interaction used more diverse language when telling stories compared to pupils who spent less time in Teacher-Focal Pupil Interaction.
  - Furthermore, a higher percentage of time spent in Teacher-Focal Pupil interaction also results in more complete stories. This effect of the percentage of time spent in teacher-focal pupil interaction on the way stories were told was larger for younger pupils than for older pupils.

### **9.2.5 Summary of the study**

To sum up the outcome of the total study: for six out of seven language measures we were able to model development. Newly arrived migrant kindergarteners' scores on the use of emotional terms and (two) measures of lexical diversity increased linearly over the two-and-a-half years of the study. The same pupils' scores on measures of receptive vocabulary, lexical richness, and the use of story structure showed non-linear development: in the

beginning there is faster growth that eventually levels off over the two-and-a-half year period.

Besides information on the second language development of newly arrived migrant pupils as a group, we were interested in the differences in development between pupils from DL2-schools and Mainstream schools. There were no differences in receptive vocabulary development between pupils from DL2-schools and Mainstream schools when the general variable Educational Facility was taken into account. The only difference in narrative ability development between the two school types was found for NDW. The effect of Educational Facility was stronger on lexical diversity for younger pupils compared to older pupils, although we refrain from interpreting this interaction effect due to large standard errors.

When we looked closer at the characteristics of the learning environment we found several differences in the learning environment between the two school types, DL2-schools and Mainstream schools. All in all, it could be stated that the pedagogical practices at DL2-schools were more teacher-centered compared to Mainstream Schools. The activities were more teachers-led and therefore the interactions in the class were more between teacher and pupils than among peers. At Mainstream schools the pedagogical practices were more pupil-centered with higher scores on taking into account students perspectives and initiatives and more interaction among peers.

Some of the differences between the two school types had an effect on the second language development of newly arrived migrant pupils. Nevertheless, it remains difficult to say whether it is better for a newly arrived migrant kindergartener to enter one school type over the other, since some favorable pedagogical practices for language development were more present at one school type and other favorable pedagogical practices were more present at the other school type. Likewise, schools of one type exhibited differences among them with regard to pedagogical practices, yet schools of different types showed similarities. Finally, due to large standard errors, we did not interpret some of the interaction effects since this would make the interpretations unreliable. These effects need to be confirmed with a larger sample. In the next Section we will discuss these outcomes more thoroughly.

### **9.3 General discussion**

This chapter continues with a general discussion of the results. The main goal of this dissertation was to answer the following research question:

*To what extent do pedagogical practices contribute to the second language development of newly arrived migrant kindergarteners in the first two-and-a-half years after arrival in the Netherlands?*

In order to answer the main research question, the language development and the school learning environment of newly arrived migrant kindergarteners were analyzed, before the relation between the two was investigated. Throughout this dissertation the following five sub-questions (RQ's) were addressed:

- (RQ1) How does the *receptive vocabulary* (in Dutch) of newly arrived migrant kindergarteners develop during the first two-and-a-half years of schooling in the Netherlands in relation to school type?
- (RQ2) How does the *narrative ability* (in Dutch) of newly arrived migrant kindergarteners develop during the first two-and-a-half years of schooling in the Netherlands in relation to school type?
- (RQ3) What are the differences in characteristics of the school learning environment *regarding teacher behavior* between DL2-schools and Mainstream schools?
- (RQ4) What are the differences in characteristics of the school learning environment *from the point of view of focal pupils' experiences* between DL2-schools and Mainstream schools?
- (RQ5) To what extent do *differences in the characteristics of the school learning environment* during the first year after arrival relate to receptive vocabulary development and to narrative ability development of newly arrived migrant kindergarteners during the first two-and-a-half years of schooling in the Netherlands?

The discussion of these five sub-questions is divided into four sub-sections. In the first sub-section (Section 9.3.1) the development of both receptive vocabulary and narrative ability is discussed, including the influence of the child characteristics (RQ1, RQ2). The second sub-section (Section 9.3.2) concerns the school learning environment of newly arrived migrant kindergarteners (RQ3 and RQ4) and the third sub-section (Section 9.3.3) zooms in on the relationship between the school learning environment and the development of the different language measures (RQ5).

### 9.3.1 Second language development

The answer to our first main sub-question is that all pupils show growth in receptive vocabulary: The older the pupils, the higher their receptive vocabulary scores. However, the growth was quadratic: in the beginning there is a steeper growth which then levels off. Receptive vocabulary development was influenced by the length of time exposed to Dutch. Since we looked at the development of only newly arrived pupils we cannot say anything about them closing the gap in receptive vocabulary scores with their first language learning peers. We did however show that there was progress in their receptive vocabulary scores.

The different microstructural features and macrostructural components of narrative ability also showed an increase over time, except for the complexity of the stories, for

which no developmental trajectory could be estimated. There were however, differences in developmental trajectories. The microstructural features improved linearly, while the macrostructural features showed a decrease in growth-rate. This was expected since Schneider, Hayward, and Vis Dubé (2006) also showed a reduction in growth rate for macrostructure around the age of 7. Contrary to Maviş, Tunçer, and Gagarina, (2016), no ceiling effect was visible in our data.

### **9.3.2 The school learning environment**

Focusing on teacher behaviour, we found many similarities in teacher behavior between the two school types, which means that pupils will encounter teachers with similar teacher behavior irrespective of school type. An explanation of this could be that teachers at both school types might have experienced similar teacher education or similar sources of professionalization, but this needs to be confirmed in a follow-up study. These many similarities also suggest that teachers at a DL2-school do not stand out: this is strange since they are supposed to be experts in education for second language learners, given that they have this specific student population in their class. Unfortunately, the DL2-school teachers do not seem to take the opportunity to provide the newly arrived migrant kindergarteners with richer, complex language supported with scaffolding.

It is especially remarkable that the DL2-school teachers did not score highly on the variable Language Modeling, which measured how rich the language is the teacher uses. One interpretation would be that mainstream school teachers, at least in our sample, are well aware of the specific needs of newly arrived migrant pupils and therefore show similar behavior to DL2-school teachers. Another interpretation is that the CLASS dimensions could not capture the distinct behavior of DL2-school teachers. However, the most likely interpretation is that in general teachers have difficulty scoring highly on the Instructional Support domain of the CLASS of which language modeling is part, which was also shown in previous international and national studies (cf. Mashburn, Justice, Downer, & Pianta, 2009; Pakarinen et al., 2010; Leseman & Veen, 2016; Henrichs & Leseman, 2016; Veen et al., 2017). In recent publications teachers are educated in and encouraged to use strategies such as scaffolding and comprehensible input in the education of newly arrived migrant pupils (see for example Smits & van Koeven, 2016 and Schrijfgroep LPTN, 2017).

Focusing on focal pupils' experiences, we found differences between the two school types in focal pupils' experiences. As might be expected in a class for children with a Dutch language gap, there were significantly more language activities at DL2-schools compared to Mainstream schools, as well as more teacher-focal pupil interactions and therefore more balanced language situations. Despite the greater focus on language learning in DL2-school, there was not more teacher instruction, not more teacher talk, or less play at DL2-schools compared to Mainstream schools. Interestingly, pupils at Mainstream schools were more engaged in peer interaction and consequently in language situations with peers

compared to pupils at DL2-schools – situations that could positively impact the pupils' language development. Nevertheless, teachers and peers in both school types seem to provide the focal pupils with mostly simple language. Apparently, this was the only way that teachers and peers were able to communicate with the newly arrived migrant pupils during our observations in the first year after arriving. This is in line with the observations of Cekaite and Aronsson (2014) and Tabors and Snow (1994).

Overall, the pattern seems to be that pupils at DL2-schools received more input from their teachers than from their peers. For pupils at Mainstream schools this pattern seems to be reversed. The pedagogical practices at DL2-schools seemed to be more teacher-centered and at Mainstream schools more pupil-centered. An explanation for this difference could be that teachers at DL2-schools feel that they need to be in control in order to reach certain goals for learning by the newly arrived pupils, they want to implicitly teach new words. At Mainstream schools teachers might rely more on implicit learning via peers.

### **9.3.3 The influence of the school learning environment on second language development**

Focusing on teacher behaviour, only one variable of teacher behavior that was significantly different between the two school types influenced the language development of newly arrived migrant pupils. The significant interaction effect between Age and Positive Climate on both receptive vocabulary and lexical richness could suggest that a classroom with more positive relationships between teachers and pupils indeed influences vocabulary development. However, it is not possible to conclude from our data what that effect was. Furthermore, the result is confusing because the effect of Positive Climate seemed to be larger for older pupils on the receptive vocabulary scores, while the effect of Positive Climate was larger for younger pupils for lexical richness. Studies with data with more variation in teacher scores on Positive Climate might shed light on the influence of Positive Climate.

Focusing on focal pupils' experiences, the results from the analysis of the influence of focal pupils' experiences on the different language measures are ambiguous which makes it difficult to draw conclusions. All aspects of microstructure, lexical diversity, and lexical richness were influenced by one or three variables concerning the focal pupils' experiences. It is impossible to find a pattern, since all six learning environment variables are significant for at least one aspect of microstructure. Peer Interaction is the only variable that influenced two microstructural measures, of which one as a main fixed effect and the other as an interaction effect.

Of the macrostructural components, it was only possible to build a reliable model for Story Structure. The variables from the focal pupils' experiences that influenced this macrostructural component were similar to the variables that influence the microstructural feature for lexical diversity, the Guiraud Index Score. Both variables included the

involvement of the teachers. Therefore, it seems that the influence of the teacher on language development is stronger since the interaction and language use variables in which the teacher is involved caused a main effect three times, and one time an interaction effect with a significant main effect – on both microstructural and macrostructural level.

In sum, the results show that besides the child characteristics Age and Exposure to Dutch, characteristics of the school learning environment also have an influence on second language development of newly arrived migrant pupils. However, it was not possible to determine what the most ideal learning environment is, since we could not create a comprehensive model. Without including both internal child characteristics, such as, Age and Exposure, and external child characteristics, such as, teacher behavior, in one model, it is impossible to determine what the ideal environment is.

## **9.4 Implications for educational practice**

In the present study, the learning environment of educational facilities attended by newly arrived migrant pupils in the Netherlands has been investigated. We found a pattern of teacher behavior that was in line with Henrichs and Leseman (2016; higher scores on Emotional Support and lowest scores on Instructional Support); however, overall, we found the scores of the teachers on the different dimensions of classroom quality to be lower in our study than in theirs. Therefore, it is recommendable that schools of all types reconsider their way of evaluating teacher behavior and especially focus on improving teacher skills that can enhance the Instructional Support domain of CLASS. That is, to improve teacher practices regarding, for example, how to ask different questions (“why,” “how,” and open-ended questions), to use scaffolding, to give feedback, to stimulate sustained conversations, and, for example, how to provide rich language examples. Based on Hamre et al. (2012) and Henrichs and Leseman (2016) process quality in classrooms can be improved when schools invest in the professional development of the teachers. When teachers receive concrete feedback with tips on their performance, improvements can be made.

We investigated the school learning environment for newly arrived migrant pupils in their first year after arrival in the Netherlands, but naturally the school learning environment continues to be important after that first year. Despite this, the Dutch government, at this time, only finances extra support in the first year after arrival for newcomers (or the first two years for asylum seeker children), even though it takes at least five years (Cummins, 1981) to approach grade norms. We can only speculate, but it might be that due to a lack of financial resources not all teachers at mainstream schools can continue to foster second language development and provide the resources necessary to keep migrant pupils engaged in their education. Researchers must disseminate their knowledge and work together with schools to guarantee that newly arrived migrant pupils receive education on a level that meets their cognitive possibilities, interests, and talents

during their whole school career. An example of such a recent cooperation is the project EDINA<sup>24</sup>.

In order for schools to foster second language development, it is highly recommendable that all teachers in primary school have background knowledge about second language learning and that they are familiar with beneficial second language pedagogies. Most importantly, all teachers should be aware that learning a new language is more than reaching conversational fluency, which takes about two years; it also includes learning academic aspects of the new language, which takes 5–7 years (Cummins, 1981). Thus, after the (sometimes separate) first year, second language learners need guidance for at least four more years (e.g., Thomas & Collier, 2002). The too often concentrated and isolated knowledge about second language pedagogies at DL2-schools should be transferred to mainstream schools, and exchanges about these practices should be encouraged. Researchers could play a role in facilitating this knowledge dissemination.

In Chapter 1 (Section 1.4.1) we explained that in the Netherlands in almost all primary schools the main language of instruction is Dutch and that few teachers take into account the multilingual repertoires of the pupils in the curriculum. In our study we observed that other language than Dutch were used by the pupils, but not by the teachers. International studies (e.g., Cummins, 2007; García, 2009; García & Wei, 2014; Lewis, Jones, & Baker, 2013; Manyak, 2004) in fact, show that one particularly powerful tool to facilitate integration into the new educational and social environment is to maximize the school's engagement in pupils' home languages. These studies showed that promotion of the home language helps to “prepare diverse pupils to negotiate successful life pathways in our increasingly diverse society” (Manyak, 2004, p. 12), as well as to support pupils' literacy development. In the Netherlands, policymakers and school boards may be reluctant to embrace home languages as part of the educational practices, particularly since the abrogation of financial support to education in migrant languages in 2004, due to criticisms of multicultural approaches (see Winter, 2010, for a more detailed historical overview of this shift in policy).

In our study, and probably nation-wide in the Netherlands, teachers seem to be unaware about practices of how to include other languages than Dutch in the curriculum. That this is changing is evidenced by a project in the north of the Netherlands where the holistic model for multilingualism in education is currently being implemented (Duarte & van der Meij, 2018). Some teachers in our study indicated that they cannot understand the home languages of the pupils and thus are not able to include these in the curriculum. However, even when teachers themselves cannot understand or speak those languages they can implement home languages in their pedagogical practices. All teachers can be curious and welcoming towards the home languages of the pupils and to encourage the use of these

---

<sup>24</sup> Education of international newly arrived migrant pupils; <https://edinaplatform.eu/en>.

languages among the pupils in the school (a.o., Gielen & Işçi, 2015; Groothoff, in press; Slembrouck, Van Avermaet, & Van Gorp, 2018).

## **9.5 Limitations and future research**

This study comes with several limitations. The following are discussed in this section: the compilation of our group of participants, the instruments and followed procedures that were used, and finally, the incomplete use of the communicative competence theory. In this discussion we will also give suggestions how to overcome these limitations in future research.

### **9.5.1 Participants**

In Chapter 3, which concerned the participants in this study, it was already mentioned that here was skewness in the distribution of the participants. The lack of differences in score between the two school types could have been caused by the fact that we had only a small and skewed population. A better representation of DL2-schools and classes and more pupils attending a Mainstream school would have been preferable. This would have only been possible if the research had been undertaken by a larger research groups since it is impossible for one researcher alone to execute all the observations and assessments necessary to test a larger.

In this study the influencing factors Age, Exposure to Dutch at School, and Educational Facility were studied, but there are many more factors that might play a role in the language development of newly arrived migrant pupils. For example, detailed information on their first language development, use and proficiency, socioeconomic status, maternal education, and quantity and quality of the input of the second language could have been included (for a list of determinants on language development, see Sorensen Duncan, 2017). Furthermore, additional information on motivation, language aptitude, learning style, learning strategies, and other general cognitive ability, preferably language-independent, might also help to explain individual differences.

In the present study the number of background variables was kept to a minimum because we did not want to burden the teachers and parents with interviews or questionnaires; teachers might not be willing to participate in our study due to their already heavy workload and to interview parents of newly arrived migrant pupils requires a lot of time and preferably interpreters in numerous home languages. Finally, the participants in the present study were very young and therefore not a reliable source of information themselves. Thus, since it is a challenge to gather information on every possible influencing factor via parents and teachers, our recommendation for future research in this area would be to additionally apply special tools for measuring variables such as motivation and language aptitude for this specific group of young learners in order to make them informants on these influencing factors themselves.

### **9.5.2 Instruments and procedure**

Since we did not find a clear-cut effect of school type in our study on our language measures: being in a segregated facility did not emerge as the better option for our language measures, it might be better for all pupils to start in an inclusive facility. However, being in a segregated facility at the start might have positive effects on other outcomes than language measures. For example, it could be that to start in a segregated facility where everybody is learning the new language helps pupils feel safer in the beginning and supports their socio-emotional development. Therefore, we recommend also including the evaluation of the well-being of the newly arrived migrant pupils in future research and the investigation of the long-term effects this has on their development.

With our data we have shown how the narratives of 42 newly arrived migrant pupils in their second language developed. However, if we had been able to assess the participants in their L1 as well, this would have provided us with the possibility to see whether or not these L1 stories would mirror the L2 stories. Furthermore, whether the L1 stories would vary based on cultural context of the L1. Although the creators of the MAIN claim that the metric is culturally appropriate, storytelling conventions are in fact influenced by culture. “Substantial cultural differences influence the ways children structure their narratives. Such differences must be carefully distinguished from individual deficits” (McCabe & Rosenthal Rollins, 1994). Unfortunately, this went beyond the scope of this dissertation, however, researchers should be aware of cultural difference and therefore it is valuable to compare stories (from the same children) in different languages in more detail.

In previous research with the MAIN, authors, to our knowledge, never mentioned children’s use of communication strategies like acting out or gesturing. In our first assessments we noticed the participants doing this and we started videotaping the sessions. However, these video recordings could not be included in the present study but can be used in future research.

The CLASS is a fairly new instrument and recently Buell, Han, and Vukelich (2017) questioned the CLASS’ accountability as an evaluation tool. They suggest that CLASS results should be interpreted based on classroom composition (it seems for example that a higher percentage of boys influences CLASS results), seasonal fluctuations, and the context of the data collection. Since in the present study CLASS has been used only once with each teacher, and the assessments could not take place within one season, differences in CLASS scores between the teachers should be interpreted with caution.

