PhD-FLSHASE-2019-14
The Faculty of Language and Literature,
Humanities, Arts and Education

Faculty of Medicine, Dentistry and Health
Department of Human Communication Sciences

DISSERTATION

Defence held on 21/06/2019 in Esch-sur-Alzette

to obtain the degrees of

DOCTEUR DE L'UNIVERSITÉ DU LUXEMBOURG
EN PSYCHOLOGIE

AND

DOCTOR IN PHILOSOPHY
IN HUMAN COMMUNICATION SCIENCES

by

Cyril WEALER
Born on 21 May 1986 in Luxembourg (Luxembourg)

EARLY LITERACY DEVELOPMENT IN A MULTILINGUAL EDUCATIONAL CONTEXT: A QUASI-EXPERIMENTAL INTERVENTION AND LONGITUDINAL STUDY

Dissertation defence committee

Dr Pascale Engel de Abreu, dissertation supervisor
Associate Professor, Université du Luxembourg

Dr Silke Fricke, dissertation supervisor
Senior Lecturer, The University of Sheffield

Dr Samuel Greiff, Chairman
Professor, Université du Luxembourg

Dr Jenny Thomson, Vice Chairman
Senior Lecturer, The University of Sheffield

Dr Paul Leseman
Professor, University Utrecht

Dr Claudine Bowyer-Crane
Senior Lecturer, University of York

Abstract

Based on concerns about literacy difficulties experienced by children learning to read and write in a second language, repeated calls have been made for more research on literacy development in multilingual educational settings. Enhanced understanding of literacy development in a second language is essential to optimize support structures for children learning to read and write in a language they have yet to fully acquire. The current thesis presents two longitudinal studies contributing towards this aim. Both studies were conducted with young children growing up in Luxembourg, a linguistically and culturally diverse country where the language spoken in preschool is Luxembourgish, but children learn to read and write in German in Grade 1.

Study 1 was a quasi-experimental intervention study exploring the efficacy of a classroom-based early literacy intervention. Children from 28 preschool classes (age 5-6) were allocated to either the intervention (n = 89) or a standard curriculum (control) group (n = 100). Classroom teachers delivered four intervention sessions (20 minutes each) per week over 12 weeks (48 sessions in total) to their whole classes. The intervention programme targeted phonological awareness, letter-sound knowledge and print awareness embedded in a language and literacy-rich context. All children were assessed before and immediately after the intervention in preschool, and at a nine months follow-up in Grade 1 (age 6-7) after having started formal literacy instruction in German for five months. The intervention group significantly outperformed the control group on early literacy measures immediately postintervention in preschool and the results generalised to measure of reading comprehension and spelling in Grade 1. The study provides clear evidence for the efficacy of the early literacy intervention, particularly for a sub-population of children with low oral language skills in Luxembourgish, many of whom were second language learners.

Study 2 was a correlational study on a subsample of the children from Study 1 (from untrained control group). The aim was to identify preschool predictors in Luxembourgish of literacy skills in German in Grade 1 for multilingual children learning German as a second language. Ninety-eight children completed measures of potential predictors in preschool (age 5-6), including phonological awareness, letter-sound knowledge, rapid automatized naming, verbal short-term memory and vocabulary knowledge in Luxembourgish, along with measures of word reading, reading comprehension and spelling in German in Grade 1 (age 6-7). While moderate to strong correlations were found between all individual preschool predictors and later literacy measures, only phonological awareness, and letter-sound knowledge emerged as unique predictors of all literacy measures. These findings suggest that, despite individual differences, learning to read in a second language may be in many aspects similar to learning to read in a first language.

Taken together, the findings of this thesis represent important steps in extending the theoretical knowledge base on second language literacy acquisition and in strengthening the evidence base for identification and prevention strategies of literacy difficulties in linguistically diverse children in Luxembourg.

Acknowledgments

First and foremost, my most profound gratitude goes to my supervisors Pascale Engel de Abreu and Silke Fricke. Your enthusiasm for both scientific excellence and real world change are unmatched and have been highly inspirational on this journey. I feel incredibly grateful that you gave me the opportunity to grow as an independent researcher and person under your wings. Thank you for making this a truly fascinating and transformative experience.

A special *grande obrigado* to the Polilux team, Carolina Nikaedo, Rute Tomás, Ariana Loff. Thank you for all your support. A special *obrigadissimo* to you Ariana for your words of encouragement. Maybe you did not realize it, but you were my lighthouse on this journey.

A thank you to all my fellow PhD students in Sheffield and Luxembourg. Special thanks to Dea Nielsen for her support in my early PhD days, and to Lucy Dyson and to Katharina Albrecht and her husband for always accommodating me at their homes during my Sheffield stays.

My deepest gratitude also goes to the children and caregivers involved in my research. This project would not have been possible without you and it was a great pleasure working with all of you. My most heartfelt appreciation goes to the teachers of the schools who invested their time and energy to make this research possible. It lightens my heart too see that the future of our children is in such caring hands.

I would like to thank the *Lions Club Mondorf-les-Bains* for their generous donation to support the development of the educational material within the context of the intervention study.

A big high five to all my closest friends (especially Georges Bley and Steph Colling), the members of my Basketball team and my Golf buddies for distracting me from academic work when I hit my cognitive limits. Thanks for reminding me that happiness existed during those rare moments I had struggled to find it in journal articles.

Finally, I would like to thank my family. Thank you mum, for planting the seed of

curiosity in me. Thank you dad, for showing me how to grow this seed into a strong tree by cultivating it with hard work and dedication. Big shout out to my three brothers as well, whose opinion I sometimes value more than my own… sometimes.

Conferences

The research referred to in this thesis has been presented at the following invited talks and conferences.

Invited talks

Wealer, C. (May, 2019). Multilingualism: the cognitive perspective. *Multilingualism, Migration and Spaces of Belonging*. Centre for Luxembourg Studies, Sheffield, (UK).

Wealer, C. (November, 2018). Mehrsprachigkeit im Fokus [Focus on multilingualism]. Opening speech of the *23. Therapietage Bad Nenndorf* [*23rd Therapy Days Bad Nenndorf,* professional development for speech and language therapists]. Institute Schlaffhorst-Andersen, Bad-Nenndorf, (DE).

Wealer, C. (May, 2016). Early literacy instruction in a multilingual setting. *Echoing Voices: Multilingual Educational Experiences*. Centre for Luxembourg Studies, Sheffield, (UK).

Conferences

Wealer, C., Fricke, S. & Engel de Abreu, P. (July 2019). Promoting literacy development in linguistically diverse learners: an early-literacy intervention study. Paper presented at the *Child Language Symposium (CLS2019).* Sheffield, (UK).

Wealer, C., Fricke, S. & Engel de Abreu, P. (May 2019). Developing literacy in linguistically diverse children: a pre-literacy intervention. Paper presented at the conference on *Bilingual Acquisition of Language and Literacy* (*BiALL*), Berlin, (DE).

Wealer, C., Fricke, S. & Engel de Abreu, P. (September 2018). Developing literacy in a multilingual setting. Paper presented at the symposium on the effects of bilingualism on cognitive development and learning at the *51st Congress of the German Association of Psychology (DGP)*, Frankfurt, (DE).

Wealer, C., Fricke, S. & Engel de Abreu, P. (July 2018). Developing literacy in language-minority children: a pre-literacy intervention study in Luxembourg. Paper presented at the *25th Annual Meeting of the Society for the Scientific Study of Reading (SSSR)*, Brighton, (UK).

Wealer, C., Fricke, S. & Engel de Abreu, P. (August 2017). A pre-literacy intervention study in a multilingual setting. Paper presented at the *17th Biennial Conference Meeting of the European Association for Research on Learning and Instruction (EARLI)*, Tampere, (FI).

Wealer, C. Fricke, S. & Engel de Abreu, P. (July 2015). A pre-literary intervention in a multilingual setting. Poster presented at the *PGR Conference*, University of Sheffield, Sheffield, (UK).

Contents

Abstract iii
Acknowledgements v
Conferences vii
Contents ix
List of Tables xiii
List of Figures xvii
List of Abbreviations xix

CHAPTER I - Framework of literacy development 4
1.1 Literacy acquisition in Luxembourg 4
1.2 The framework of reading 6
1.3 The Simple View of Reading 7
1.4 Framework of spelling 8
1.5 Development of reading and spelling 10
1.6 External factors influencing literacy acquisition 12
1.6.1 Socio-economic factors influencing literacy acquisition 13
1.6.2 Orthographic differences influencing literacy acquisition 14

STUDY ONE
FOSTERING LITERACY DEVELOPMENT IN LINGUISTICALLY DIVERSE LEARNERS:
A QUASI-EXPERIMENTAL INTERVENTION STUDY

CHAPTER II - Literature review: Intervention study 17
2.1 Foundational skills of reading and spelling 20
2.2 Foundational skills of reading and spelling in L2 learners 21
2.3 Supporting early literacy skills 23
2.3.1 Supporting phonological awareness in L1 learners 24
2.3.2 Supporting phonological awareness in L2 learners 27
2.3.3 Supporting letter-sound knowledge in L1 learners 28
2.3.4 Supporting letter-sound knowledge in L2 learners 28
2.3.5 Supporting phonological awareness and letter-sound knowledge in L1 learners 28
2.3.6 Supporting phonological awareness and letter-sound knowledge in L2 learners 30
2.4 Summary of effective early literacy interventions 32
2.5 Importance of evidence-based interventions 33
2.6 The current study 34

Chapter III - Methods: Intervention study 36
3.1 Participants 37
3.1.1 Group allocation 38
3.1.2 Background variables of participants 39