In the present study the CLASS and a snapshot method were used once to investigate the first year of education of migrant pupils newly arrived in the Netherlands. Due to practical limitations including limited time, we could not include multiple observations per teacher or per student, especially not when all of the pupils were dispersed over 35 Mainstream schools in the final part of the study. Taking into account that learning

the academic aspects of a new language requires a period of 5–7 years (Cummins, 2008), it is important that future studies with newly arrived migrant pupils keep detailed track not only of the activities and interactions these pupils have on different occasions within that first year, but also in later years when they attend a mainstream school. This is especially so since it is expected that the interactions in the school learning environment change over time. For example Cekaite and Aronsson (2014) and Tabors and Snow (1994) found that during the first months at school a second language learner might only experience basic interaction.

Furthermore, we knew that by using the snapshot method the way we did, important information was lost since it is an on the spot method without recordings of the actual interactions and language use on tape. For our study we believe that this method was sufficient, but it is recommendable that future research will record the actual speech of and around the focal pupil and investigate how language learning can occur from the input a language learner receives per day, which is of course really complicated to achieve with current privacy regulations. This is however especially important since peer learning is an argument I often heard during the orientation phase of the present study used by mainstream schools for not having segregated facilities. Nowadays, in my new job as peripatetic language support advisor (in Dutch: *ambulant taalondersteuner*) I still hear teachers talk about the positive influence of peers on the second language development of newly arrived migrant pupils. However teachers do not seem to realize that peers can only be helpful to second language learners when these peers have developed specific capacities to be helpful for these learners. Researchers should make teachers aware of this and give them tools to support peers in this. Additionally, researchers should provide teachers with information on how to specifically target effective communicative attempts of the second learners themselves (Blum-Kulka & Snow, 2004).

The use of CLASS and the snapshot method provided us with valuable information about the pedagogical practices in classrooms of newly arrived migrant pupils. However, we feel that these observational tools could be more specified for working with second language learners and should focus more on specific pedagogical practices which are proven to be effective for the language learning of newly arrived migrant pupils, such as the functional use of the home languages in the classroom and the use of scaffolding to make the language input at the same time richer but also more comprehensible. Since the scores on Emotional Support are already in the higher ranges it might be interesting to focus more on the Instructional support domain of CLASS and give observers more indicators of good teacher behavior to evaluate this domain when second language learners part of the group.

In this dissertation the focus was on differences between the learning environment of DL2-schools and Mainstream schools. However, our data could also be interpreted without this distinction, thereby analyzing second language development of newly arrived

migrant pupils in general. Including all variables of CLASS and Snapshot in the analysis of our data might reveal more significant effects of the school learning environment for newly arrived migrant pupils irrespective of the school type. Due to time limitations we could not include these analyses in the current dissertation.

All in all, it was difficult to draw firm conclusions about the effects of the school learning environment based on the language measures since we could only include the variables one-by-one in our models. With a larger population it might be possible to build a model that includes multiple explanatory variables at the same time. Furthermore, it would be desirable to develop learning profiles that account for both internal child characteristics (like age, home language, and motivation) and external child characteristics (like educational program, language use of peers, and percentage of language activities) are taken into account, which might result in different teaching approaches for learners with various backgrounds.

### **9.5.3 Language development and communicative competence**

Even though it is acknowledged in this present study that looking at language development through a communicative competence lens will benefit language development assessment, this study could not cover all segments of communicative competence. Based on the data that were collected in the present study, all segments of communicative competence (linguistic competence, discourse competence, sociocultural competence, and strategic competence) could have been discussed. However, due to time limitations and the fact that this study data collection was carried out by one researcher alone, not all segments were covered. Nevertheless, we agree with Sridhar and Sridhar (2018) who encourage researchers to work towards a holistic theory of second language acquisition. They encourage not only focusing on cognitive and structural dimensions of language acquisition, but also incorporating social, functional, and multilingual perspectives.

In Chapter 2 the theory about communicative competence was explained and throughout the remaining of this dissertation the relation between the data and communicative competence has been discussed. The addition of a more communicative assessment instrument to the receptive vocabulary task – as was done by using the MAIN – already added value to the analysis of the language development. The data was collected in a more communicative context and children were freer to show their language abilities. Nevertheless, the data collected with the tool used to assess narrative ability, MAIN, can be a further source of information about especially the strategic competence of the pupils which in the current study has not focused on.

In previous research on MAIN, only rarely video-recording was used (Bohnacker, 2016). Nevertheless, in previous studies with the MAIN no observations were presented about acting out parts of the story, the use of gestures, or other strategies to improve comprehensibility of the story. Video-taped narratives provide more natural language and

the in-depth analysis can shed a light on more authentic behavior and include information about coping mechanisms that pupils use in every day communication (Schaefer & Bowyer-Crane, 2016). In the present study the narratives of the participants were video-recorded and with that a lot of information was gathered about the communication strategies the pupils used, their strategic competence, and how these developed over time. Unfortunately, it was beyond the scope of the present study to report the data from those observations.

Sociocultural competence was only briefly mentioned in Chapter 5 and 7 and was not a specific focus of the analysis in the current set-up of the study. Nevertheless, it is interesting to mention that the participants were almost never observed using other languages than Dutch, except for interactions with pupils with a shared language background. It would be interesting to study if these pupils would use code-switching during story generation in situations with an interlocutor who understands their home language.

The use of a narrative instrument such as the MAIN seems to be suitable to measure several parts of communicative competence. However, for the educational practice it might be difficult to use this instrument to its full extent because of its time consuming analysis. Just like in the present study we expect that teachers will select certain aspects that are most informative or most useful for the situation at hand. It would be recommendable for further research to focus on developing holistic ways to assess language development, but at the same time to take into account the limited time teachers have for analyzing results. Furthermore, suggestions should be made on feasible follow-up after the assessments: how can teachers stimulate different aspects of language development?

## 9.6 Concluding remarks

To conclude, this thesis contributes to a better understanding of second language development, specifically, the receptive vocabulary development and narrative ability development of newly arrived migrant kindergarteners. We have found that different language measures show different growth trajectories and that the child characteristics Age and Exposure to Dutch at School influence these trajectories differently. We also found factors in the school learning environment that influenced different aspects of the language development.

A beginning has been made with the investigation of the specific learning environment for newly arrived migrant kindergarteners, and this study can be a starting point for extended research on this matter, since at least two questions remain: (1) which factors in the school learning environment are the *most important* for the language development of newly arrived migrant kindergarteners? and (2) is it possible to ensure that those factors are part of the school learning environment of newly arrived migrant pupils *irrespective of the type of schooling?*

We encourage researchers to develop teaching materials and observational instruments in cooperation with the educational field. These materials and instruments should include good practices that answer to the specific needs of (young) newly arrived migrant pupils. Furthermore, such instruments should be used by teacher educators in their training for (prospective) teachers. In the next decade more and more classes will include second language learners and thus every teacher should know how to create the best learning environment for these children, how to measure language development appropriately, and how to take into account the language background of the child and other child characteristics that influence second language development.

---

## References

---

- Altman, C., Armon-Lotem, S., Fichman, S., & Walters, J. (2016). Macrostructure, microstructure and mental state terms in the narratives of English–Hebrew bilingual preschool children with and without specific language impairment. *Applied Psycholinguistics*, 37, 165–193. <https://doi.org/10.1017/S0142716415000466>
- Appel, R., & Lalleman, J. (1989). Het Nederlands van Turkse kinderen. *Gramma*, 13(2), 113–137.
- Appose, A., & Karuppali, S. (2018). Decoding the macrostructural form of oral narratives in typically developing children between 6-11 years of age: Using story grammar analysis. *Online Journal of Health and Allied Sciences*, 17(1). Retrieved from <https://www.ojhas.org>
- August, D., Carlo, M., Dressler, C., & Snow, C. (2005). The critical role of vocabulary development for English language learners. *Learning Disabilities Research & Practice*, 20, 50–57. <https://doi.org/10.1111/j.1540-5826.2005.00120.x>
- Bachman L. F., & Palmer A. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford, England: Oxford University Press.
- Baker, C. (2011). *Foundations of bilingual education and bilingualism* (5th. ed.). Bristol, England: Multilingual Matters.
- Berman, R. A. (1988). On the ability to relate events in narrative. *Discourse Processes*, 11, 469–497. <https://doi.org/10.1080/01638538809544714>
- Berman, R. A. (2001). Setting the narrative scene: How children begin to tell a story. In K. E. Nelson, & A. Aksu-Koc (Eds.), *Children's Language: Developing Narrative and Discourse Competence* (pp. 1–30). Mahwah, NJ: Lawrence Erlbaum Associates.
- Berman, R. A., & Slobin, D.I. (1994). Narrative structure. In R. A. Berman, & D. I. Slobin (Eds.), *Relating events in narrative: A cross linguistic developmental study* (pp. 39–84). Mahwah, NJ: Lawrence Erlbaum Associates.
- Bialystok, E., Luk, G., Peets, K., & Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition*, 13, 525–531. <https://doi.org/10.1017/S1366728909990423>

- Bishop, D. V. M., & Edmundson, A. (1987). Language-impaired 4-year-olds: Distinguishing transient from persistent impairment. *Journal of Speech and Hearing Disorders*, 52, 156–173. <https://doi.org/10.1044/jshd.5202.156>
- Blankenstijn, C., & Scheper, A. R. (2003). *Language development in children with psychiatric impairment*. Utrecht, the Netherlands: LOT. Retrieved from [https://pure.uva.nl/ws/files/3647880/33372\\_22.pdf](https://pure.uva.nl/ws/files/3647880/33372_22.pdf)
- Bliss, L. S., McCabe, A., & Miranda, A. E. (1998). Narrative assessment profile: Discourse analysis for school-age children. *Journal of Communication Disorders*, 31, 347–363. [https://doi.org/10.1016/S0021-9924\(98\)00009-4](https://doi.org/10.1016/S0021-9924(98)00009-4)
- Blom, E. (2010). Effects of input on the early grammatical development of bilingual children. *International Journal of Bilingualism*, 14, 422–446. <https://doi.org/10.1177/1367006910370917>
- Blom, E., & Boerma, T. (2016). Why do children with language impairment have difficulties with narrative macrostructure? *Research in Developmental Disabilities*, 55, 301–311. <https://doi.org/10.1016/j.ridd.2016.05.001>
- Blum-Kulka, S., & Gorbatt, N. (2014). “Say princess”: The challenges and affordances of young Hebrew L2 novices’ interaction with their peers. In A. Cekaite, S. Blum-Kulka, V. Grøver, & E. Teunal (Eds.), *Children’s Peer Talk. Learning from each other* (pp. 169–193). Cambridge, England: Cambridge University Press.
- Blum-Kulka, S., & Snow, C.E. (2004). Introduction: the potential of peer talk. *Discourse Studies*, 6, 291–306. <https://doi.org/10.1177/1461445604044290>
- Boerma, I. (2005). *Woordenschatontwikkeling van jonge Marokkaanse kinderen in Nederland* (Master’s thesis, Universiteit van Amsterdam, the Netherlands). Retrieved from <http://www.studiotaalwetenschap.nl/Bestanden/Scriptie%20I%20Boerma.pdf>
- Boerma, T., Leseman, P. P. M., Timmermeister, M., Wijnen, F., & Blom, E. (2016). Narrative abilities of monolingual and bilingual children with and without language impairment: Implications for clinical practice. *International Journal of Language and Communication Disorders*, 51, 626–638. <https://doi.org/10.1111/1460-6984.12234>
- Bohnacker, U. (2016). Tell me a story in English or Swedish: Narrative production and comprehension in bilingual preschoolers and first graders. *Applied Psycholinguistics*, 37, 19–48. <https://doi.org/10.1017/S0142716415000405>
- Bonifacci, P., Barbieri, M., Tomassini, M., & Roch, M. (2017). In few words: Linguistic gap but adequate narrative structure in preschool bilingual children. *Journal of Child Language*, 45, 120–147. <https://doi.org/10.1017/S0305000917000149>

- Bonnes Bowne, J., Yoshikawa, H., & Snow, C. E. (2017). Relationships of teachers' language and explicit vocabulary instruction to students' vocabulary growth in kindergarten. *Reading Research Quarterly*, 52, 7–29. <https://doi.org/10.1002/rrq.151>
- Botting, N. (2002). Narrative as a tool for the assessment of linguistic and pragmatic impairments. *Child Language Teaching and Therapy*, 18, 1–21. <https://doi.org/10.1191/0265659002ct224oa>
- Bronfenbrenner, U. (1979). *The Ecology of human development: Experiments by nature and design*. Cambridge, MA, and London, England: Harvard University Press.
- Bronfenbrenner U., & Morris, P. A. (1998). The ecology of developmental processes. In R. M. Lerner (Ed.), *Handbook of child psychology* (5th ed.), Vol. 1, (pp. 993–1028). New York, NY: Wiley.
- Bronfenbrenner U., & Morris, P. A. (2006). The bioecological model of human development. In R. M. Lerner (Ed.), *Handbook of child psychology* (6th ed.), Vol. 1, (pp. 793–828). Hoboken, NJ: Wiley.
- Brophy, J. (1986). Teacher influences on student achievement. *American Psychologist*, 41, 1069–1077. <https://psycnet.apa.org/doi/10.1037/0003-066X.41.10.1069>
- Bryant, D., Clifford, R., Early, D., Pianta, R., Howes, C., Barbarin, O., ... Burchinal, M. (2002, November). Findings from the NCEDL multi-state pre-kindergarten study. *Annual meeting of the National Association of the Education of Young Children*. New York, NY.
- Buell, M., Han, M., & Vukelich, C. (2017). Factors affecting variance in Classroom Assessment Scoring System scores: Season, context, and classroom composition. *Early Child Development and Care*, 187, 1635–1648. <https://doi.org/10.1080/03004430.2016.1178245>
- Bulters, I., & Vermeer, A. (2007). Leerkrachtvaardigheden en woordenschatonderwijs. Het effect van training op leerkrachtgedrag en leerlingprestaties. *Toegepaste Taalwetenschap in Artikelen*, 77, 9–19. Retrieved from <https://www.rezulto.nl/wp-content/uploads/2018/10/Leerkrachtvaardigheden-woordenschatonderwijs.pdf>
- Burchinal, M., Kainz, K., & Cai, Y. (2011). How well do our measures of quality predict child outcomes? A meta-analysis and coordinated analysis of data from large-scale studies of early childhood settings. In M. Zaslow (Ed.), *Reasons to take stock and strengthen our measures of quality* (pp. 11–31). Baltimore, MD: Paul H. Brookes.
- Buysse, V., Peisner-Feinberg, E., Páez, M., Scheffer Hammer, C., & Knowles, M. (2014). Effects of early education programs and practices on the development and learning

- of dual language learners: A review of the literature. *Early Childhood Research Quarterly*, 29, 765–785. <https://doi.org/10.1016/j.ecresq.2013.08.004>
- Cadima, J., Leal, T., & Burchinal, M. (2010). The quality of teacher-student interactions: Associations with first graders' academic and behavioral outcomes. *Journal of School Psychology*, 48, 457–482. <https://doi.org/10.1016/j.jsp.2010.09.001>
- Canale, M. (1983). From communicative competence to communicative language pedagogy. In J. C. Richards & R. W. Schmidt (Eds.), *Language and communication* (pp. 2–27). London, England: Longman.
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1, 1–47. <https://doi.org/10.1093/applin/1.1.1>
- CBS (2003). *Jeugd 2003, cijfers en feiten*. Retrieved from <https://www.cbs.nl/nl-nl/publicatie/2003/20/jeugd-2003-cijfers-en-feiten>
- Cekaite, A., & Aronsson, K. (2014). Language play, peer group improvisations, and L2 learning. In Cekaite, A., Blum-Kulka, S., Grøver, V., & Teunel, E. (Eds.), *Children's peer talk: Learning from each other* (pp. 194–213). Cambridge, England: Cambridge University Press.
- Celce-Murcia, M. (2007). Rethinking the role of communicative competence in language teaching. In E. Alcón Soler, & M. P. Safont Jordà (Eds.), *Intercultural Language use and language learning* (pp. 41–57). Dordrecht, the Netherlands: Springer Netherlands.
- Celce-Murcia, M., Dörnyei, Z., & Thurrell, S. (1995). Communicative Competence: A pedagogically motivated model with content specifications. *Issues in Applied Linguistics*, 6(2), 5–35.
- Cenoz, J., & Genesee, F. (1998). Psycholinguistic perspectives on multilingualism and multilingual education. In J. Cenoz, & F. Genesee (Eds.), *Beyond Bilingualism: Multilingualism and Multilingual Education* (pp. 16–32). Clevedon, England: Multilingual Matters.
- Chang, F., Crawford, G., Early, D., Bryant, D., Howes, C., Burchinal, M., ... Planta, R. (2007). Spanish-speaking children's social and language development in pre-kindergarten classrooms. *Early Education and Development*, 18, 243–269. <https://doi.org/10.1080/10409280701282959>
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
- Cleave, P., Girolametto, L., Chen, X., & Johnson, C. (2010). Narrative abilities in monolingual and dual language learning children with specific language