3.2 Testing procedure 40
3.3 Measures 41
3.4 Pilot study 44
3.5 Primary outcome measures 46
3.5.1 Phonological awareness in Luxembourgish 46
3.5.2 Letter-sound knowledge 49
3.6 Secondary outcome measures 50
3.6.1 Phonological awareness in German 50
3.6.2 Literacy measures 51
3.7 Control Measures 54
3.7.1 Vocabulary 54
3.7.2 Non-verbal reasoning 55
3.7.3 Number knowledge 55
3.8 Reliability of the instruments 55
3.9 Intervention programme 57
3.9.1 Intervention structure 58
3.9.2 Supporting phonological awareness 59
3.9.3 Supporting letter-sound knowledge 60
3.9.4 Promoting print and book awareness 62
3.9.5 Encouraging parent involvement 62
3.9.6 Delivery of intervention 63
3.10 Measures on fidelity of implementation of intervention 64

CHAPTER IV - Results and Discussion: Intervention study 65

4.1 Fidelity of intervention implementation 66
4.1.1 Adherence to the intervention 66
4.1.2 Quality of intervention implementation 67
4.1.3 Dosage of the intervention 67
4.1.4 Participant responsiveness to the intervention 68
4.1.5 Programme differentiation 69
4.2 Preliminary data considerations 70
4.3 Data screening 70
4.4 Data analysis procedures 71
4.5 Examination of possible study confounds 73
4.5.1 Group differences on background and control variables 74
4.5.2 Group differences on vocabulary knowledge 75
4.5.3 Group differences on outcome measures at pre-test 76
4.6 Data preparation 77
4.7 Efficacy of the early literacy intervention for the entire sample 79
4.7.1 Effects on primary outcome measures 79
4.7.2 Effects on phonological awareness in German in Grade 1 83
4.7.3 Effects on literacy measures 84
4.7.4 Specificity of intervention effects 88
4.8 Efficacy of the early literacy intervention for children with low oral language proficiency in Luxembourgish 89
4.8.1 Identifying children with low oral language proficiency in Luxembourgish 89
4.9 Examination of possible study confounds 94
4.9.1 Group differences on background and control variables 94

4.9.2 Measures of vocabulary ... 95
4.9.3 Group differences on outcome measures at pre-test ... 96
4.10 Efficacy of the intervention for children with LOL ... 98
4.10.1 Effects on primary outcome measures ... 98
4.10.2 Effects on phonological awareness in German in Grade 1 ... 102
4.10.3 Effects on literacy measures ... 103
4.10.4 Specificity of intervention effects ... 106

4.11 Discussion: Efficacy of the intervention ... 107
4.11.1 Effects on primary outcome measures at post-test for the entire sample ... 108
4.11.2 Effects on primary outcomes at delayed follow-up for the entire sample ... 111
4.11.3 Effects on phonological awareness in German for the entire sample ... 112
4.11.4 Effects on literacy measures at post-test for the entire sample ... 113
4.11.5 Effects on literacy measures at delayed follow-up for the entire sample ... 114
4.12 Efficacy of the intervention for children with low oral language (LOL) skills in the language of instruction ... 118
4.12.1 Effects on primary outcome measures for the LOL group. ... 119
4.12.2 Effects on primary phonological awarness in German for the LOL group ... 120
4.12.3 Effects on literacy measures for the LOL group ... 120
4.13 Strengths, weaknesses and future directions ... 121

Study Two
Early predictors of learning to read and spell in a second language: a longitudinal study with linguistically diverse children

Chapter V - Literature review: Predictor study ... 127
5.1 Phonological awareness ... 127
5.2 Phonological awareness as a predictor of reading and spelling ... 131
5.3 Letter-sound knowledge ... 133
5.4 Letter-sound knowledge as a predictor of reading and spelling ... 134
5.5 Oral language ... 136
5.6 Oral langue as a predictor of reading and spelling ... 138
5.7 Rapid automatized naming ... 141
5.8 Rapid automatized naming as a predictor of reading and spelling ... 142
5.9 Verbal short-term memory ... 144
5.10 Verbal short-term memory as a predictor of reading and spelling ... 146
5.11 Summary on predictors ... 148
5.12 The current study ... 149

Chapter VI - Methods: Predictor study ... 150
6.1 Participants ... 150
6.2 Study design ... 151
6.3 Measures ... 152
6.3.1 Rapid automatized naming ... 152
6.3.2 Verbal short-term memory ... 153
6.4 Testing procedure ... 153

Chapter VII - Results and discussion: Predictor study 154
7.1 Performance on measures in preschool and Grade 1 155
7.2 Concurrent correlations between potential preschool predictors 157
7.3 Concurrent correlations between literacy measures in Grade 1 158
7.4 Longitudinal correlations between potential preschool predictors and literacy outcome measures in Grade 1 158
7.5 Predicting Grade 1 reading and spelling from preschool predictors 160
7.5.1 Information on data preparation 160
7.5.2 Predicting Grade 1 word reading from preschool predictors 163
7.5.3 Predicting Grade 1 reading comprehension from preschool predictors 164
7.5.4 Predicting Grade 1 spelling from preschool predictors 165