- impairment. *Journal of Communication Disorders*, 43, 511–522.  
<https://doi.org/10.1016/j.jcomdis.2010.05.005>
- Collier, V. P. (1987). Age and rate of acquisition of second language for academic purposes. *TESOL Quarterly*, 21, 617–641. <https://doi.org/10.2307/3586986>
- Collier, V. P., & Thomas, W. P. (1989). How quickly can immigrants become proficient in school English? *Journal of Educational Issues of Language Minority Students*, 5, 26–38.
- Conboy, B. T., & Thal, D. J. (2006). Ties between the lexicon and grammar: Cross-sectional and longitudinal studies of bilingual children. *Child Development*, 77, 712–735. <https://doi.org/10.1111/j.1467-8624.2006.00899.x>
- Cook, V. (1992). Evidence for multi-competence. *Language Learning*, 42, 557–91.  
<https://doi.org/10.1111/j.1467-1770.1992.tb01044.x>
- Cummins, J. (1981). Age on arrival and immigrant second language learning in Canada: A reassessment. *Applied Linguistics*, II, 132–149.
- Cummins, J. (1984). *Bilingualism and special education: Issues in assessment and pedagogy*. Clevedon, England: Multilingual Matters.
- Cummins, J. (1989). *Empowering minority students*. Sacramento, CA: California Association for Bilingual Education.
- Cummins, J. (2000). *Language, power and pedagogy: Bilingual children in the crossfire*. Clevedon, England: Multilingual Matters.
- Cummins, J. (2001). Assessment and Intervention with culturally and linguistically diverse learners. In S. Hurley, & J. Tinajero (Eds.), *Literacy assessment for second language learners* (pp. 115-129). Boston, MA: Allyn & Bacon.
- Cummins, J. (2007). Rethinking monolingual instructional strategies in multilingual classrooms. *The Canadian Journal of Applied Linguistics*, 10, 221–240.
- Cummins, J. (2008). BICS and CALP: Empirical and theoretical status of the distinction. In B. Street & N. H. Hornberger (Eds), *Encyclopedia of language and education* (2nd. ed.), Vol. 2, (pp. 71–83). Dordrecht, the Netherlands: Springer.
- Daller, H., Van Hout, R., & Treffers-Daller, J. (2003). Lexical Richness in the Spontaneous Speech of Bilinguals. *Applied Linguistics*, 24, 197–222.  
<https://doi.org/10.1093/applin/24.2.197>
- de Haan, A., Elbers, E., Hoofs, H., & Leseman, P. P. M. (2013). Targeted versus mixed preschools and kindergartens: Effects of class composition and teacher-managed activities on disadvantaged children's emergent academic skills. *School*

- Effectiveness and School Improvement*, 24, 177–194.  
<https://doi.org/10.1080/09243453.2012.749792>
- de Haan, A., Elbers, E., & Leseman, P. P. M. (2014). Teacher- and child-managed academic activities in preschool and kindergarten and their influence on children's gains in emergent academic skills. *Journal of Research in Childhood Education*, 28, 43–58. <https://doi.org/10.1080/02568543.2013.851750>
- de Haan, A., Leseman, P. P. M., & Elbers, E. (2011). *Pilot gemengde groepen 2007-2010. Onderzoeksrapportage oktober 2011*. Utrecht, the Netherlands: Utrecht University. Retrieved from <https://www.spelenderwijsutrecht.nl/Portals/0/DocumentenExtern/Pilot%20gemengde%20groepen%20eindrapport.pdf>
- de Villiers, P. A., & de Villiers, J. G. (1979). *Early language*. Cambridge, MA: Harvard University Press.
- Diamond, A., & Lee, K. (2011). Interventions shown to aid executive function development in children 4 to 12 years old. *Science*, 333, 959–964.  
<https://doi.org/10.1126/science.1204529>
- Downer, J. T., López, M. L., Grimm, K., Hamagami, A., Pianta, R. C., & Howes, C. (2012). Observations of teacher-child interactions in classrooms serving Latinos and dual language learners: Applicability of the Classroom Assessment Scoring System in diverse settings. *Early Childhood Research Quarterly*, 27, 21–32.  
<https://doi.org/10.1016/j.ecresq.2011.07.005>
- Dryden-Peterson, S. (2015). *The educational experiences of refugee children in countries of first asylum*. Washington, D.C.: Migration Policy Institute.
- Driessen, G. (1996). De taalvaardigheid Nederlands van allochtone en autochtone leerlingen. De ontwikkeling in het basisonderwijs in kaart gebracht. *Gramma/TT*, 5(1), 31–40.
- Duarte, J., & van der Meij, M. (2018). A holistic model for multilingualism in education. *EuroAmerican Journal of Applied Linguistics and Languages*, 5(2), 24–43.  
<https://doi.org/10.21283/2376905X.9.153>
- Dunn, L. M., & Dunn, D. M. (2005). *Peabody Picture Vocabulary Test (PPVT-3)* (3rd ed.). Minneapolis, MN: Pearson.
- Early, D. M., Bryant, D. M., Pianta, R. C., Clifford, R. M., Burchinal, M. R., Ritchie, S., ... Barbarin, O. (2006). Are teachers' education, major, and credentials related to classroom quality and children's academic gains in pre-kindergarten? *Early Childhood Research Quarterly*, 21, 174–195.  
<https://doi.org/10.1016/j.ecresq.2006.04.004>

- Early, D. M., Iruka, I. U., Ritchie, S., Barbarin, O. A., Winn, D-M. C, Crawford, G. M., ... Pianta, R. C. (2010). How do pre-kindergarteners spend their time? Gender, ethnicity, and income as predictors of experiences in pre-kindergarten classrooms. *Early Childhood Research Quarterly*, 25, 177–193.  
<https://doi.org/10.1016/j.ecresq.2009.10.003>
- [Educational Inspectorate] Inspectie van het onderwijs (2019). *De staat van het onderwijs 2019*. Retrieved on April 19<sup>th</sup> 2019 from <http://www.destaatvanhetonderwijs.nl>
- Elias, C. L., & Berk, L. E. (2002). Self-regulation in young children: Is there a role for sociodramatic play? *Early Childhood Research Quarterly*, 17, 216–238.  
[https://doi.org/10.1016/S0885-2006\(02\)00146-1](https://doi.org/10.1016/S0885-2006(02)00146-1)
- Emmelot, Y., van Schooten, E., Timman, Y., Verhallen, M., & Verhallen, S. (2001). *Nieuwe kansen voor taalonderwijs aan anderstaligen*. Den Haag, the Netherlands: Wetenschappelijke raad voor het regeringsbeleid.
- European Parliament (2009). *Minimum standards for the reception of asylum seekers (recast)*. European Parliament legislative resolution of 7 May 2009 on the proposal for a directive of the European Parliament and of the Council laying down minimum standards for the reception of asylum seekers (recast) (COM(2008)0815 – C6-0477/2008 – 2008/0244(COD)). Retrieved from <http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//NONSGML+TA+P6-TA-2009-0376+0+DOC+PDF+V0//EN>
- European Union (2013). *Study on educational support for newly arrives migrant children. Final report*. Retrieved from <https://publications.europa.eu/en/publication-detail/-/publication/96c97b6b-a31b-4d94-a22a-14c0859a8bea>
- Fiestas, C. E., & Peña, E. D. (2004). Narrative Discourse in Bilingual children: Language and task effects. *Language, Speech, and Hearing Services in Schools*, 35, 155-168  
[https://doi.org/10.1044/0161-1461\(2004/016\)](https://doi.org/10.1044/0161-1461(2004/016))
- Fivush, R., Haden, C., & Adam, S. (1995). Structure and coherence of preschoolers' personal narratives over time: Implications for childhood amnesia. *Journal of Experimental Child Psychology*, 60, 32–56.  
<https://doi.org/10.1006/jecp.1995.1030>
- Gagarina, N. (2016). Narratives of Russian-German preschool and primary school bilinguals: Rasskaz and Erzählung. *Applied Psycholinguistics*, 37, 91–122.  
<https://doi.org/10.1017/S0142716415000430>
- Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiūnienė, I., ... Walters, J. (2012). *Multilingual Assessment Instrument for Narratives (MAIN)*. ZAS papers in linguistics 56. Berlin, Germany: ZAS. Retrieved from <http://www.zas.gwz-berlin.de/zaspil56.html>

- Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiūnienė, I., ... Walters, J. (2015). Assessment of narrative abilities in bilingual children. In S. Armon-Lotem, J. de Jong, & N. Meir (Eds.), *Assessing multilingual children. Disentangling bilingualism from language impairment* (pp. 241–274). Bristol, England: Multilingual Matters.
- Gathercole, V. C. M., & Thomas, E. M. (2009). Bilingual first-language development: Dominant language takeover, threatened minority language take-up. *Bilingualism: Language and Cognition*, 12, 213–237.  
<https://doi.org/10.1017/S1366728909004015>
- García, O. (2009). *Bilingual education in the 21st century: A global perspective*. Malden, MA: Wiley-Blackwell.
- García, O., & Wei, L. (2014). *Translanguaging: Language, bilingualism and education*. Basingstoke, England: Palgrave Macmillan.
- Gass, S. M., & Selinker, L. (2008). *Second language acquisition. An introductory course*. (3rd. ed.). New York, NY: Routledge.
- Genesee, F., Paradis, J., & Crago, M. (2004). *Dual language development and disorders: A handbook on bilingualism and second languages learning*. Baltimore, MD: Paul H. Brookes.
- Gielen, S., & Işçi, A. (2015). *Meertaligheid: een troef!* Sint-Niklaas: Abimo Uitgeverij.
- Gilmore, A. (2016). Language learning in context: Complex dynamic systems and the role of mixed methods research. In J. King (Ed.), *The dynamic interplay between context and the language learner* (pp. 194–224). Leicester, England: University of Leicester.
- Golberg, H., Paradis, J., & Crago, M. (2008). Lexical acquisition over time in minority first language children learning English as a second language. *Applied Psycholinguistics*, 29, 41–65. <https://doi.org/10.1017/S014271640808003X>
- Groothoff, F. C. (in press). From a monolingual to a multilingual approach in primary schools in the Netherlands. In Mary. L., Kruger A. B., & Young, A. S. (Eds.), *Migration, Multilingualism and Education: Critical perspectives on inclusion*. Bristol: Multilingual Matters.
- Grosjean, F. (1992). Another view of bilingualism. In R. J. Harris (Ed.), *Cognitive processing in bilinguals* (pp. 51–62). Amsterdam, the Netherlands: North Holland.
- Gutiérrez-Clellen, V. F. (2002). Narratives in two languages: Assessing performance of bilingual children. *Linguistics and Education*, 13, 175–197.  
[https://doi.org/10.1016/S0898-5898\(01\)00061-4](https://doi.org/10.1016/S0898-5898(01)00061-4)

- Hamre, B. K., & Pianta, R. C. (2005). Can instructional and emotional support in the first grade classroom make a difference for children at risk of school failure? *Child Development, 76*, 949-967. <https://doi.org/10.1111/j.1467-8624.2005.00889.x>
- Hamre, B. K., & Pianta, R. C. (2007). Learning opportunities in preschool and early elementary classrooms. In R. Pianta, M. Cox, & K. Snow (Eds.), *School readiness and the transition to kindergarten in the era of accountability* (pp. 49–84). Baltimore, MD: Paul H. Brookes.
- Hamre, B. K., Pianta, R. C., Burchinal, M. R., Field, S., LoCasale-Crouch, J., Downer, J. T., ... Scott-Little, C. (2012). A course on effective teacher-child interactions: Effects on teacher beliefs, knowledge, and observed practice. *American Educational Research Journal, 49*, 88-123  
<https://doi.org/10.3102/0002831211434596>
- Hamre, B. K., Pianta, R. C., Downer, J. T., DeCoster, J., Mashburn, A. J., Jones, S. M., ... Hamagami, A. (2013). Teaching through interactions: Testing a developmental framework of teacher effectiveness in over 4,000 classrooms. *The Elementary School Journal, 113*, 461–487. <https://doi.org/10.1086/669616>
- Harms, T., Clifford, R. M., & Cryer, D. (1998). *The early childhood environment rating scale* (Rev. ed.). New York, NY: Teachers College Press.
- Hatfield, B. E., Burchinal, M. R., Pianta, R. C., & Sideris, J. (2016). Thresholds in the association between quality of teacher–child interactions and preschool children’s school readiness skills. *Early Childhood Research Quarterly, 36*, 561–571.  
<https://doi.org/10.1016/j.ecresq.2015.09.005>
- Hayward, D., & Schneider. P. (2000). Effectiveness of teaching story grammar knowledge to pre-school children with language impairment. An exploratory study. *Child Language Teaching and Therapy, 16*, 255–284.  
<https://doi.org/10.1177/026565900001600303>
- Helmke, A., & Schrader, F. W. (1988). Successful student practice during seatwork: Efficient management and active supervision not enough. *Journal of Educational Research, 82*, 70–76. <https://doi.org/10.1080/00220671.1988.10885870>
- Henrichs, L. F., Damhuis, C. M. P., Trooster, C. S. M., van Look, C.J., Be, M. S. Y., & Schot, W. D. (2017). *Interactievaardigheden in de kinderopvang. Een onderzoek naar ervaren effecten van de trainingen 'TINK' en 'Oog voor interactie'*. Utrecht, the Netherlands: Utrecht University. Retrieved from  
<https://www.kennisplatformwerkeninkomen.nl/documenten/rapporten/2017/06/13/interactievaardigheden-in-de-kinderopvang>
- Henrichs, L., & Leseman, P. P. M. (2016). *Onderzoeksrapport VVE Utrecht 2012-2015. Wetenschap en Werkveld ontmoeten elkaar in actieonderzoek*. Utrecht, the

- Netherlands Utrecht University. Retrieved from [https://www.uu.nl/sites/default/files/vve\\_utrecht\\_onderzoek\\_henrichsleseman.pdf](https://www.uu.nl/sites/default/files/vve_utrecht_onderzoek_henrichsleseman.pdf)
- Herdina, P., & Jessner, U. (2000). The dynamics of third language acquisition. In J. Cenoz, & U. Jessner (Eds.), *English in Europe: The acquisition of a third language* (pp. 84–98). Clevedon, England: Multilingual Matters.
- Herzog-Punzenberger, B. (2016). *Successful integration of migrant children in EU member states: Examples of good practice*. NESET II ad hoc question 1 / 2016. Retrieved from <https://nesetweb.eu/en/resources/library/successful-integration-of-migrant-children-in-eu-member-states/>
- Herzog-Punzenberger, B., Le Pichon-Vorstman, E., & Siarova, H. (2017). *Multilingual education in the light of diversity: Lessons learned, NESET II report*, Luxembourg, Luxembourg: Publications Office of the European Union. <https://doi.org/10.2766/71255>
- Hindman, A. H., Wasik, B. A., & Snell, E. K. (2016). Closing the 30 million word gap: Next steps in designing research to inform practice. *Child Development Perspectives*, 10, 1334-139 . <https://doi.org/10.1111/cdep.12177>
- Hoff, E., Core, C., Place, S., Rumiche, R., Señor, M., & Parra, M. (2012). Dual language exposure and early bilingual development. *Journal of Child Language*, 39, 1–27. <https://doi.org/10.1017/S0305000910000759>
- Howes, C., Burchinal, M., Pianta, R., Bryant, D., Early, D., Clifford, R., & Barbarin, O. (2008). Ready to learn? Children’s pre-academic achievement in pre-Kindergarten programs. *Early Childhood Research Quarterly*, 23, 27–50. <https://doi.org/10.1016/j.ecresq.2007.05.002>
- Howes, C., & Ritchie, S. (2002). *A matter of trust: Connecting teachers and learners in the early childhood classroom*. New York, NY: Teachers College.
- Howes, C., Shivers, E. M., & Ritchie, S. (2004). Improving social relationships in child care through a researcher-program partnership. *Early Education and Development*, 15, 57–78. [https://doi.org/10.1207/s15566935eed1501\\_4](https://doi.org/10.1207/s15566935eed1501_4)
- Howes, C., Wishard Guerra, A., Fuligni, A., Zucker, E., Lee, L., Obregon, N. B., & Spivak, A. (2011). Classroom dimensions predict early peer interaction when children are diverse in ethnicity, race, and home language? *Early Childhood Research Quarterly*, 26, 399–408. <https://doi.org/10.1016/j.ecresq.2011.02.004>
- Hymes, D. H. (1972). On communicative competence. In J. B. Pride and J. Holmes (Eds.), *Sociolinguistics. Selected eadings*, Pt.1, (pp. 269–293). Harmondsworth, England: Penguin.