7.6 Discussion 166
7.7 Cross-sectional relationships between preschool measures 166
7.7.1 Cross-sectional relationships for phonological awareness at preschool 166
7.7.2 Cross-sectional relationships for letter-sound knowledge at preschool 169
7.7.3 Cross-sectional relationships for vocabulary at preschool 170
7.7.4 Cross-sectional relationships for rapid automatized naming at preschool 171
7.7.5 Cross-sectional relationships for verbal short-term memory at preschool 171
7.8 Cross-sectional relationships between Grade 1 literacy measures 172
7.9 Longitudinal relationships between predictor and literacy variables 173
7.9.1 Preschool phonological awareness as a predictor of Grade 1 literacy 173
7.9.2 Preschool letter-sound knowledge as a predictor of Grade 1 literacy 175
7.9.3 Preschool vocabulary as a predictor of Grade 1 literacy 175
7.9.4 Preschool rapid automatized naming as a predictor of Grade 1 literacy 177
7.9.5 Preschool verbal short-term memory as a predictor of Grade 1 literacy 179
7.10 Summary of findings 181
7.10.1 Strengths, weaknesses and future directions 182

Chapter VIII - General discussion 184
8.1 Supporting early literacy skills in linguistically diverse children 185
8.2 Predictors of reading and spelling in second language learners 188
8.3 Implications 192
8.4 Conclusion 195

Glossary 196

Appendices 199

References .... Error! Bookmark not defined.

List of tables

Table 3.1: Overview of the demographic data of the preschool pairings based on the data from the Ministry of Education (number of children per school, number of L2 learners (%) and SES 38

Table 3.2: Overview of the languages spoken at home by the children in the intervention and control group 40

Table 3.3: Overview of primary, secondary and control measures for each testing point (t1, t2, t3) 43

Table 3.4: Overview of the linguistic units and levels of explicitness tapped by the subtests of the PA assessment battery in Luxembourgish for each testing point 46

Table 3.5: Item examples of the rhyme identification subtest 47

Table 3.6: Item examples of the onset identification subtest 48

Table 3.7: Overview of the linguistic unit and level of explicitness of the PA tests in German in Grade 1 50

Table 3.8: Internal consistency for all measures at t1, t2 and t3 56

Table 3.9: Overview of the 23 sounds with the corresponding letter and the mnemonic word, in the order of introduction 61

Table 4.1: Number (%) of male students and Luxembourgish language learners in the intervention and in the control group at pre-test, with chi-square significance levels and Cramer's V effect sizes for group differences 74

Table 4.2: Mean raw scores (SD) for age (t1, t2, t3), SES (t1) and NVR (t1) for children of the intervention and control group, with t-test significance levels and Cohen's d effect sizes for group differences 74

Table 4.3: Mean raw scores (SD) for vocabulary measures in Luxembourgish (t1, t2, t3) and German (t3) for children of the intervention and control group, with t-test significance levels and Cohen's d effect sizes for group differences 75

Table 4.4: Mean raw scores (SD) for early literacy measures at pre-test for children of the intervention and control group, with significance levels and effect sizes for group differences 76

Table 4.5: Mean raw scores (SD), with ANCOVA significance values and Cohen's d effect sizes for intervention effects at pre-test (t1), at post-test (2) and at delayed follow-up test (t3) for primary outcome measures.............................80

Table 4.6: Mean raw scores (SD), with ANCOVA / ANOVA significance values and Cohen's d effect sizes for intervention effects at delayed follow-up test (t3) for PA in German..83

Table 4.7: Mean raw scores (SD), with ANCOVA / ANOVA significance values and effect sizes for intervention effects at pre-test (t1), at post-test (2) and at delayed follow-up test (t3) for literacy measures 85

Table 4.8: Mean raw scores (SD), ANCOVA significant values and Cohen's d effect size for group differences at pre-test (t1) and at post-test (t2) for number naming 88

Table 4.9: Number (%) of children with LOL per preschool, with chi-square test of independence and Cramer's V effect size...90

Table 4.10: Language backgrounds of children with LOL and their no-LOL peers 91

Table 4.11: Means (SD) for age and SES for the LOL and the no-LOL group, with t-test significance values and Cohen's d effect sizes for group differences ...92

Table 4.12: Gender and number (%) of Luxembourgish language learners of the LOL and the peer group, with chi-square significance levels and Cramer's V effect sizes for group differences 92

Table 4.13: Means (SD) for vocabulary in Luxembourgish and German for children with LOL compared to their no-LOL peers, with t-test significance levels and Cohen's d effect sizes for group differences 93

Table 4.14: Number (%) of male students and Luxembourgish language learners in the intervention and in the control group with LOL, with t-test significance levels and Cohen's d effect sizes for group differences 94

Table 4.15: Mean raw scores (SD) for age, SES and NVR for children of the intervention and control group with LOL at pre-test, with significance levels and effect sizes for group differences.. 95

Table 4.16: Mean raw scores (SD) for intervention and control children with LOL for vocabulary measures at pre-test (t1), immediately postintervention (t2) and delayed follow-up (t3), with test of differences and effect sizes..................... 95

Table 4.17: Mean raw scores (SD) for outcome measures at pre-test for children with LOL of the intervention and control group, with t-test significance levels and Cohen's d effect sizes for group differences... 96

Table 4.18: Mean raw scores (SD), with ANCOVA significance values and Cohen's d effect sizes for intervention effects at pre-test (t1), at post-test (2) and at delayed follow-up test (t3) for primary outcome measures............................ 99

Table 4.19: Mean raw scores (SD), with ANOVA significance values and effect sizes for intervention effects at delayed follow-up test (t3) for PA in German 102

Table 4.20: Mean raw scores (SD), with ANCOVA / ANOVA significance values and effect sizes for intervention effects at pre-test (t1), at post-test (2) and at delayed follow-up test (t3) for literacy measures 103

Table 4.21: Mean raw scores (SD), ANCOVA significant values and effect size for group differences at pre-test (t1) and at post-test (t2) for number naming .. 106

Table 6.1: Overview of the main home languages spoken 151

Table 6.2: Overview of measures with cross-references to the task description 152

Table 7.1: Mean raw scores (SD) and range of potential predictors in preschool and literacy measures in Grade 1.. 155

Table 7.2: Pearson's correlations between potential preschool predictor measures 157

Table 7.3: Pearson's correlations between literacy measures in Grade 1 158

Table 7.4: Pearson's correlations between potential predictors in Luxembourgish and Grade 1 literacy measures in German... 159

List of Figures

Figure 2.1: The virtuous circle between theory and practice 33

Figure 3.1: Timeline of study showing the three testing points, the educational phase of the children, mean ages and SDs (in months) 36

Figure 3.1: Developmental sequence of PA 129

Figure 6.1: Timeline of the study design showing the two test points, grades, mean ages and SDs (in months)..152

Figure 7.1: Path analyses model predicting word reading in Grade 1 from preschool predictor variables 164

Figure 7.2: Path analysis model predicting reading comprehension in Grade 1 from preschool predictor variables 165

Figure 7.3: Path analysis model predicting spelling in Grade 1 from preschool predictor variables..166

List of abbreviations

ELFE	*Ein Leseverständnistest für Erst- bis Sechstklässler*
HSP	*Hamburger Schreib-Probe*
ISCO	*International* Standard Classification of Occupations
ISEI	International Socio-Economic Index of occupational status
L1	First language
L2	Second language
LALA	Name of the hand puppet used in the intervention
LOL	Low oral language in Luxembourgish
LSK	Letter-Sound Knowledge
MCAR	Missing Completely at Random
MENJE	*Ministère de l'Éducation nationale, de l'Enfance et de la Jeunesse*
NRP	*National Reading Panel*
NELP	*National Early Literacy Panel*
NVR	Non-Verbal Reasoning
OL	Oral Language
OECD	*Organisation for Economic Co-operation and Development*
PA	Phonological Awareness
PCA	Principal Component Analysis
PISA	*Programme for International Student Assessment*
PPVT	*Peabody Picture Vocabulary Test*
RAN	Rapid Automatized Naming
RCT	Randomised Controlled Trial
SEM	Structural Equation Modelling
SES	Socio-Economic Status
SVR	Simple View of Reading
SLRT	*Salzburger Lese- und Rechtschreibtest*
TPB	*Test für Phonologische Bewusstheitsfähigkeiten*
VSTM	Verbal Short-Term Memory
WM	Working Memory
WPPSI	*Wechsler Preschool and Primary Scale of Intelligence*

INTRODUCTION

Becoming literate is, arguably, one of the most important milestone of early academic learning for children growing up in our information-driven society. However, while most children acquire oral language naturally through interaction with their environment, there is no biological predisposition *per se* for written language acquisition. Instead, instruction is required to become literate (Bowman & Treiman, 2004). All over the world more and more children are immersed in school settings where instruction takes place in a language that is not their first language. These so called "second language learners" (L2 learners), do not only face challenges to follow general school instructions in a language they have yet to fully acquire, but they are often also disadvantaged in comparison to their monolingual peers with regards to learning how to read (August & Shanahan, 2010; Lervåg & Aukrust, 2010; Sonnleitner et al., 2018). A large corpus of early reading research suggests that preparing children optimally for literacy development before they are formally introduced to reading can have long lasting positive effects (Bowyer-Crane et al., 2008; Fricke, Bowyer-Crane, Haley, Hulme, & Snowling, 2013). Particularly systematic and explicit training of phonological awareness and letters sound knowledge have been linked to improvements in early word-level literacy-acquisition in monolingual and multilingual speakers (Hatcher, Hulme, & Snowling, 2004; Huo & Wang, 2017; Piasta & Wagner, 2010; Snowling & Hulme, 2012; Stuart, 1999; Torgerson, Brooks, & Hall, 2006; Yeung, Siegel, & Chan, 2013).