- IBM Corp. Released 2013. IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY: IBM Corp.
- Iluz-Cohen, P., & Walters, J. (2012) Telling stories in two languages: Narratives of bilingual children with typical and impaired language. *Bilingualism: Language and Cognition*, 15, 58–74. <https://doi.org/10.1017/S1366728911000538>
- Jacobs, B. (2016). *Woordenschat van nieuwkomers in het basisonderwijs. Onderzoek naar de geschikte toets om de woordenschat van nieuwkomers in de bovenbouw van het basisonderwijs te meten.* (Master's thesis, Universiteit van Amsterdam, the Netherlands). Retrieved from <https://www.lowan.nl/wp-content/uploads/2017/04/Masterscriptie-Babet-Jacobs.pdf>
- Jacoby, J. W., & Lesaux, N. K. (2017). Language and literacy instruction in preschool classes that serve Latino dual language learners. *Early Childhood Research Quarterly*, 40, 77–86. <https://doi.org/10.1016/j.ecresq.2016.10.001>
- Jelsma, B. (2015). *Productieve taalontwikkeling in het Nederlands van nieuwkomers.* (Master's thesis, Utrecht University, the Netherlands). Retrieved from <https://dspace.library.uu.nl/handle/1874/320019>
- Jia, G., & Fuse, A. (2007). Acquisition of English grammatical morphology by native Mandarin-speaking children and adolescents. Age-related differences. *Journal of Speech, Language and Hearing Research*, 50, 1280–1299. [https://doi.org/10.1044/1092-4388\(2007\)090](https://doi.org/10.1044/1092-4388(2007)090)
- John, S. F., Lui, M., & Tannock, R. (2003). Children's story retelling and comprehension using a new narrative resource. *Canadian Journal of School Psychology*, 18, 91–113. <https://doi.org/10.1177/082957350301800105>
- Justice, L. M., Mashburn, A. J., Hamre, B. K., & Pianta, R. C. (2008). Quality of language and literacy instruction in preschool classrooms serving at-risk pupils? *Early Childhood Research Quarterly*, 23, 51–68. <https://doi.org/10.1016/j.ecresq.2007.09.004>
- Justice, L. M., Petscher, Y., Schatschneider, C., & Mashburn, A. (2011). Peer effects in preschool classrooms: Is children's language growth associated with their classmates' skills? *Child Development*, 82, 1768–1777. <https://doi.org/10.1111/j.1467-8624.2011.01665.x>
- Kapalková, S., Polišenská, K., Marková, L., & Fenton, J. (2016). Narrative abilities in early successive bilingual Slovak–English children: A cross-language comparison. *Applied Psycholinguistics*, 37, 145–164. <https://doi.org/10.1017/S0142716415000454>

- Kaplan, I., Stolk, Y., Valibhoy, M., Tucker, A., & Baker, J. (2016). Cognitive assessment of refugee children: Effects of trauma and new language acquisition. *Transcultural Psychiatry*, 53, 81–109. <https://doi.org/10.1177%2F1363461515612933>
- Kemper, S. (1984). The development of narrative skills: Explanations and entertainments. In S. A. Kuczaj (Ed.), *Discourse development* (pp. 99–124). New York, NY: Academic Press.
- Kunnari, S., Välimaa, T., & Laukkanen-Nevala, P. (2016). Macrostructure in the narratives of monolingual Finnish and bilingual Finnish–Swedish children. *Applied Psycholinguistics*, 37, 123–144. <https://doi.org/10.1017/S0142716415000442>
- Kutnick, P., Brighi, A., with co-directors Avgitidou, S., Genta, M. L., Hännikäinen, M., Karlsson-Lohmander, M., ... Lofqvist, M. (2007). The role and practice of interpersonal relationships in European early education settings: Sites for enhancing social inclusion, personal growth and learning? *European Early Childhood Education Research Journal*, 15, 379–406. <https://doi.org/10.1080/13502930701679429>
- Labov, W. (1972). *Language in the inner city: Studies in the black English vernacular*. Philadelphia, PA: University of Pennsylvania Press.
- Labov, W., & Waletzky, J. (1967). Narrative analysis: Oral versions of personal experience. In J. Helm (Ed.), *Essays on the verbal and visual arts* (pp. 12–44). Seattle, WA: University of Washington.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159–174. <https://doi.org/10.2307/2529310>
- La Paro, L. M., Pianta, R. C., & Stuhlman, M. (2004). Classroom Assessment Scoring System (CLASS). Findings from the pre-K year. *Elementary School Journal*, 104, 409–426. <https://doi.org/10.1086/499760>
- Law, J., & Roy, P. (2008). Parental report of infant language skills: A review of the development and application of the communicative development inventories. *Child and Adolescent Mental Health*, 13(4), p. 198–206. <https://doi.org/10.1111/j.1475-3588.2008.00503.x>
- Schrijfgroep Lectoreninitiatief Professionalisering Taalonderwijs Nieuwkomers (2017). *Ruimte voor nieuwe talenten. Keuzes rond nieuwkomers op de basisschool*. Utrecht: POraad. Retrieved from <https://www.taallectoren.nl>
- Le Pichon, E. & Baauw, S. (2015). *Onderzoeksrapport Taalschool Utrecht*. [Research report Language school Utrecht]. Utrecht, the Netherlands: Utrecht University.

- Le Pichon, E., van Erning, R., & Baauw, S. (2016). *Nederland. Edina*. Utrecht, the Netherlands: Utrecht University. Retrieved from <https://edinaplatform.eu/wp-content/uploads/sites/77/2016/05/Country-Report-Nederland-1.pdf>
- Leseman, P. P. M. (2000). Bilingual vocabulary development of Turkish preschoolers in the Netherlands. *Journal of Multilingual and Multicultural Development*, 21, 93–112. <https://doi.org/10.1080/01434630008666396>
- Leseman, P. P. M., & Slot, P. L. (2013). *Kwaliteit en curriculum van voorschoolse opvang en educatie in Nederland. Relaties met structurele kenmerken, organisatiekenmerken en gebruik van educatieve programma's*. Utrecht, the Netherlands: Utrecht University, Onderzoeksprogramma Education & Learning - Centrum voor Cognitieve en Motorische Ontwikkelingsproblemen. Retrieved from <https://www.nro.nl/vve-leseman-en-slot/>
- Leseman, P. P. M., & Veen, A. (Eds.) (2016). *Ontwikkeling van kinderen en relatie met kwaliteit van voorschoolse instellingen. Resultaten uit het Pre-Cool cohortonderzoek*. Amsterdam, the Netherlands: Kohnstamm Instituut. Retrieved from <https://www.rijksoverheid.nl/documenten/rapporten/2016/06/06/ontwikkeling-van-kinderen-en-relatie-met-kwaliteit-van-voorschoolse-instellingen>
- Leseman, P.P.M., Henrichs, L.F., Blom, E. & Verhagen, J. (2019). Young mono- and bilingual children's exposure to academic language as related to language development and school achievement. In V. Grøver, P. Uccelli, M. Rowe & E. Lieven (Eds), *Learning through language* (pp. 205–217). Cambridge, MA: Cambridge University Press.
- Lewis, W. G., Jones, B., & Baker, C. (2013). 100 bilingual lessons: Distributing two languages in classrooms. In C. Abello-Contesse, P. M. Chandler, M. D. López-Jiménez, & R. Chacón-Beltrán (Eds.), *Bilingual and multilingual education in the 21st century* (pp. 107–135). Bristol, UK: Multilingual Matters.
- Liddicoat, A. J. & Scarino, A. (2013). *Intercultural Language Teaching and Learning*. Hoboken, NJ: Wiley-Blackwell.
- Lightbown, P. M., & Spada, N. (2013). *How languages are learned* (4th ed.). Oxford, England: Oxford University Press.
- Liles, B. Z. (1993). Narrative discourse in children with language disorders and children with normal language: A critical review of the literature. *Journal of Speech and Hearing Research*, 36, 868–882. <https://doi.org/10.1044/jshr.3605.868>
- Liles, B. Z., Duffy, R. J., Merritt, D. D., & Purcell, S. L. (1995). Measurement of narrative discourse ability in children with language disorders. *Journal of Speech and Hearing Research*, 38, 415–425. <https://doi.org/10.1044/jshr.3802.415>

- Lindgren, J. (2018). *Developing narrative competence Swedish, Swedish-German and Swedish-Turkish children aged 4–6* (Doctoral dissertation, Uppsala, Sweden). *Studia Linguistica Upsaliensia*, 19. <https://www.lingfil.uu.se/research/publikationsserier/studia-linguistica-upsaliensia>
- Love, J. M., Meckstroth, A., & Sprachman, S. (1997). Measuring the quality of program environments in Head Start and other early childhood programs: A review and recommendations for future research (Contract RN 94094001). Washington, DC: U.S. Department of Education, National Center for Education Statistics. Retrieved from <https://eric.ed.gov/?id=ED416030>
- Maarse, J., & Muller, P. (2017). Anticiperen op een aanbod? Naar effectief taalonderwijs voor nieuwkomerskinderen in het onderwijs. *Les*, 203, 46–50.
- MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk* (3rd. ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Mandler, G. (1979). Organization and repetition: Organization principles with special reference to rote learning. In L.-G. Nilsson (Ed.), *Perspectives on memory research* (pp. 293–327). Hillsdale, NJ: Erlbaum.
- Maneva, B., & Genesee, F. (2001). Bilingual babbling: Evidence for language differentiation in bilingual first language acquisition. Paper presented at the *Boston University Conference on Language Development*, Nov. 3, 2001. Retrieved from [https://www.researchgate.net/publication/301690867\\_Bilingual\\_babbling\\_Evidence\\_for\\_language\\_differentiation\\_in\\_dual\\_language\\_acquisition](https://www.researchgate.net/publication/301690867_Bilingual_babbling_Evidence_for_language_differentiation_in_dual_language_acquisition)
- Manyak, P. C. (2004). "What did she say?": Translation in a primary-grade English immersion class. *Multicultural Perspectives*, 6(1), 12–18. [https://doi.org/10.1207/S15327892mcp0601\\_3](https://doi.org/10.1207/S15327892mcp0601_3)
- Marchman, V. A., Martínez-Sussmann, C., & Dale, P. S. (2004). The language-specific nature of grammatical development: Evidence from bilingual language learners. *Developmental Science*, 7, 212–224. <https://doi.org/10.1111/j.1467-7687.2004.00340.x>
- Mashburn, A. J., Justice, L. M., Downer, J. T., & Pianta, R. C. (2009). Peer effects on children's language achievement during pre-kindergarten. *Child Development*, 80, 686–702. <https://doi.org/10.1111/j.1467-8624.2009.01291.x>
- Mashburn, A. J., Pianta, R. C., Hamre, B. K., Downer, J. T., Barbarin, O. A., Bryant, D. ... C. Howes (2008). Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills. *Child Development*, 79, 732–749. <https://doi.org/10.1111/j.1467-8624.2008.01154.x>

- Maviş, İ., Tunçer, M., & Gagarina, N. (2016). Macrostructure components in narrations of Turkish–German bilingual children. *Applied Psycholinguistics*, 37, 69–89. <https://doi.org/10.1017/S0142716415000429>
- Mayer, M., (1973). *Frog On His Own*. New York, NY: Dial.
- Mayer, M., (1975). *One Frog Too Many*. New York, NY: Dial.
- McCabe, A., & Rosenthal Rollins, P. (1994). Assessment of preschool narrative skills. *American Journal of Speech-Language Pathology*, 3(1), 45–56. <https://doi.org/10.1044/1058-0360.0301.45>
- McDonald Connor, C. M. D., Morrison, F. J., Fishman, B. J., Ponitz, C. C., Glasney, S., Underwood, P. S., ... Schatschneider, C. (2009). The ISI classroom observation system: Examining the literacy instruction provided to individual students. *Educational Researcher*, 38, 85–99. <https://doi.org/10.3102%2F0013189X09332373>
- Melhuish, E., Ereky-Stevens, K., Petrogiannis, K., Ariescu, A., Penderi, E., Rentzou, K., ... Leseman, P. P. M. (2015). *A review of research on the effects of early childhood education and care (ECEC) upon child development*. CARE project: Curriculum quality analysis and impact review of European early childhood education and care (ECEC). Retrieved from <http://ecec-care.org/resources/publications/>
- Mellenbergh, G. J. (1976). Bekend, maar onbemind (deel I). [Syllabus Known, but unloved (part I)]. Amsterdam, the Netherlands: Universiteit van Amsterdam, Subfaculteit Psychologie.
- Meyer, L. A., Wardrop, J. L., Hastings, C. N., & Linn, R. L. (1993). Effects of ability and settings on kindergarteners' reading performance. *Journal of Educational Research*, 86, 142–160. <https://doi.org/10.1080/00220671.1993.9941153>
- [Ministry for Primary and Secondary Education and Media] Minister voor Basis- en Voortgezet Onderwijs en Media (2015a). *Kamerbrief 1 december: Beantwoording van vragen naar aanleiding van Volkskrantartikel over vluchtelingenonderwijs* (1st of December, 2015). Retrieved from <https://docplayer.nl/12466686-Datum-1-december-2015-beantwoording-van-vragen-naar-aanleiding-van-volkskrantartikel-over-vluchtelingenonderwijs.html>
- [Ministry for Primary and Secondary Education and Media] Minister voor Basis- en Voortgezet Onderwijs en Media (2015b). *Onderwijs aan asielzoekers* (26th of October, 2015). Retrieved from <https://www.rijksoverheid.nl/documenten/kamerstukken/2015/10/26/kamerbrief-over-onderwijs-aan-asielzoekers>

- [Ministry for Primary and Secondary Education and Media] Minister voor Basis- en Voortgezet Onderwijs en Media (2018). *Eerste Regeling bekostiging personeel PO 2018–2019 en vaststelling bedragen voor ondersteuning van leerlingen in het PO en VO 2018–2019*. Retrieved from <http://wetten.overheid.nl/BWBR0040726/2018-03-21#Hoofdstuk6>
- Montie, J. E., Xiang, Z., & Schweinhart, L. J. (2006). Preschool experience in 10 countries: Cognitive and language performance at age 7. *Early Childhood Research Quarterly*, 21, 313–331. <https://doi.org/10.1016/j.ecresq.2006.07.007>
- Montrul, S. A. (2008). *Incomplete acquisition in bilingualism*. Amsterdam, the Netherlands: John Benjamins.
- Muñoz, C. (2008). Symmetries and asymmetries of age effects in naturalistic and instructed L2 learning. *Applied Linguistics*, 29, 578–596. <https://doi.org/10.1093/applin/amm056>
- Muñoz, M. L., Gillam, R. B., Peña, E. D., & Gulley-Faehnle, A. G. (2003). Measures of language development in fictional narratives of Latino children. *Language, Speech, and Hearing Services in Schools*, 34, 332–342. [https://doi.org/10.1044/0161-1461\(2003/027\)](https://doi.org/10.1044/0161-1461(2003/027))
- Nakamura, K. (2009). Language and affect. Japanese Children's use of evaluative expressions in narratives. In J. Guo, E. Lieven, N. Budwig, S. Ervin-Tripp, K. Nakamura, & Ş. Özçalışkan, (Eds.), *Crosslinguistic approaches to the psychology of language: Research in the tradition of Dan Isaac Slobin* (pp. 225–240). New York, NJ: Psychology Press.
- [National Institute for Curriculum Development in the Netherlands] Stichting Leerplan Ontwikkeling (2013). Retrieved from <http://www.doelennieuwkomers.slo.nl/> in 2013 (this website can still be found via <https://web.archive.org/>)
- NICHD Early Child Care Research Network (2002a). Early child care and children's development prior to school entry: Results from the NICHD study of early child care. *American Educational Research Journal*, 39, 133–164. <https://doi.org/10.3102%2F00028312039001133>
- NICHD Early Child Care Research Network (2002b). The relation of global first grade classroom environment to structural classroom features, teacher, and student behaviors. *The Elementary School Journal*, 102, 367–387. <https://doi.org/10.1086/499709>
- NICHD Early Child Care Research Network (2004). Does class size in first grade relate to children's academic and social performance or observed classroom processes? *Developmental Psychology*, 40, 651–664. <https://doi.org/10.1037/0012-1649.40.5.651>

- NICHD Early Child Care Research Network (2005). A day in third grade: A large-scale study of classroom quality and teacher and student behavior. *The Elementary School Journal*, 105, 305–323. <https://doi.org/10.1086/428746>
- Nicoladis, E. (2001). Finding first words in the input. In J. Cenoz, & F. Genesee (Eds.), *Trends in bilingual acquisition* (pp. 131–147). Amsterdam, the Netherlands: John Benjamins.
- Nippold, M. A., Ward-Loneragan, J. M., & Fanning, J. L. (2005). Persuasive writing in children, adolescents, and adults: A study of syntactic, semantic, and pragmatic development. *Language, Speech, and Hearing Services in Schools*, 36, 125–128. [https://doi.org/10.1044/0161-1461\(2005/012\)](https://doi.org/10.1044/0161-1461(2005/012))
- [OECD] Organisation for Economic Co-operation and Development (2015). *Immigrant students at school: Easing the journey towards integration*. OECD Publishing. Retrieved from <https://doi.org/10.1787/9789264249509-en>
- Oller, D. K., Eilers, R. E., Urbano, R., & Cobo-Lewis, A. B. (1997). Development of precursors to speech in infants exposed to two languages. *Journal of Child Language*, 24, 407–425. <https://doi.org/10.1017/S0305000997003097>
- Oller, D. K., Pearson, B. Z., & Cobo-Lewis, A. B. (2007). Profile effects in early bilingual language and literacy. *Applied Psycholinguistics*, 28, 191–230. <https://doi.org/10.1017/S0142716407070117>
- Ortega, L., & Iberri-Shea, G. (2005). Longitudinal research in second language acquisition: Recent trends and future directions. *Annual Review of Applied Linguistics*, 25, 26–45. <https://doi.org/10.1017/S0267190505000024>
- Osborn, A. Q. (2012). *The Relation between high-quality prekindergarten classroom environments and literacy outcomes for students learning English as a second language* (Doctoral dissertation, University of Nebraska). Retrieved from <http://digitalcommons.unl.edu/cehssdiss/152>
- Özcan, F. H., Altinkamiş, F., & Gillis, S. (2016). Early lexical composition of Turkish-Dutch bilinguals: Nouns before verbs or verbs before nouns. *Poznan Studies in Contemporary Linguistics*, 52, 583–604. <https://doi.org/10.1515/psicl-2016-0023>
- Pakarinen, E., Lerkkanen, M.-K., Poikkeus, A.-M., Kiuru, N., Siekkinen, M., Rasku-Puttonen, H., & Nurmi, J.-E. (2010). A validation of the Classroom Assessment Scoring System in Finnish kindergartens. *Early Education and Development*, 21, 95–124. <https://doi.org/10.1080/10409280902858764>
- Paradis, J. (2011). Individual differences in child English second language acquisition: Comparing child-internal and child-external factors. *Linguistic Approaches to Bilingualism*, 1, 213–237. <https://doi.org/10.1075/lab.1.3.01par>

- Paradis, J., Genesee, F., & Crago, M. B. (2011). *Dual language development and disorders: A Handbook on bilingualism & second language learning* (2nd. ed.). Baltimore, MD: Paul H. Brookes.
- Pearson, B. Z. (2002). Narrative competence among monolingual and bilingual school children in Miami. In K. Oller, & R. Eilers (Eds.), *Language and literacy in bilingual children: Language and literacy in bilingual children* (pp.135–174). [ProQuest Ebook Central].  
<https://ebookcentral.proquest.com/lib/uunl/detail.action?docID=3007702>
- Pearson, B. Z., Fernández, S. C., Lewedeg, V., & Oller, D. K. (1997). The relation of input factors to lexical learning by bilingual infants. *Applied Psycholinguistics*, 18, 41–58. <https://doi.org/10.1017/S0142716400009863>
- Pearson, B. Z., Oller, D. K., Umbel, V. M., & Fernández, M. C. (1996, October). The relationship of lexical knowledge to measures of literacy and narrative discourse in monolingual and bilingual children. Paper presented at the *Second Language Research Forum*. Tucson, AZ.
- Peisner-Feinberg, E., Buysse, V., Fuligni, A., Burchinal, M., Espinosa, L., Halle, T., & Castro, D. C. (2014). Using early care and education quality measures with dual language learners: A review of the research. *Early Childhood Research Quarterly*, 29, 786–803. <https://doi.org/10.1016/j.ecresq.2014.04.013>
- Peristeri, E., Andreou, M., & Tsimpli, I. M. (2017). Syntactic and story structure complexity in the narratives of high- and low-language ability children with autism spectrum disorder. *Frontiers in Psychology*, 8, 1–16.  
<https://doi.org/10.3389/fpsyg.2017.02027>
- Pesco, D., & Kay-Raining Bird, E. (2016). Perspectives on bilingual children’s narratives elicited with the Multilingual Assessment Instrument for Narratives. *Applied Psycholinguistics*, 37, 1–9. <https://doi.org/10.1017/S0142716415000387>
- Phillips, D. A., & Lowenstein, A. E. (2011). Early care, education and child development. *Annual Review of Psychology*, 62, 483–500.  
<https://doi.org/10.1146/annurev.psych.031809.130707>
- Pianta, R. C. (1999). *Enhancing relationships between children and teachers*. Washington, DC: American Psychological Association.
- Pianta, R. C. (2003). *Standardized classroom observations from pre-K to 3rd grade: A mechanism for improving access to high-quality classroom experiences and practices during the p-3 years*. Retrieved from <https://www.fcd-us.org/assets/2016/04/StandardizedClassroomObservations.pdf>, 2019-04-24

- Pianta, R. C., Barnett, W. S., Burchinal, M., & Thornburg, K. R. (2009). The effects of preschool education: What we know, how public policy is or is not aligned with the evidence base and what we need to know. *Psychological Science in the Public Interest*, 10, 49–88. <https://doi.org/10.1177/1529100610381908>
- Pianta, R. C., DeCoster, J., Cabell, S., Burchinal, M., Hamre, B. K., Downer J., ... Howes, C. (2014). Dose–response relations between preschool teachers’ exposure to components of professional development and increases in quality of their interactions with children? *Early Childhood Research Quarterly*, 29, 499–508. <https://doi.org/10.1016/j.ecresq.2014.06.001>
- Pianta, R. C., Howes, C., Burchinal, M. R., Bryant, D., Clifford, R., Early, D., & Barbarin, O. (2005). Features of pre-kindergarten programs, classrooms, and teachers: Do they predict observed classroom quality and child-teacher interactions? *Applied Developmental Science*, 9, 144–159. [https://doi.org/10.1207/s1532480xads0903\\_2](https://doi.org/10.1207/s1532480xads0903_2)
- Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom Assessment Scoring System Manual, pre-K*. Baltimore, MD: Paul H. Brookes.
- Price, J. R., Roberts, J. E., & Jackson, S. C. (2006). Structural development of the fictional narratives of African American preschoolers. *Language, Speech, and Hearing Services in School*, 37, 178–190. [https://doi.org/10.1044/0161-1461\(2006/020\)](https://doi.org/10.1044/0161-1461(2006/020))
- Quené, H., & van den Bergh, H. (2004). On multi-level modeling of data from repeated measures designs: A tutorial. *Speech Communication*, 43, 103–121. <https://doi.org/10.1016/j.specom.2004.02.004>
- Rasbash, J., Charlton, C., Browne, W., Healy, M., & Cameron, B. (2016). *MLwiN version 2.36*. Bristol, England: University of Bristol, Centre for Multilevel Modelling.
- Reuterskiöld, C., Hansson, K., Sahlén, B. (2011). Narrative skills in Swedish children with language impairment. *Journal of Communication Disorders*, 44(6), 733–744. <https://doi.org/10.1016/j.jcomdis.2011.04.010>
- Reuterskiöld-Wagner, C., Sahlén, B., & Nettelbladt, U. (1999). What’s the story? Narration and comprehension in Swedish preschool children with language impairment. *Child Language Teaching and Therapy*, 15, 113–137. <https://doi.org/10.1177/026565909901500202>
- Ritchie, S., Howes, C., Kraft-Sayre, M., & Weiser, B. (2001). *Emerging Academics Snapshot*. Los Angeles: University of California at Los Angeles.
- Rydland, V. Grøver, V., & Lawrence, J. (2014). The second-language vocabulary trajectories of Turkish immigrant children in Norway from ages five to ten: The role of preschool talk exposure, maternal education, and co-ethnic concentration in

- the neighborhood. *Journal of Child Language*, 41, 352–381.  
<https://doi.org/10.1017/S0305000912000712>
- Savignon, S. J. (1976, April). Communicative Competence Theory and classroom practice. Paper presented at the *Central States Conference on the Teaching of Foreign Languages* (Detroit, MI, April 23, 1976)
- Schaefer, B., & Bowyer-Crane, C. (2015). Development of a tablet application for the screening of receptive vocabulary skills in multilingual children: A pilot study. *Child Language Teaching and Therapy*, 32, 179–191.  
<https://doi.org/10.1177/0265659015591634>
- Scheele, A.F. (2010). *Home language and mono- and bilingual children's emergent academic language : a longitudinal study of Dutch, Moroccan-Dutch, and Turkish-Dutch 3- to 6-year-old children* (Doctoral dissertation, Utrecht University, the Netherlands). Retrieved from  
<https://dspace.library.uu.nl/bitstream/handle/1874/44570/scheele.pdf;sequence=2>
- Schlichting, J. E. P. T. (2005). *Peabody Picture Vocabulary Test-III-NL. Dutch version*. Amsterdam, the Netherlands: Harcourt.
- Schneider, P., Hayward, D., & Vis Dubé, R. (2006). Storytelling from pictures using the Edmonton Narrative Norms Instrument. *Canadian Journal of Speech-Language Pathology and Audiology*, 30, 224–238.
- Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime user's guide*. Pittsburgh, PA: Psychology Software Tools.
- Skutnabb-Kangas, T. (2000). *Linguistic genocide in education-or worldwide diversity and human rights?* Mahwah, NJ: Lawrence Erlbaum Associates.
- Slembrouck, S., Van Avermaet, P., & Van Gorp, K. (2018). Strategies of multilingualism in education for minority children. In P. Van Avermaet, S. Slembrouck, K. Van Gorp, S. Sierens, & K. Maryns (Eds.), *The multilingual edge of education* (pp. 9–39). London, England: Palgrave Macmillan.
- Slot, P. L. (2014). *Early childhood education and care in the Netherlands. Quality, curriculum and relations with child development* (Doctoral dissertation, Utrecht University, the Netherlands). Retrieved from  
<https://dspace.library.uu.nl/handle/1874/301535>
- Slot, P. L., Boom, J., Verhagen, J., & Leseman, P. P. M. (2017). Measurement properties of the CLASS Toddler in ECEC in The Netherlands. *Journal of Applied Developmental Psychology*, 48, 79–91.  
<http://dx.doi.org/10.1016/j.appdev.2016.11.008>

- Slot, P. L., Broekhuizen, M. L., Leseman, P. P. M., & Veen, A. (2015a). Kwaliteit van voorzieningen en effecten op ontwikkeling van kinderen. In A. Veen, & P. P. M. Leseman (Eds.), *Pre-COOL cohortonderzoek: Resultaten over de voorschoolse periode* (pp. 73–88). Amsterdam, the Netherlands: Kohnstamm Instituut
- Slot, P. L., Lerkkanen, M., & Leseman, P. P. M. (2015b). *The relations between structural quality and process quality in European early childhood education and care provisions: Secondary analysis of large scale studies in five countries*. Retrieved from [http://ecec-care.org/fileadmin/careproject/Publications/reports/CARE\\_WP2\\_D2\\_\\_2\\_Secondary\\_data\\_analyses.pdf](http://ecec-care.org/fileadmin/careproject/Publications/reports/CARE_WP2_D2__2_Secondary_data_analyses.pdf)
- Slot, P. L., Leseman, P. P. M., Verhagen, J., & Mulder, H. (2015). Associations between structural quality aspects and process quality in Dutch early childhood education and care settings. *Early Childhood Research Quarterly*, 33, 64–76. <https://doi.org/10.1016/j.ecresq.2015.06.001>
- Smits, A. & van Koeven, E. (2016, July 12). Geletterdheid en schoolsucces. Goed taalonderwijs is goed NT2-onderwijs [Blog post]. Retrieved from [https://geletterdheidenschoolsucces.blogspot.com/2016/07/goed-taalonderwijs-is-goed-nt2-onderwijs\\_84.html](https://geletterdheidenschoolsucces.blogspot.com/2016/07/goed-taalonderwijs-is-goed-nt2-onderwijs_84.html)
- Sorenson Duncan, T. G. (2017). *How does maternal education influence language acquisition? Interdependencies between environment and input in the bilingual development of immigrant and refugee children* (Doctoral dissertation, University of Alberta, Canada). Retrieved from [https://sites.ualberta.ca/~tgs/DocumentsForWebsite/SorensonDuncan\\_Dissertation\\_Abstract.pdf](https://sites.ualberta.ca/~tgs/DocumentsForWebsite/SorensonDuncan_Dissertation_Abstract.pdf)
- Spencer, T. D., & Slocum, T. A. (2010). The effect of a narrative intervention on story retelling and personal story generation skills of preschoolers with risk factors and narrative language delays. *Journal of Early Intervention*, 32, 178–199. <https://doi.org/10.1177/1053815110379124>
- Squires, K. E., Lugo-Neris, M. J., Peña, E. D., Bedore, L. M., Bohman, T. M., & Gillam, R. B. (2014). Story retelling by bilingual children with language impairments and typically developing controls. *International Journal of Language & Communication Disorders*, 49, 60–74. <https://doi.org/10.1111/1460-6984.12044>
- Sridar, S.N. & Sridar, K.K. (2018). A bridge half-built: Toward a holistic theory of Second Language Acquisition and world Englishes. *World Englishes*, 17, 127–139. <https://doi.org/10.1111/weng.12308>
- Stein, N., & Glenn, C. (1979). An analysis of story comprehension in elementary school children. In R. Freedle (Ed.), *New directions in discourse processing*, Vol. 2, (pp.55–120). Norwood, NJ: Ablex.

- Stegelin, D. A., Anderson, D., Kemper, K., Wagner, J., & Evans, K. (2014). Exploring daily physical activity and nutrition patterns in early learning settings: Snapshots of young children in head start, primary, and after-school settings. *Early Childhood Education Journal*, 42, 133–142. [https://doi.org/10.1007/978-3-319-76006-3\\_10](https://doi.org/10.1007/978-3-319-76006-3_10)
- Strating-Keurentjes, H. (2000). *Lexicale ontwikkeling in het Nederlands van autochtone en allochtone kleuters* (Doctoral dissertation, Tilburg University, the Netherlands). Retrieved from <https://research.tilburguniversity.edu/en/publications/lexicale-ontwikkeling-in-het-nederlands-van-autochtone-en-allocht>
- Swanson, L. A., Fey, M. E., Mills, C. E., & Hood, L. S. (2005). Use of narrative-based language intervention with children who have specific language impairment. *American Journal of Speech-Language Pathology*, 14, 131–143. [https://doi.org/10.1044/1058-0360\(2005/014\)](https://doi.org/10.1044/1058-0360(2005/014))
- Tabors, P.O., & Snow, C. (1994). English as a second language in preschools. In F. Genesee (Ed.), *Educating second language children: The whole child, the whole curriculum, the whole community* (pp. 103–125). New York: Cambridge University Press.
- Thomas, W. P., & Collier, V. P. (2002). *A national study of school effectiveness for language minority students' long-term academic achievement*. <https://eric.ed.gov/?id=ED475048>
- Torrance, N., & Olson, D. R. (1984). Oral language competence and the acquisition of literacy. In A. Pellegrini, & T. Yawkey (Eds.), *The development of oral and written language in social contexts* (pp. 167–181). Norwood, NJ: Ablex.
- Trabasso, T., & Rodkin, P. C. (1994). Knowledge of goal/plans: A conceptual basis for narrating "Frog where are you?". In R. A. Berman, & D. I. Slobin (Eds.), *Relating events in narrative: A cross linguistic developmental study* (pp. 85–106). Mahwah, NJ: Lawrence Erlbaum Associates.
- Tsimpli, I. M., Peristeri, E., & Andreou, M. (2016). Narrative production in monolingual and bilingual children with specific language impairment. *Applied Psycholinguistics*, 37, 195–216. <https://doi.org/10.1017/S0142716415000478>
- Uccelli, P., & Páez, M. M. (2007). Narrative and vocabulary development of bilingual children from kindergarten to first grade: Developmental changes and associations among English and Spanish skills. *Language, Speech, and Hearing Services in Schools*, 38, 225–236. [https://doi.org/10.1044/0161-1461\(2007/024\)](https://doi.org/10.1044/0161-1461(2007/024))
- Unsworth, S., Hulk, A., & Marinis, T. (2011). Internal and external factors in child second language acquisition. *Linguistic Approaches to Bilingualism*, 1, 207–212. <https://doi.org/10.1075/lab.1.3.00int>

- Vandell, D. L., Belsky, J., Burchinal, M., Steinberg, L., & Vandergrift, N., &. (2010). Do effects of early child care extend to age 15 years? Results from the NIDHD study of early child care and youth development. *Child Development*, 81, 737–756. <https://doi.org/10.1111/j.1467-8624.2010.01431.x>
- van den Akker, J. (2003). Curriculum perspectives: An introduction. In J. van den Akker, W. Kuiper, & U. Hameyer (Eds.), *Curriculum landscapes and trends* (pp. 1–10). Dordrecht: Kluwer Academic.
- van den Bergh, M., Schmittmann, V. D., Hofman, A. D., & van der Maas, H. L. J. (2015). Tracing the development of typewriting skills in an adaptive e-learning environment. *Perceptual & Motor Skills*, 121, 727–745. <https://doi.org/10.2466%2F23.25.PMS.121c26x6>
- Van Druten-Frietman, L., Denessen, E., Gijssels, M., & Verhoeven, L. (2015). Child, home and institutional predictors of preschool vocabulary growth. *Learning and Individual Differences*, 43, 92–99. <https://doi.org/10.1016/j.lindif.2015.08.032>
- van Lier, L. (2002). An ecological-semiotic perspective on language and linguistics. In C. Kramsch (Ed.), *Language Acquisition and Language Socialization: Ecological Perspective* (pp. 140–164). London, England: Continuum.
- van Schaik, S. D. M., Leseman, P. P. M., & de Haan, M. (2018). Using a group-centered approach to observe interactions in early childhood education. *Child Development*, 89, 897–913. <https://doi.org/10.1111/cdev.12814>
- Veen, A., Karssen, A. M., Daalen, M. M. van, Roeleveld, J., Triesscheijn, J., & Elshof, D. (2013). *De aansluiting tussen voor- en vroegschoolse educatie en tussen vroegschoolse educatie en groep 3*. Amsterdam, the Netherlands: Kohnstamm Instituut. Retrieved from [http://pre-cool.kohnstammstituut.nl/assets/kohnstamm\\_892.pdf](http://pre-cool.kohnstammstituut.nl/assets/kohnstamm_892.pdf)
- Veen, A., van der Veen, I., Heurter, A. M. H., & Paas, T. (2012). *Pre-COOL cohortonderzoek. Technisch rapport vierjarigencohort, eerste meting 2009–2010* (Rapport 877, projectnummer 20379). Amsterdam, the Netherlands: Kohnstamm Instituut. Retrieved from <http://pre-cool.kohnstammstituut.nl/resultaten-en-publicaties.html>
- Veen, A., van der Veen, I., Heurter, A. M. H., Ledoux, G., Leseman, P. P. M., Mulder, H., ... Slot, P. L. (2012). *Pre-COOL cohortonderzoek. Technisch rapport tweejarigencohort, eerste meting 2010–2011* (Rapport 877, projectnummer 20379). Amsterdam, the Netherlands: Kohnstamm Instituut. Retrieved from <http://pre-cool.kohnstammstituut.nl/resultaten-en-publicaties.html>
- Veen, A., van der Veen, I., van Schaik, S., & Leseman, P.P.M. (2017). *Kwaliteit in kleutergroepen en de relatie met ontwikkeling van kinderen*. Amsterdam, the