However, the field of early literacy research has been overly shaped by work in English-speaking countries (Share, 2008). It has become increasingly clear that there are fundamental differences between orthographies and that learning to read in English may be qualitatively different from learning to read in more consistent orthographies (Caravolas, Lervåg, Defior, Seidlová Málková, & Hulme, 2013; Dombey, 2006; Share, 2008). Another general concern has been the scarcity of research on L2 literacy development and the fact that L2 learners are often overrepresented amongst children that don't acquire literacy as expected (Kieffer & Vukovic, 2013; Lesaux, Crosson, Kieffer, & Pierce, 2011). In addition, only very few studies have examined whether early literacy interventions that have shown to be efficient for monolinguals are also effective for L2 learners (for review see Richards-Tutor, Baker, Gersten, Baker, & Smith, 2016). In recent reviews, Murphy and Unthiah, (2015) and Oxley and de Cat (2019) concluded that the lack of intervention studies aimed at supporting L2 learners' literacy is alarming. Due to differences in the linguistic and orthographic characteristics of languages taken together with differences in social and educational settings between countries, it cannot be assumed that approaches that have been shown to be effective in one country would be equally effective in another. Clearly more research on the developmental trajectories and on how to support literacy acquisition of children learning to read in an L2 in different settings is needed. This thesis is an urgently needed step towards improving the early and accurate identification of students at risk for reading difficulties and towards best practice to prevent early literacy difficulties in children who are learning to reading in an L2.

The wealth of scientific knowledge generated from basic research does often not translate to significant public health benefits (Drolet & Lorenzi, 2011). The so called “research-practice” gap is substantial. It has been estimated that it takes an average of 17 years for research evidence to reach practice (Morris, Wooding, & Grant, 2011). Butler (2008), a senior reporter for the journal Nature, describes the gap between the ivory tower and the real world as the valley of death. He argues that one of the main reasons for the “research-practice” gap is that basic scientists have few incentives to move outside their labs and that applied or translational research “is not the sort of research that gets published by the top journals and spurs promotion” (Butler, 2008, p. 842). In the field of language and literacy development the need for more applied research has also been expressed with regards to the accurate identification of young children at risk of language difficulties and the prevention or remediation of literacy difficulties (Beeghly, 2006; Snowling & Hulme, 2011). There is still a dearth of evidence on the effectiveness of many early literacy interventions in classrooms all over the world. This raises concerns about what children are missing out in terms of early support and about many interventions being administered to vulnerable populations despite the absence of evidence of effectiveness (cf. Strong, Torgerson, Torgerson, & Hulme, 2011).

A first aim of this PhD project is to explore the effects of an early literacy intervention in the multilingual educational context of Luxembourg (Study 1). A particular interest will lie upon the efficacy of the intervention for children with low oral language proficiency in the language of instruction, most of whom are L2-learners. A second aim of thesis is to add to the understanding of L2 literacy development by examining the longitudinal cross-linguistic predictors of reading and spelling in a group of linguistically diverse children growing up in Luxembourg, who are all learning to read in an L2 (Study 2). In line with the two major aims, the thesis is split into two parts, each presenting one study. Chapter 1 gives a brief overview of the educational context in Luxembourg and establishes the overall framework of literacy development for the two studies:

Study 1 is an intervention study evaluating the efficacy of an early literacy programme in multilingual Luxembourg (Chapters 2 to 4). Chapter 2 reviews types of preventative and support strategies to decrease the early risk of reading failure. Chapter 3 describes the methodology of the intervention study (i.e. Study 1). Chapter 4 presents and discusses the results of the intervention study, and considers study strengths, limitations, as well as directions for future research.

Study 2 is a longitudinal study on concurrent correlations and predictors of reading and spelling in children learning to read in an L2 (Chapters 5 to 7). Chapter 5 reviews previous work on potential predictor skills in children acquiring literacy in their first and second language. Chapter 6 describes the methodological aspects of Study 2. Chapter 7 presents the results of the longitudinal study and discusses the findings, study strengths, limitations and directions for future research.

Chapter 8 brings together the findings from Study 1 and Study 2 and presents overall results, implications and concluding remarks.

CHAPTER I - Framework of literacy development

The early literacy intervention study (Study 1) and the longitudinal cohort study (Study 2) of the thesis concern the development of early literacy and literacy skills in children in Luxembourg who grow up speaking more than one language. This first chapter begins with a brief introduction into the multilingual educational system in Luxembourg, with a short overview of how children learn to read and write in mainstream schools in Luxembourg. This is followed by a literature review on early literacy development to set the overall framework for the two subsequent studies of the thesis.

1.1 Literacy acquisition in Luxembourg

Luxembourg is a trilingual country in central Europe. The national and official language is Luxembourgish, but German and French act as additional official languages. Although, an official orthographic system exists for Luxembourgish it is not the language of literacy instruction in Luxembourg and is only taught on a rudimentary level in schools (Gilles, 2014). Instead, children are introduced to literacy in Grade 1 in German.

The Luxembourgish orthography consists of the 26 letters of the Latin alphabet plus three letters with diacritics: "à", "é" and "ë". No research has yet looked at the consistency and orthographic depth of Luxembourg's official orthographic system. However, although, Luxemburgish is regarded as a language of its own, in its origin Luxembourgish has to be considered as a Central Franconian dialect that shares many linguistic features with Standard German (Gilles & Trouvain, 2013). Luxembourg is a culturally and linguistically heterogeneous country, where approximately 59 % of the total pupil population comes from a non-Luxembourgish-speaking background (Lenz & Heinz, 2018). National studies have repeatedly revealed that children from non-Luxembourgish speaking backgrounds show a greater risk of encountering difficulties in reading and writing development than children who are Luxembourgish native speakers (Hoffmann et al., 2018; Sonnleitner et al., 2018).