- Netherlands: Kohnstamm Instituut. Retrieved from <http://pre-cool.kohnstammstituut.nl/resultaten-en-publicaties.html>
- Verhagen, J., Mulder, H., Boom, J., Hoofs, H., Slot, P. L., & Leseman, P. P. M. (2016). Woordenschat. In P. P. M. Leseman, A. & Veen (Eds.), *Ontwikkeling van kinderen en relatie met kwaliteit van voorschoolse instellingen: Resultaten uit het Pre-Cool cohortonderzoek*. (pp. 55–66). Amsterdam, the Netherlands: Kohnstamm Instituut. Retrieved from <http://pre-cool.kohnstammstituut.nl/resultaten-en-publicaties.html>
- Verhallen, M. (1994). *Lexicale vaardigheden van Turkse en Nederlandse kinderen. Een vergelijkend onderzoek naar betekenisstoekenning* (Doctoral dissertation). Amsterdam, the Netherlands: Instituut voor Functioneel Onderzoek van Taal en Taalgebruik (IFOTT).
- Verhallen, M., & Schoonen, R. (1993). Lexical knowledge of monolingual and bilingual children. *Applied Linguistics*, 14, 344–363. <https://doi.org/10.1093/applin/14.4.344>
- Verhallen, S. (2009). *Basiswoordenlijst Amsterdamse Kleuters. Toelichting voor belangstellenden*. Amsterdam, the Netherlands: Gemeente Amsterdam, Dienst Maatschappelijke Ontwikkeling. <http://www.onderwijsconsument.nl/wordpress/wp-content/uploads/tekstvoorbelangstellendenwoordenlijstdef.pdf>
- Verhoeven, L. T. (1991). Predicting minority children's bilingual proficiency: Child, family, and institutional factors. *Language Learning*, 41, 205–233. <https://doi.org/10.1111/j.1467-1770.1991.tb00684.x>
- Verhoeven, L. T., & Vermeer, A. (1996). *Taalvaardigheid in de bovenbouw. Nederlands van autochtone en allochtone leerlingen in het basis- en mlk-onderwijs*. Tilburg: Tilburg University Press.
- Vermeer, A. (2015). De context van tweedetaalverwerving en tweedetaalleren. Handboek Nederlands als Tweede taal. In B. Bossers, F. Kuiken, & A. Vermeer (Eds.), *Handboek Nederlands als tweede taal in het volwassenenonderwijs* (pp. 16–47). Bussum, the Netherlands: Coutinho.
- Vermeer, A. (2016). Breadth and depth of vocabulary in relation to L1/L2 acquisition and frequency of input. *Applied Psycholinguistics*, 22, 217–234. <https://doi.org/10.1017/S0142716401002041>
- Viberg, Å. (2001). Age-related and L2-related features in bilingual narrative development in Sweden. In L. Verhoeven, & S. Strömquist (Eds.), *Narrative development in a multilingual context*, (pp. 87–128). Amsterdam, the Netherlands: John Benjamins.

- von Suchodeltz, A., Fäsche, A., Gunzenhauser, C., & Hamre, B. K. (2014). A typical morning in preschool: Observations of teacher-child interactions in German preschools. *Early Childhood Research Quarterly*, 28, 509–519. <https://doi.org/10.1016/j.ecresq.2014.05.010>
- Wallach, G. P. (2008). *Language intervention for school-age students: Setting goals for academic success*. St. Louis, MO: Mosby.
- Wasik, B. A., & Hindman, A. H. (2011). Improving vocabulary and pre-literacy skills of at-risk preschoolers through teacher professional development. *Journal of Educational Psychology*, 103, 455–469. <https://psycnet.apa.org/doi/10.1037/a0023067>
- Weiland, C., Ulvestad, K., Sachs, J., & Yoshikawa, H. (2013). Associations between classroom quality and children's vocabulary and executive function skills in an urban public prekindergarten program? *Early Childhood Research Quarterly*, 28, 199–209 <https://doi.org/10.1016/j.ecresq.2012.12.002>
- Westby, C. E. (2012). Assessing and remediating text comprehension problems. In A. G. Kamhi, & H. W. Catts (Eds.), *Language and reading disabilities* (3rd. ed.), (pp. 163–225). Boston, MA: Pearson.
- Westby, C. E., Van Dongen, R., & Maggart, Z. (1989). Assessing narrative competence. *Seminars in speech and language*, 10, 63–76.
- Whitebook, M. (1989). *Who cares? Child care teachers and the quality of care in America. Final report, National Child Care Staffing Study*. Berkeley, CA: Child Care Employee Project. Retrieved from <https://eric.ed.gov/?id=ED323031>
- Winter, E. (2010). Trajectories of multiculturalism in Germany, the Netherlands and Canada: In search of common patterns. *Government and Opposition*, 45, 166–186. <https://doi.org/10.1111/j.1477-7053.2009.01309.x>
- Wong Fillmore, L. (1983). The language learner as an individual: Implications of research on individual differences for the ESL teacher. In M. A. Clarke, & J. Handcombe (Eds.), *On TESOL '82 Pacific perspectives on language learning and teaching* (pp. 157–174). Washington, DC: TESOL
- Yates, G. C. R., & Yates, S. M. (1990). Teacher effectiveness research: Towards describing user-friendly classroom instruction. *Educational Psychology*, 10, 225–238. <https://doi.org/10.1080/0144341900100304>
- Yeager Pelatti, C., Piasta, S. B., Justice, L. M., & O'Connell, A. (2014). Language- and literacy-learning opportunities in early childhood classrooms: Children's typical experiences and within-classroom variability, *Early Childhood Research Quarterly*, 29, 445–456. <https://doi.org/10.1016/j.ecresq.2014.05.004>

250 *References*

- Yoshikawa, H., Weiland, C., Brooks-Gunn, J., Burchinal, M. R., Espinosa, L. M., Gormley, W. T., ... Zaslow, M. J. (2013). *Investing in our future: The evidence base on preschool education*. Ann Arbor, MI: Society for Research in Child Development. Retrieved from <https://www.fcd-us.org/the-evidence-base-on-preschool/>
- Yoshikawa, H., Weiland, C., & Brooks-Gunn, J. (2016). When does preschool matter? *The Future of Children*, 26(2), 21–35. <https://doi.org/10.1353/foc.2016.0010>
- Zaslow, M., Anderson, R., Redd, Z., Wessel, J., Tarullo, L., & Burchinal, M. (2010). *Quality, dosage, thresholds, and features in early childhood settings: A review of the literature* (OPRE 2011-5). Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. Retrieved from [http://www.acf.hhs.gov/sites/default/files/opre/quality\\_review\\_0.pdf](http://www.acf.hhs.gov/sites/default/files/opre/quality_review_0.pdf)
- Zuniga, S. A., & Howes, C. (2009). Predictions of children's experiences with Latina family child care providers. *Early Education and Development*, 20, 265–284. <https://doi.org/10.1080/10409280802595433>

## APPENDIX 1

---

### **Table with parameters settings Chapter 4**

---

Table 1: Fixed and Random Parameters of the General Development Model to Describe the Development of Receptive Vocabulary + Extra Explanatory Child Characteristic Variables.

	GDM PPVT		GDM + Exposure to Dutch at School	
	Est.	SE	Est.	SE
<b>Fixed Part</b>				
Intercept	64.39	2.30	36.69	3.33
Age	1.60	0.10	1.93	0.26
Age <sup>2</sup>	-0.01	0.05	0.07	0.01
Exposure			2.03	0.21
Age*Exposure			-0.14	0.02
Educational Facility				
Age*Educational Facility				
<b>Random Part</b>				
Variance within individuals				
S <sup>2</sup> Intercept	149.84	20.90	93.78	14.89
S Age, Intercept	-4.42	1.15	-2.19	0.73
S <sup>2</sup> Age	0.21	0.09	0.10	0.07
Variance between individuals				
S <sup>2</sup> Intercept	163.34	42.48	68.06	20.92
S Age, Intercept	-2.78	0.96	-0.48	0.62
S <sup>2</sup> Age				

## APPENDIX 2

---

### **Tables with parameters settings Chapter 5**

---

Table 2.1: Fixed and Random Parameters of the General Development Model to Describe the Development of Number of Different Words (NDW) + Extra Explanatory Child Characteristic Variables.

	General Development Model		GDM + Exposure to Dutch at School		GDM + Educational Facility	
	Est.	SE	Est.	SE	Est.	SE
<b>Fixed Part</b>						
Intercept	36.27	1.48	29.46	3.11	31.29	4.47
Age	0.73	0.07	3.34	1.74	1.22	0.21
Age <sup>2</sup>						
Exposure			0.46	0.19		
Age*Exposure						
Educational Facility					4.02	4.32
Age*Educational Facility					-0.397	0.16
<b>Random Part</b>						
Variance within individuals						
S <sup>2</sup> Intercept	85.90	10.83	85.27	10.74	81.83	10.32
S Age, Intercept						
S <sup>2</sup> Age						
Variance between individuals						
S <sup>2</sup> Intercept	69.96	20.10	60.24	18.00	68.52	19.54
S Age, Intercept						
S <sup>2</sup>						

Table 2.2: Fixed and Random Parameters of the General Development Model to Describe the Development of Guiraud Index Score (GIS) + Extra Explanatory Child Characteristic Variables.

	General Development Model		GDM + Exposure to Dutch at School	
	Est.	SE	Est.	SE
<b>Fixed Part</b>				
Intercept	3.63	0.09	3.14	0.15
Age	0.04	0.00	0.02	0.01
Age <sup>2</sup>				
Exposure			0.04	0.01
Age*Exposure				
Educational Facility				
Age*Educational Facility				
Random Part				
Variance within individuals				
S <sup>2</sup> Intercept	0.41	0.05	0.34	0.05
S Age, Intercept				
S <sup>2</sup> Age				
Variance between individuals				
S <sup>2</sup> Intercept	0.22	0.07	0.13	0.05
S Age, Intercept	-0.01	0.00	-0.00	0.00
S <sup>2</sup> Age				

Table 2.3: Fixed and Random Parameters of the General Development Model to Describe the Development of Measure of Lexical Richness (MLR) + Extra Explanatory Child Characteristic Variables.

	General Development Model		GDM + Exposure to Dutch at School	
	Est.	SE	Est.	SE
Fixed Part				
Intercept	2.25	0.24	1.04	0.44
Age	0.07	0.02	-0.00	0.03
Age^2	0.003	0.001	0.00	0.00
Exposure			0.08	0.03
Age*Exposure				
Educational Facility				
Age*Educational Facility				
Random Part				
Variance within individuals				
S <sup>2</sup> Intercept	0.99	0.21	0.95	0.20
S Age, Intercept	0.11	0.02	0.11	0.02
S <sup>2</sup> Age	0.02	0.02	0.01	0.00
Variance between individuals				
S <sup>2</sup> Intercept	1.89	0.44	1.68	0.39
S Age, Intercept	0.07	0.02	0.06	0.01
S <sup>2</sup> Age				

Table 2.4: Fixed and Random Parameters of the General Development Model to Describe the Development of Story Structure (SS) + Extra Explanatory Child Characteristic Variables.

	General Development Model		GDM + Exposure to Dutch at School	
	Est.	SE	Est.	SE
<b>Fixed Part</b>				
Intercept	7.01	0.33	3.40	0.56
Age	0.24	0.02	0.32	0.05
Age^2	-0.01	0.001	0.01	0.002
Exposure			0.25	0.04
Age*Exposure			-0.02	0.004
Educational Facility				
Age*Educational Facility				
<b>Random Part</b>				
Variance within individuals				
S <sup>2</sup> Intercept	3.76	0.53	3.33	0.47
S Age, Intercept	-0.06	0.02	-0.03	0.02
S <sup>2</sup> Age	0.00	0.00	0.00	0.00
Variance between individuals				
S <sup>2</sup> Intercept	3.11	0.92	1.20	0.47
S Age, Intercept	-0.11	0.04	-0.03	0.02
S <sup>2</sup> Age	0.01	0.003	0.002	0.002

Table 2.5: Fixed and Random Parameters of the General Development Model to Describe the Development of Internal State Terms (IST) + Extra Explanatory Child Characteristic Variables.

General Development Model	
	<b>IST</b>
	Est.      SE
Fixed Part	
Intercept	5.45      0.55
Age	0.14      0.03
Age <sup>2</sup>	
Exposure	
Age*Exposure	
Educational Facility	
Age*Educational Facility	
Random Part	
Variance within individuals	16.82      2.09
S <sup>2</sup> Intercept	
S Age, Intercept	
S <sup>2</sup> Age	
Variance between individuals	
S <sup>2</sup> Intercept	8.81      2.83
S Age, Intercept	-0.24      0.08
S <sup>2</sup> Age	

## APPENDIX 3

---

### **Description of the CLASS**

---

**The ten dimensions from the Classroom Assessment Scoring System (CLASS; Pianta, La Paro, & Hamre, 2008).**

(1) *Positive Climate*; is about the relations between the teacher and pupils and among pupils. It reflects enthusiasm, affection, and respect

(2) *Negative Climate*; is the degree in which there are displays of anger, aggression and/ or harshness in the classroom done by the teacher or the pupils

(3) *Teacher Sensitivity*; is the extent to which teachers are aware of and anticipate on problems and provide comfort, reassurance, and encouragement

(4) *Regard for Student Perspectives*; reflects the extent to which classroom activities are rigidly structured and whether pupils can show initiative, leadership or responsibility

(5) *Behavior Management*; is about the teacher's capacity to use effective methods to prevent and redirect pupils' misbehavior by formulating clear expectations and rules

(6) *Productivity*; is the extent to which pupils' time is well-managed so that they have the chance to be constantly involved in learning activities

(7) *Instructional Learning Formats*; focuses on what a teacher does to maximize pupils' attention, interest, and active engagement in classroom activities

(8) *Concept Development* considers the strategies teachers must promote pupils' higher order thinking skills and creativity

(9) *Quality of Feedback*; concerns the quality of verbal encouragements provided to pupils about their work, comments, and ideas

(10) *Language Modeling*; is about the amount of sustained conversations and open questions in the classroom and how rich the language the teacher uses is

## APPENDIX 4

---

### **Boxplot CLASS**

---

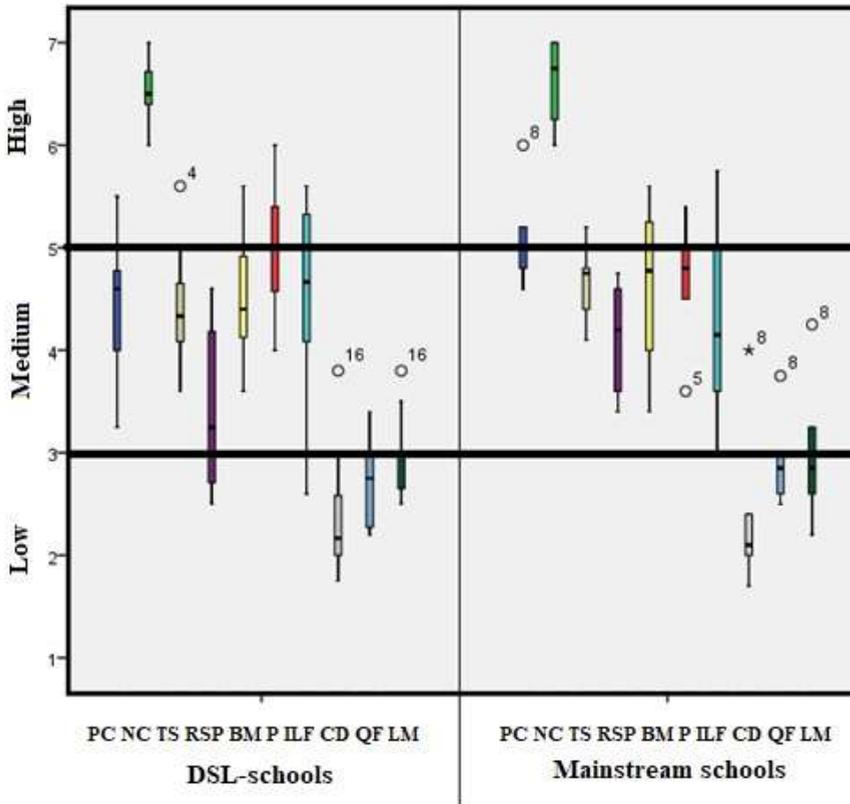


Figure 5.1: Boxplot for the Observed Scores on the Ten Dimensions of CLASS, Separated by School Type. PC = Positive Climate, NG = Negative Climate, TS = Teacher Sensitivity, RSP = Regard for Student Perspectives, BM = Behavior Management, P = Productivity, ILF = Instructional Learning Formats, CD = Concept Development, QF = Quality of Feedback, and LM = Language Modeling.

## APPENDIX 5

---

### **Snapshot items**

---

### **Activities**

*In what kinds of activities are the focal pupils engaged?*

- (1) *Routines*; like cleaning, transitions to a new activity, eating and drinking, changing clothes for sports or out-door activities
- (2) *Mathematical Activities*; like pre-mathematic games, working with mathematic materials or computer games focused on mathematics
- (3) *Language Activities*; like conversations, instructions, vocabulary lessons, social emotional lessons and watching television
- (4) *Literacy activities*; like reading to the pupils, reading of the pupils, lessons focused on learning to discriminate and pronounce sounds, language games (rhyme), pre-literacy activities, working with materials for language, computer games focused on literacy, working with materials for motor skills as in practicing writing
- (5) *Play*; like games, computer games, television for fun, working with construction materials and exploring with sand, water and clay, fantasy game or role play
- (6) *Physical activities*; like sports or playing outdoor in the schoolyard
- (7) *Expressional activities*; like singing, drawing, or drama/ expression lessons
- (8) *Waiting*
- (9) *Other*; like taking a test, nature lessons, evaluation of activities and energizers.

### **Interactions**

*What types of interactions do the focal pupils have?*

- (1) *Teacher-Focal pupil Interaction*; there is interaction between only the teacher and the focal child
- (2) *Peer Interaction*; there is interaction between the focal pupil and (one of his or her) peers
- (3) *Teacher Instruction*; the teacher gives instruction to the whole group or a smaller group with the pupil
- (4) *No interaction or with parent*; the pupil is working alone, is wondering around, or has interaction with his parent

When there was teacher interaction the follow-up question was: *what kinds of teaching interactions are taking place?*

- (1) *Didactic*
- (2) *Scaffolding*

### **Language**

*What kinds of language situations are there?*

- (1) *Dialogue*; there was a dialogue between the observed pupil and his or her teacher

- (2) *Language Situation with Peers*; the focal pupil was engaged in a language situation with (one of his or her) peers
- (3) *Teacher Talk*; the teacher spoke to the whole group or a small group with the focal pupil in this group
- (4) *Balanced Language Situation*; there was a balanced situation where teacher and pupils (including the focal pupil) were speaking to each other, during for example circle time
- (5) *Self-talk*; the focal pupil was speaking to him- or herself
- (6) *No language*; the focal pupil was silent and no one spoke to him or her
- (7) *Other*; like sound from the television, the computer, or from parents

*What kinds of languages are there?*

- (1) *Dutch*
- (2) *Language Other than Dutch*
- (3) *Non-verbal language*
- (4) *Silence*

*What types of language are used by the focal pupils, the peers with whom the focal pupils are interacting, and the teachers?*

- (1) *Complex language*; longer sentences, precise referring, use of infrequent words with explanation
- (2) *Simple language*; which is with short sentences, mostly 1-word sentences. No explicit reference, use of “this”, “that”, and “like this”, use of diminutives, no infrequent words
- (3) *Language Other than Dutch*; the pupils spoke in their home language or a lingua franca (impossible for the observer to decode what was said in that language other than Dutch)
- (4) *Non-verbal*; which could be non-verbal signs like pointing or nodding, or it could contain sounds but no words such as singing, laughing or whistling
- (5) *No language*; the focal pupil is silent



## APPENDIX 6

---

### **Tables with parameters settings Chapter 8**

---

Table 6.1: Fixed and Random Parameters of the General Development Model to Describe the Development of Receptive Vocabulary + Extra Explanatory Learning Environment Characteristic Variables – Focusing on Teacher Behavior (118 cases).

	General Development Model		GDM + Positive Climate	
	Est.	SE	Est.	SE
<b>Fixed Part</b>				
Intercept	68.46	1.910	80.74	14.19
Age	1.11	0.08	-0.02	0.57
Age <sup>2</sup>	-	-	-	-
Positive Climate			-2.68	3.06
Age * Positive Climate			2.47	1.23
<b>Random Part</b>				
Regard for Student Perspectives				
Age * Regard for Student Perspectives				
<b>Random Part</b>				
Variance within individuals				
S <sup>2</sup> Intercept	85.96	17.12	79.47	16.04
S Age, Intercept	-3.75	1.09	-3.47	1.02
S <sup>2</sup> Age	0.23	0.10	0.22	0.09
Variance between individuals				
S <sup>2</sup> Intercept	98.82	27.74	104.20	28.63
S Age, Intercept	-	-	-	-
S <sup>2</sup> Age	-	-	-	-

Table 6.2: Fixed and Random Parameters of the General Development Model to Describe the Development of Receptive Vocabulary – Focusing on Focal Pupils' Experiences (127 cases).

	ppVT	
	Est.	SE
Fixed Part		
Intercept	68.48	1.88
Age	0.99	0.08
Age <sup>2</sup>	-	-
Language Activities		
Age * Language Activities		
Math Activities		
Age * Math Activities		
Language Situation with Peers		
Age * Language Situation with Peers		
Balanced Language Situations		
Age * Balanced Language Situations		
Peer Interaction		
Age * Peer Interaction		
Teacher-Focal Pupil Interaction		
Age * Teacher-Focal Pupil Interaction		
Random Part		
Variance within individuals		
S <sup>2</sup> Intercept	49.85	9.70
S Age, Intercept	-	-
S <sup>2</sup> Age	-	-
Variance between individuals		
S <sup>2</sup> Intercept	125.52	32.45
S Age, Intercept	-2.64	1.12
S <sup>2</sup> Age	0.08	0.07

Table 6.3: Fixed and Random Parameters of the General Development Model to Describe the Development of Number of Different Words (NDW) – Focusing on Teacher Behavior (114 cases).

General Development Model		
	Est.	SE
Fixed Part		
Intercept	37.37	1.64
Age	0.63	0.08
Age <sup>2</sup>		
Positive Climate		
Age * Positive Climate		
Regard for Student Perspectives		
Age * Regard for Student Perspectives		
Random Part		
Variance within individuals	67.44	10.94
S <sup>2</sup> Intercept		
S Age, Intercept		
S <sup>2</sup> Age		
Variance between individuals	76.96	23.19
S <sup>2</sup> Intercept		
S Age, Intercept		
S <sup>2</sup> Age		

Table 6.4: Fixed and Random Parameters of the General Development Model to Describe the Development of Guiraud Index Score (GIS) – Focusing on Teacher Behavior (114 cases).

	GIS	
	Est.	SE
Fixed Part		
Intercept	3.78	0.10
Age	0.04	0.01
Age <sup>2</sup>		
Positive Climate		
Age * Positive Climate		
Regard for Student Perspectives		
Age * Regard for Student Perspectives		
Random Part		
Variance within individuals	0.28	0.04
S <sup>2</sup> Intercept		
S Age, Intercept		
S <sup>2</sup> Age		
Variance between individuals		
S <sup>2</sup> Intercept	0.25	0.08
S Age, Intercept	-0.01	0.00
S <sup>2</sup> Age	0.00	0.00

Table 6.5: Fixed and Random Parameters of the General Development Model to Describe the Development of Measure of Lexical Richness (MLR) + Extra Explanatory Learning Environment Characteristic Variables – Focusing on Teacher Behavior (114 cases).

	General Development Model		GDM + Positive Climate	
	Est.	SE	Est.	SE
<b>Fixed Part</b>				
Intercept	2.73	0.29	-0.95	2.00
Age	0.09	0.02	0.35	0.14
Age <sup>2</sup>				
<b>Positive Climate</b>			0.80	0.43
Age * Positive Climate			-0.06	0.03
<b>Regard for Student Perspectives</b>				
Age * Regard for Student Perspectives				
<b>Random Part</b>				
Variance within individuals				
S <sup>2</sup> Intercept	4.30	0.70	4.26	0.69
S Age, Intercept				
S <sup>2</sup> Age				
Variance between individuals				
S <sup>2</sup> Intercept	1.62	0.74	1.20	0.65
S Age, Intercept				
S <sup>2</sup> Age				

Table 6.6: Fixed and Random Parameters of the General Development Model to Describe the Development of Story Structure (SS) – Focusing on Teacher Behavior (114 cases).

	SS	
	Est.	SE
Fixed Part		
Intercept	7.39	0.32
Age	0.19	0.02
Age <sup>2</sup>	-0.002	0.001
Positive Climate		
Age * Positive Climate		
Regard for Student Perspectives		
Age * Regard for Student Perspectives		
Random Part		
Variance within individuals		
S <sup>2</sup> Intercept	3.79	0.71
S Age, Intercept	-0.12	0.04
S <sup>2</sup> Age	0.01	0.00
Variance between individuals		
S <sup>2</sup> Intercept	2.10	0.73
S Age, Intercept		
S <sup>2</sup> Age		





Table 6.9: Fixed and Random Parameters of the General Development Model to Describe the Development of Measure of Lexical Richness (MLR) + Extra Explanatory Learning Environment Characteristic Variables – Focusing on Focal Pupils' Experiences (162 cases).

	General Development Model		GDM + Peer Interaction	
	Est.	SE	Est.	SE
<b>Fixed Part</b>				
Intercept	2.26	0.25	3.31	0.75
Age	0.07	0.02	0.16	0.04
Age <sup>2</sup>	0.003	0.001	0.002	0.001
Language Activities				
Age * Language Activities				
Math Activities				
Age * Math Activities				
Language Situation with Peers				
Age * Language Situation with Peers				
Balanced Language Situations				
Age * Balanced Language Situations				
Peer Interaction			-3.73	2.51
Age * Peer Interaction			-0.32	0.15
Teacher-Focal Pupil Interaction				
Age * Teacher-Focal Pupil Interaction				
<b>Random Part</b>				
Variance within individuals				
S <sup>2</sup> Intercept	1.02	0.23	1.00	0.22
S Age, Intercept	0.11	0.03	0.10	0.02
S <sup>2</sup> Age	0.02	0.004	0.02	0.003
Variance between individuals				
S <sup>2</sup> Intercept	1.98	0.55	1.882	0.45
S Age, Intercept	0.08	0.03	0.07	0.02
S <sup>2</sup> Age	0.001	0.002	0.00	0.00

Table 6.10: Fixed and Random Parameters of the General Development Model to Describe the Development of Story Structure (SS) + Extra Explanatory Learning Environment Characteristic Variables – Focusing on Focal Pupils' Experiences (162 cases).

	General Development Model		GDM + Balanced Language Situations		GDM + Teacher Focal Pupil Interaction	
	Est.	SE	Est.	SE	Est.	SE
Fixed Part						
Intercept	6.91	0.34	5.42	0.76	4.94	0.71
Age	0.21	0.02	0.21	0.02	0.21	0.02
Age <sup>2</sup>	-0.002	0.001	-0.002	0.001	-0.002	0.001
Language Activities						
Age * Language Activities						
Math Activities						
Age * Math Activities						
Language Situation with Peers						
Age * Language Situation with Peers						
Balanced Language Situations			7.27	3.40		
Age * Balanced Language Situations			-	-		
Peer Interaction						
Age * Peer Interaction						
Teacher-Focal Pupil Interaction					8.66	3.22
Age * Teacher-Focal Pupil Interaction					-0.45	0.18
Random Part						
Variance within individuals	4.16	0.54	4.17	0.54	3.95	0.51
S <sup>2</sup> Intercept						
S, Age, Intercept						
S <sup>2</sup> Age						
Variance between individuals	2.75	0.85	2.36	0.77	2.11	0.70
S <sup>2</sup> Intercept						
S, Age, Intercept						
S <sup>2</sup> Age						



---

## Nederlandse samenvatting

---

### **Hoofdstuk 1: Introductie van de Nederlandse onderwijscontext voor nieuwkomers in de kleuterleeftijd**

Als leerkracht op een school voor nieuwkomers heb ik me veel bezig gehouden met de ontwikkeling van het onderwijs voor nieuwkomers in de kleuterleeftijd. Met nieuwkomers bedoelen we in deze dissertatie: (1) de leerling is niet in Nederland geboren, (2) de leerling heeft nog geen onderwijs in het Nederlands ontvangen, (3) de leerling was minimaal vier en maximaal zes jaar oud, en (4) de leerling was aan het begin van de studie niet langer dan één jaar in Nederland. De aanleiding voor deze dissertatie is het feit dat veel scholen voor nieuwkomers geen aparte klassen hebben voor nieuwkomers in de kleuterleeftijd. De vraag is of het voor kleuters die nieuw zijn in Nederland en het Nederlands nog moeten verwerven geen meerwaarde zou hebben als ze eerst in een speciale taalvoorziening les krijgen. Als leerkracht zag ik de taalontwikkeling die deze kleuters doormaakten op een tijdelijke speciale school voor nieuwkomers, maar ik vroeg me af of ze een soortgelijke ontwikkeling zouden doormaken op een reguliere Nederlandse basisschool.

Globalisatie heeft in Nederland, net als in veel andere landen in de wereld, effect op de samenstelling van de leerlingenpopulatie op school. In bijna elke klas in Nederland zit tegenwoordig wel een leerling voor wie het Nederlands niet de thuistaal is en waarvoor Nederlands niet de eerste taal was die de leerling leerde. Er is in Nederland al veel onderzoek gedaan naar de Nederlandse taalontwikkeling van deze tweedetaalleerders, maar meestal waren dit leerlingen die wel in Nederland geboren zijn en waarvan we aan kunnen nemen dat zij vóór ze naar school gingen, op vier- of vijfjarige leeftijd, al een beetje kennis van het Nederlands hadden opgedaan. In deze dissertatie ligt de focus op kleuters met een andere taalachtergrond dan het Nederlands die nog geen jaar in Nederland waren toen zij met school en tegelijk ook met het leren van het Nederlands begonnen.

Zoals gezegd gaat niet elke nieuwkomer in de kleuterleeftijd tijdelijk naar een speciale school voor nieuwkomers. Er zijn grofweg twee soorten scholen waar een nieuwkomer in de kleuterleeftijd naar toe kan gaan, afhankelijk van de keuze van het schoolbestuur in een bepaalde regio:

- (1) Naar een school met speciale voorzieningen voor leerlingen die het Nederlands als tweede taal leren (*DL2-school* genaamd in deze dissertatie):
  - a. in een aparte taalschool, of
  - b. in een aparte taalklas binnen een reguliere school.

- (2) Naar een reguliere kleuterklas op een basisschool (waar wellicht speciale aandacht gegeven wordt aan hun meertalige ontwikkeling; *Mainstream school* genaamd in deze dissertatie).

De keuze voor een van deze twee vormen is vrij voor elk schoolbestuur in Nederland, maar het is onduidelijk op basis waarvan deze keuze gemaakt wordt. Maarse en Muller (2017) hebben dit recentelijk onderzocht en zagen dat die besluiten voornamelijk gebaseerd zijn op praktische overwegingen betreffende de logistiek, infrastructuur en het leerlingaantal. Het lijkt erop dat de overweging niet gemaakt wordt op basis van wat het beste is voor de ontwikkeling van deze leerlingen (Vermeer, 2015).

Het type school waar de nieuwkomer in de kleuterleeftijd onderwijs volgt, zou dus per stad kunnen verschillen en de vraag is in hoeverre deze twee typen scholen van elkaar verschillen qua pedagogische en didactische omstandigheden. Daarom staat in deze dissertatie de volgende onderzoeksvraag centraal:

*In welke mate dragen pedagogische en didactische omstandigheden bij aan de tweede-taalontwikkeling van nieuwkomers in de kleuterleeftijd in de eerste twee en een half jaar na aankomst in Nederland.*

Om deze vraag te beantwoorden, was het nodig om een interdisciplinaire aanpak te hanteren aangezien de taalontwikkeling van de leerlingen onderzocht moest worden vanuit de taalkundige kant, terwijl de pedagogische en didactische omstandigheden werden onderzocht vanuit de kant van de pedagogiek. We wilden namelijk niet alleen weten *hoe* de tweede taal van jonge nieuwkomers zich ontwikkelde, ook wilden we kijken naar de *mogelijke invloed van de leeromgeving* op deze taalontwikkeling.

## **Hoofdstuk 2: Theoretische perspectieven**

In Hoofdstuk 2 van deze dissertatie wordt het theoretisch kader van de studie beschreven. De interdisciplinaire aanpak wordt ondersteund door de ecologische kijk op ontwikkeling van een leerling. Verder worden de keuzes in de datacollectie onderbouwd aan de hand van de *communicative competence* theorie.

Een ecologische kijk op de ontwikkeling van een leerling is gebaseerd op werk van Bronfenbrenner (1979) waarin hij wijst op het betrekken van de bredere omgeving van de leerling bij het onderzoek naar de ontwikkeling van leerlingen. Een onderdeel van die bredere omgeving is de leeromgeving en dus is het verstandig om factoren uit die context, zoals leerkrachtgedrag en peer-interactie mee te nemen in het taalkundige onderzoek.

De communicative competence theorie borduurt hier eigenlijk op voort. Met deze theorie willen onderzoekers taal eigenlijk op een andere manier definiëren en onderzoek naar taalontwikkeling verbreden. Waar eerst voornamelijk de focus lag op taalkundige

aspecten (zoals woordenschat en grammatica) voegt de benadering van communicative competence meer sociale aspecten toe, omdat zonder die sociale aspecten de grammaticaregels weinig zin hebben. Het is namelijk belangrijk om tijdens de communicatie rekening te houden met wie en tegen wie er gesproken wordt, wat de relatie tussen de mensen is en wat de context en de bedoeling van de communicatie is.

Communicative competence is dus een aanpak waarbij verschillende competenties in samenhang worden bekeken. In deze dissertatie staan voornamelijk *linguistic competence* (de vaardigheid om goede zinnen te maken met de juiste woordenschat en grammaticale constructies) en *discourse competence* (de vaardigheid om een coherent verhaal te vertellen dat te volgen is voor de luisteraar) centraal. Er is ook data verzameld over de *sociocultural competence* en *strategic competence*, maar binnen de begrenzings van dit onderzoek hebben we deze niet kunnen analyseren.

### **Hoofdstuk 3: Participanten in het onderzoek**

In de datacollectie van Hoofdstuk 4 tot en met 7 wordt gebruikt gemaakt van data van dezelfde basisscholen, leerkrachten en leerlingen. In Hoofdstuk 3 wordt beschreven hoe de participerende scholen, leerkrachten en leerlingen benaderd zijn. Ook worden de algemene karakteristieken van de participanten gepresenteerd. Er deden aan het begin van het onderzoek, in het eerste intensieve jaar van de dataverzameling 10 scholen (waarvan 5 DL2-scholen), 17 leerkrachten (waarvan 11 DL2 leerkrachten) en 42 leerlingen (waarvan 32 op een DL2-school) mee.