In Luxembourg children start formal literacy instruction in German in Grade 1 of primary school at the age of 6 years. Before these children attend preschool education that consists of one non-compulsory year of *précoce* (age 3) and two compulsory years of *préscolaire* [preschool] (from age 4 upwards). The language of instruction in preschool is Luxembourgish. A national curriculum for preschools exists, which, does not incorporate explicit formal instruction of letter-sound correspondences. It states that preschool teaching is foremost a social experience and that preschool teaching should focus on a global, holistic approach through immersing children in stimulating contexts and not on the explicit teaching of skills (MENFP, 2011). However, the preschool curriculum states that by the end of Year 2 of preschool, children should be able to: identify rhymes and initial sounds and segment words; differentiate different written signs; handle a book, discover the social use of writing, discover their first name among other names; recognize well-known pictograms; follow the course of events in an easy text that is read to them.

In Grade 1 (age 6), a method focusing on code-related skills (Beck & Beck, 2013) is used to formally introduce children to the task of learning to read and write in all the 156 public schools in Luxembourg (MENFP, 2011). There are around 10 private primary schools in Luxembourg, where less is known about the method of literacy instruction, but which are also likely to follow code-emphasis teaching method. Over the first two years of public primary school, children are systematically introduced one by one to single-letters and later also to multi-letter graphemes. Teaching letters is supported by a “Fibel”, an alphabetic book that sequentially presents letters and graphemes of the alphabet with corresponding words and images. Most teachers use the same material to introduce literacy and this was published by the National Ministry of Education, i.e. the *Karibu* fibel (Berg et al., 2009), or the *MILA* fibel (Biltgen et al., 2013). Children are taught to blend and write letters in German, which is the second language for 98% of the school population (MENFP, 2015).

1.2 The framework of reading

Reading in alphabetic languages is often conceptualized as operating either on word-level or text-level. Word-level skills incorporate skills needed for word decoding and word identification and are perceived to develop prior to advanced text-level skills. Text-level skills build upon world-level skills, but revolve around the comprehension of connected text (Lesaux, Geva, Koda, Siegel, & Timothy, 2008). Before considering the development of reading (see section 1.5), different pathways to proficient word-level reading skills are discussed.

Ehri (1991, 2005) distinguishes between four types of word reading: reading by phonological recording, by analogizing, by prediction and sight word reading. The first three reading techniques help with reading unfamiliar words, while the fourth technique describes how we read familiar words. Reading by phonological recoding consists of reading by converting graphemes or spelling patterns into a blend of phonemes and by searching our mental lexicon for the word that matches the blend (Ehri, 2005, 2014). The second way to read words is so-called reading by analogizing, which consists of the usage of familiar words to read new words. For instance, the reader searches his mental lexicon for parallel spelling of a familiar word and adapts the pronunciation to match letters in the unfamiliar word that s/he is trying to read. For example, the reader decodes the word “snail” by analogizing its spelling pattern to a known word like ”tail” (Ehri, 2014; Goswami, 1986). A third way of reading is by prediction, which consists of using initial letter clues combined with context clues to anticipate the word. Once the word is predicted, its pronunciation is matched to the spelling to confirm whether the sounds fit the letters (Tummer & Chapman, 1998). The above-mentioned three reading strategies vary in terms of the extent letters, letter cues or spelling patterns are used to decode words. However, there is a fourth way to read words that is restricted to reading familiar words, i.e. sight reading or reading by memory. This reading strategy entails that the sight of a familiar spelling directly activates the word’s pronunciation and meaning in the mental lexicon. According to Ehri (2005), reading by memory or by sight can be considered a more advanced skill, or at least a skill that develops later as a consequence of repeated exposure to a written word.

While reading by sight or memory is considered an inherent part of more advanced reading, it is reading by phonologically recoding (i.e. converting graphemes or spelling patterns into a blend of phonemes) that is thought to be the most accurate decoding strategy in early readers as it is the least ambiguous and the most direct manner to read new words (Ehri & Wilce, 1985; Ehri, 2014; Tracey & Morrow, 2017). As will be discussed later (see section 1.6.2**),** the strategy of reading by phonological recording seems to be a particularly efficient route of entry to word reading in consistent alphabetic orthographies (Caravolas et al., 2012; Seymour, Aro, & Erskine, 2003).

1.3 The Simple View of Reading

One of the most parsimonious and wide-spread frameworks of the complex cognitive task of reading is the Simple View of Reading (SVR) (Gough & Tunmer, 1986; Hoover & Gough, 1990). According to the SVR, skilled reading comprehension (R) is the product of two independent components, word decoding (D) and listening comprehension skills (C). A simple equation fully synthesises the framework of the SVR, R = D x C. Decoding (D) refers to the ability to apply the grapheme to phoneme correspondence rules to read printed words and non-words, whereas linguistic comprehension (C) refers to the ability to interpret words, sentences and discourses in the oral domain. As specified by the equation of the SVR, neither of the two components by itself is sufficient for skilled reading.