### **Hoofdstuk 4 & 5: De Nederlandse taalontwikkeling van jonge nieuwkomers**

In Hoofdstuk 4 en 5 onderzoeken we de Nederlandse taalontwikkeling van nieuwkomers in de kleuterleeftijd over een periode van tweeënehalf jaar. Omdat ontwikkeling gemeten wordt, houden we rekening met de variabele *leeftijd* maar we voegen ook de variabele *blootstelling* toe omdat we willen zien of leerlingen die langer in Nederland zijn een andere score of ontwikkeling hebben dan leerlingen die korter in Nederland zijn. Bovendien kijken we of die Nederlandse taalontwikkeling van deze tweedetaalleerders anders verloopt per *schooltype*. Oftewel, is de ontwikkeling anders bij leerlingen op een speciale taalschool in vergelijking met leerlingen op een reguliere basisschool?

Hoofdstuk 4 gaat over de ontwikkeling van de receptieve woordenschat (hoeveel woorden de leerlingen begrijpen). In Hoofdstuk 4 staat deze vraag centraal: hoe ontwikkelt zich de receptieve woordenschat van nieuwkomers in de kleuterleeftijd in de eerste tweeënehalf jaar na aankomst in Nederland, in relatie tot schooltype. De receptieve woordenschat wordt gemeten aan de hand van de Nederlandse versie van de *Peabody Picture Vocabulary Task* (PPVT; PPVT-III-NL; Schlichting, 2005). Bij deze taak horen

leerlingen een woord en moeten ze uit vier opties het bijbehorende plaatje kiezen. De uitkomst van onze analyse is dat de receptieve woordenschat van de leerlingen groeit; hoe ouder de leerlingen hoe hoger de scores op de PPVT, maar dat er wel een afname in groei is te zien. Ditzelfde geldt voor de invloed van blootstelling (gemeten aan de hand van het aantal maanden op een Nederlandse basisschool); hoe meer blootstelling hoe hoger de scores op de PPVT, maar er is een afname in groei te zien. We hebben geen effect gevonden van schooltype; de leerlingen in beide schooltypen lieten dezelfde scores en eenzelfde ontwikkeling in receptieve woordenschat zien.

Hoofdstuk 5 gaat over de narratieve capaciteiten van deze leerlingen: kunnen ze een duidelijk en compleet verhaal vertellen? In Hoofdstuk 5 staat deze vraag centraal: hoe ontwikkelt zich de narratieve vaardigheid van nieuwkomers in de kleuterleeftijd in de eerste tweeënehalf jaar na aankomst in Nederlands, in relatie tot het schooltype? De narratieve vaardigheid van de participanten wordt gemeten met de *Multilingual Assessment Instrument for Narratives* (MAIN; Gagarina et al., 2012). Met dit instrument kun je de narratieve vaardigheid op verschillende niveaus meten. Een verhaal kan namelijk geanalyseerd worden op microstructuur en op macrostructuur. *Microstructuur* richt zich onder andere op het woordniveau en wij keken naar de lexicale diversiteit van deze woorden (hoeveel verschillende woorden worden er gebruikt in het verhaal) en naar de lexicale rijkdom van de woorden (gebruikt een leerling hoog frequente woorden of laag frequente woorden?). *Macrostructuur* richt zich op het verhaal zelf; is het verhaal begrijpelijk, compleet en coherent? De macrostructuur van een verhaal is meer of minder compleet of complex als de volgende onderdelen in het verhaal zitten: een setting (met plaats- en tijdsaanduiding), een doel, een poging, een uitkomst en termen van interne toestand (worden de gedachten en gevoelens van de verhaalkarakters onder woorden gebracht?).

Voor vijf van de zes onderdelen van narratieve vaardigheid die we hebben onderzocht konden we een ontwikkelingsmodel bouwen. Dit waren de drie maten van de microstructuur (het aantal verschillende woorden, de Guiraud index score en de maat voor lexicale rijkheid<sup>25</sup>) en twee van de drie maten van macrostructuur (verhaal structuur en interne toestand termen<sup>26</sup>). Alle vijf deze onderdelen lieten groei zien over de tijd: de leerlingen scoorden hoger naar mate ze ouder werden. Voor twee onderdelen bleek dat de groei wel afvlakte in de loop der tijd. Voor vier van de vijf onderdelen bleek de blootstelling aan Nederlands er ook toe te doen. Voor één onderdeel, het aantal

---

<sup>25</sup> In het proefschrift worden de volgende Engelse termen hiervoor gebruikt: het aantal verschillende woorden = Number of Different Words, NDW; de Guiraud index score = Guiraud Index Score, GIS; en de maat voor lexicale rijkheid = Measure of Lexical Richness, MLR.

<sup>26</sup> In het proefschrift worden de volgende Engelse termen hiervoor gebruikt: verhaal structuur = Story Structure, SS; interne toestand termen = Internal State Terms, IST.

verschillende woorden in een verhaal, vonden we een effect van schooltype. Er bleek een interactie effect tussen schooltype en leeftijd te zijn op het gebied van het aantal verschillende woorden. Het effect van schooltype op het aantal verschillende woorden in een verhaal was groter voor jonge leerlingen dan voor oudere leerlingen, maar doordat er geen hoofdeffect was, de standaard error was te groot, zou de interpretatie van dit interactie-effect onbetrouwbaar zijn. Om het interactie-effect beter te kunnen interpreteren zou een grotere sample wellicht de oplossing zijn.

Hoofdstuk 4 en 5 samenvattend, voor zes van de zeven taalonderdelen die we onderzocht hebben, konden we groei vaststellen over de periode van tweeënehalf jaar. De hoeveelheid blootstelling aan het Nederlands bleek er toe te doen en de invloed van schooltype was erg klein. We vonden op maar één van de zeven onderdelen een effect van schooltype, welke te onbetrouwbaar was om te interpreteren. .

## **Hoofdstuk 6 & 7: Observaties van de leeromgeving**

In Hoofdstuk 6 en 7 onderzoeken we de leeromgeving van nieuwkomers in de kleuterleeftijd in het eerste jaar dat ze op een Nederlandse school zitten. Hoofdstuk 6 focust op het gedrag van de leerkracht en wordt gemeten met de *Classroom Assessment Scoring System* (CLASS; Pianta, La Paro, & Hamre, 2008). De vraag die centraal staat in Hoofdstuk 6 is: wat zijn de verschillen in karakteristieken van de leeromgeving van nieuwkomers in de kleuterleeftijd als het gaat om leerkracht- gedrag tussen DL2-scholen en Mainstream scholen? Het algemene patroon van leerkracht-gedrag, gemeten met de CLASS, was dat leerkrachten hoog scoren wat betreft *Emotional Support* (ze weten bijvoorbeeld een goede sfeer te creëren in de klas, goede relaties aan te gaan met de leerlingen en geven ruimte aan initiatieven van de leerlingen), maar ook dat leerkrachten lager scoren op *Instructional Support* (leerkrachten hebben moeite met goede feedback te geven, echte gesprekken te stimuleren, op een goede manier vragen te stellen en een rijk taalaanbod te geven). Al was het algemene patroon vergelijkbaar met ander Nederlands en internationaal onderzoek, de scores waren wel lager dan in vergelijkbare studies. CLASS bestaat uit tien onderdelen en op twee van deze onderdelen was er een significant verschil tussen leerkrachten op een DL2-school en op een Mainstream school. Leerkrachten op een Mainstream school creëerden over het algemeen een hoger positief klimaat in de klas en hielden meer rekening met het kind-perspectief dan leerkrachten op DL2-scholen.

In Hoofdstuk 7 richten we ons op de specifieke activiteiten, interacties en taal rondom de focus leerlingen. We gebruiken hiervoor een snapshot-methode waarbij we steeds de leerling voor een periode van 10 seconde observeren en we daarna een vragenlijst op de laptop invullen. De vraag die centraal staat in Hoofdstuk 7 is: wat zijn de verschillen in karakteristieken van de leeromgeving van nieuwkomers in de kleuterleeftijd als het gaat om de ervaringen van de focus-leerlingen tussen DL2-scholen en Mainstream scholen? We benoemen hier alleen de activiteiten, taalsituaties en interactie-soorten waarvoor

significante verschillen werden gevonden tussen de twee schooltypes. Aan de ene kant deden leerlingen op een DL2-school significant meer Taalactiviteiten en Rekenactiviteiten in vergelijking met leerlingen op een Mainstream school. Leerlingen op een DL2-school ervoeren ook significant vaker taalsituaties waarin er een balans was tussen inbreng van de leerkracht en de leerlingen en de leerlingen ervoeren ook meer interacties met alleen de leerkracht dan leerlingen in Mainstream scholen. Aan de andere kant ervoeren leerlingen op een Mainstream school vaker taalsituaties met alleen klasgenoten en vaker interactie met alleen klasgenoten dan leerlingen op een DL2-school.

Al met al kan er gezegd worden dat de pedagogische praktijken in de DL2-scholen meer leerkracht-gericht waren en op een Mainstream school meer leerlinggericht. De leerkrachten op een DL2-school leidden de activiteiten met minder aandacht voor het leerling-perspectief en leerling initiatieven dan op een Mainstream school. Ook was er op een DL2-school meer interactie tussen de leerkrachten en de leerlingen dan op Mainstream scholen, terwijl er op een Mainstream school weer meer interactie was tussen leerlingen onderling.

## **Hoofdstuk 8: Relatie tussen de leeromgeving en de Nederlandse taalontwikkeling**

In Hoofdstuk 8 worden de uitkomsten uit Hoofdstuk 4 en Hoofdstuk 5 over de Nederlandse taalontwikkeling van de jonge nieuwkomers gerelateerd aan de uitkomsten van de observaties in de klas uit Hoofdstuk 6 en 7. Hiermee willen we antwoord geven op de vraag in hoeverre de verschillen in de karakteristieken van de leeromgeving gerelateerd kunnen worden aan de ontwikkeling van de receptieve woordenschat dan wel aan de ontwikkeling van de narratieve vaardigheden van nieuwkomers in de kleuterleeftijd in de eerste tweeënehalf jaar na aankomst in Nederland. Er kon gekeken worden naar de invloed van de leeromgeving op vijf van de zeven taalvaardigheidsscores. De uitkomsten zijn erg wisselend en de effecten niet heel erg groot: geen enkele variabele van de leeromgeving heeft een effect op meer dan twee taalvaardigheidsscores. Het meest opvallende is dat de twee variabelen die te maken hebben met de betrokkenheid van de leerkracht in de taalsituaties en interacties van invloed zijn op zowel microstructuur als macrostructuur.

## **Hoofdstuk 9: Samenvatting, discussie en conclusies**

In het laatste hoofdstuk worden alle resultaten van onze analyses van zowel de taalontwikkeling, de leeromgeving en de samenhang tussen deze twee samengevat. In Hoofdstuk 9 worden ook aanbevelingen gedaan voor de onderwijspraktijk. Om te beginnen is het aan te bevelen om leerkrachten beter te begeleiden bij het verbeteren van hun vaardigheden in het *Instructional Support* domein, oftewel, leerkrachten kunnen zich verbeteren wat betreft het gebruik van *scaffolding*, het gebruik van goede vragen, het geven

van feedback, het aangaan van echte gesprekken en het de leerlingen voorzien van taalrijke input.

Verder wordt er ook een aanbeveling gedaan wat betreft het ruimte geven aan de thuishalen van de leerlingen. Tijdens de observaties in deze studie werd het gebruik van een andere taal dan het Nederlands nauwelijks aangetroffen en als leerlingen onderling een andere taal gebruikten, werd dat vaak ontmoedigd. Internationaal onderzoek heeft uitgewezen dat het waarderen en functioneel inzetten van de thuistaal verschillende voordelen kan hebben op de ontwikkeling van leerlingen, maar daar wordt in het Nederlands onderwijssysteem nog nauwelijks ruimte voor gegeven.

Het onderzoek in deze dissertatie heeft noodzakelijkerwijs zijn begrenzingen gekend. In vervolgonderzoek zou gekeken kunnen worden of de selectie en de grootte van de steekproef aangepast kunnen worden. Een groter aantal leerlingen, beter verdeeld over de verschillende schooltypen zou er voor kunnen zorgen dat de betrouwbaarheid van de data wordt vergroot. Bovendien zou het aan te bevelen zijn om in vervolgonderzoek de leeromgeving van de leerlingen over een langere tijd in beeld te brengen. De leerlingen veranderen van school, van groep, van leerkracht en de vraag is hoe stabiel de leeromgeving daardoor is. Metingen door het jaar heen in de verschillende klassen kunnen een nog beter beeld geven van de invloed van de leeromgeving op de taalontwikkeling van leerlingen.

Concluderend kunnen we zeggen dat deze dissertatie bijdraagt aan het beter begrijpen van tweede-taalontwikkeling en dan vooral van de ontwikkeling van de receptieve woordenschat en de narratieve vaardigheden van nieuwkomers in de kleuterleeftijd. We hebben ontdekt dat deze taalvaardigheden en de onderdelen hiervan verschillende maten van groei laten zien en dat kindfactoren deze groei beïnvloeden. We hebben ook ontdekt dat bepaalde factoren van de leeromgeving (het hebben van een positief klimaat in de klas en de percentages voor taalactiviteiten, rekenactiviteiten, taalsituaties met klasgenoten, gebalanceerde taalsituaties, klasgenoot interactie en leerkracht-focus leerling interactie) van invloed zijn op de ontwikkeling van deze taalvaardigheden. Al kwam hier niet een consistent profiel uit, want terwijl het ene aspect van de leeromgeving invloed had op een bepaalde vaardigheid, had een ander aspect van de leeromgeving weer invloed op een andere vaardigheid.

Ons onderzoek kan als startpunt dienen voor vervolgonderzoek naar de specifieke karakteristieken van de leeromgeving van nieuwkomers. Naar onze mening zouden in zulk vervolgonderzoek de volgende twee vragen centraal moeten staan: (1) welke factoren in de leeromgeving zijn het *meest belangrijk* voor de taalontwikkeling van jonge nieuwkomers? (2) is het mogelijk om deze factoren onderdeel te laten uitmaken van de leeromgeving van jonge nieuwkomers *ongeacht het type school* waar deze leerlingen deel van uitmaken?

Naar ons idee zou het goed zijn als het onderwijsveld, in samenwerking met onderzoekers, leermiddelen en observatie-instrumenten gaat ontwikkelen gericht op

onderwijs aan nieuwkomers. Deze materialen zouden gebaseerd moeten zijn op “good practices” en gericht moeten zijn op het inspelen op de specifieke onderwijsbehoeften van (jonge) nieuwkomers. Verder zouden deze materialen centraal moeten staan bij lerarenopleiders van de PABO en bij nascholingen. In de komende tien jaar zullen er steeds meer tweedetaalleerders zijn en dus zou elke leerkracht moeten weten hoe je een goede leeromgeving voor deze leerlingen kunt creëren. Ook zouden alle leerkrachten op de hoogte moeten zijn van hoe je het best de taalontwikkeling van deze leerlingen kunt volgen en hoe je rekening kunt houden met de specifieke achtergronden van deze leerlingen zodat hun tweede-taalontwikkeling het best tot zijn recht kan komen.

---

## Curriculum Vitae

---

Frederike Christine Groothoff completed her bachelor's degree in Education in 2004. After teaching at a primary school in Aruba, she returned to Utrecht to study Language and Culture Studies until 2007. During her subsequent research master (2007-2011) in Language and Cognition based in Nijmegen and Tilburg, she worked as a student assistant at the Center for Language and Cognition in Groningen within Prof. Dr. Petra Hendriks' project "Asymmetries in Grammar." In 2014 at Utrecht University, she began her PhD project on second language development of newly arrived migrant kindergarteners with a teacher promotion grant awarded by the Dutch scientific organization NWO. During her PhD research she was a part-time teacher at Taalschool Utrecht, a primary school for newly arrived migrant pupils. In 2017 Frederike successfully applied for a Small Grant for Doctoral Research in Second or Foreign Language Assessment by the TOEFL. In order to expand her knowledge about second language development and to work on the transcription of her data, she visited Johanne Paradis, Elena Nicoladis, and Anna Kirova at the University of Alberta in Edmonton, Canada. Frederike gave numerous presentations at national and international conferences, regularly gave guest lectures for courses on second language development and multilingualism, and contributed to several research projects with her knowledge about the education of multilingual pupils. Since 2018 she is the owner of LangWhich, a company that advises parents, teachers, and policy makers about second language development and multilingual practices. As a freelance consultant, she trains teachers, gives workshops, and develops programs to improve the cultural and language appropriate practices in Dutch primary schools. In 2019 Frederike started to work as a peripatetic language support advisor to improve the teacher skills and language policies at schools in Amstelveen, an area of the Netherlands with a high number of multilingual pupils.







The population in primary schools in the Netherlands is rapidly changing. Teachers welcome more and more multilingual pupils into their classes. Researchers all over the world investigate the language development of second language learners from different angles, but often only include 2<sup>nd</sup> or 3<sup>rd</sup> generation migrant pupils or pupils who already have had considerable exposure to the second language. This dissertation investigates specifically the second language development of newly arrived migrant kindergarteners in different educational settings in the Netherlands (separate language schools, language classes, and mainstream schools).

In this dissertation, data is collected via a receptive vocabulary task, a narrative instrument, and classroom observations. These different methods were used in order to gather a variety of language data, but also to be able to take into account the pedagogical practices these pupils encounter in their classes.

Results show that the various language aspects that were analyzed have different developmental trajectories and that age and amount of exposure influence second language development of newly arrived migrant kindergarteners. Furthermore, the outcomes suggest that the organizational structure of a school as such is less influential than teacher behavior and pedagogical practices in the classroom.

ISBN: 978-94-6332-595-0