Since 1986, myriad studies have provided support for the two component framework of the SVR and findings suggest that a large proportion of variance in reading comprehension is indeed accounted for by decoding and listening comprehension skills (Joshi, Tao, Aaron, & Quiroz, 2012; Kendeou, Savage, & Broek, 2009; Kendeou, van den Broek, White, & Lynch, 2009; Nation & Snowling, 1997; Verhoeven & van Leeuwe, 2012). However, the SVR also faced criticism for overly simplifying the complex task of reading (Joshi & Aaron, 2000; Kirby & Savage, 2008). The critics argue that the SVR cannot be considered a comprehensive framework incorporating all the cognitive processes involved in the task of reading, and that for example, phonological memory, rapid automatized naming and non-verbal reasoning play an important role in proficient reading as well (Kershaw & Schatschneider, 2012; NELP, 2008). In addition, Florit and Cain (2011) questioned the generalizability of the oversimplified model across different types of orthographic systems. The authors highlighted that different patterns of relations and weightings between reading comprehension and decoding exist in writing systems with varying degrees of orthographic consistency. For instance, as will be discussed more in-depth later, it has been questioned whether phonological awareness is an equally important contributor to early reading in consistent and inconsistent orthographies (Aarnoutse & van Leeuwe, 2000; Georgiou, Torppa, Manolitsis, Lyytinen, & Parrila, 2012). According to the SVR, the main sub-skills needed for skilled decoding in alphabetic languages are phonological awareness (PA) and letter-sound knowledge (LSK). It follows that strong foundations in those two sub-skills are important prerequisites for proficient decoding. Indeed, PA and LSK have been repeatedly related to individual differences in reading performances (Carroll & Snowling, 2004; Fricke et al., 2015; Muter, Hulme, Snowling, & Stevenson, 2004) and experimental research has suggested that this relationship implies causality (Bowyer-Crane et al., 2008; Hatcher, Hulme, & Ellis, 1994; Hulme, Bowyer-Crane, Carroll, Duff, & Snowling, 2012; Lundberg, Frost, & Petersen, 1988).

1.4 Framework of spelling

Spelling is a quintessential part of literacy and consists of the encoding of the spoken language into written forms (Perfetti, 1997). Reading and spelling can be considered closely related coding processes, though in the opposite direction, that rely on the same underlying fundamental knowledge of the workings of the alphabetic principles (Abbott, Berninger, & Fayol, 2010; Ehri, 1997). Phonological processing skills enable spellers to segment the sound structure of spoken words into individual phonemes. Subsequently, the application of phoneme to grapheme rules allows them to write down the letters to the corresponding sounds. Yet, in the same way that mere decoding skills are not sufficient for proficient reading, pure encoding skills are not sufficient for proficient spelling. Whereas, early spelling is indeed essentially characterized by reliance on sound based information, skilled spelling requires both detailed orthographic and morphological knowledge to spell more efficiently and to spell irregular words accurately (Juel, 1983; Treiman &

Bourassa, 2000). Because spelling and reading are predicated on similar underlying components required for proficient reading and proficient spelling, it is unsurprising that reading and spelling are intimately associated (Georgiou et al., 2019) and that spelling and reading tend to be predicted by the same early literacy skills (Aarnoutse, van Leeuwe, & Verhoeven, 2005; Fricke, Szczerbinski, Fox-Boyer, & Stackhouse, 2015; Harrison et al., 2016; Hulme, Nash, Gooch, Lervåg, & Snowling, 2015; Jongejan, Verhoeven, & Siegel, 2007; Nation & Hulme, 1997; National Reading Panel, 2000; Verhoeven, 2000). In addition, reading has shown to improve spelling and spelling has also shown to improve reading (Ehri & Wilce, 1987; Graham & Hebert, 2011), and while not being a perfect relationship, poor readers tend to show poor spelling skills (Caravolas, Hulme, & Snowling, 2001)

It has been debated whether it is necessary to systematically and explicitly teach spelling, or whether spelling can be "caught" through immersion into a literacy rich environment (Graham, 2000). According to the latter view, spelling is the indirect results of the act of reading (Graham, 2000; Graham & Santangelo, 2014; Treiman, 2018). However, research has shown that reading experience in itself is generally not sufficient to acquire proficient spelling. Phonics[1] instruction combined with systematic teaching of the spellings of specific inconsistent words and combined with strategies to spell unknown words should be a major part of spelling instruction in alphabetic languages (Graham & Santangelo, 2014; Rebecca Treiman, 2018).

Learning to spell is often considered to be more difficult than learning to read because the grapheme to phoneme relations in many alphabetic languages (e.g. English, Dutch, French, German, and Spanish) are more consistent than the phoneme to grapheme associations (Bosman & van Orden, 1997). As a consequence, children face more code-related ambiguity when trying to spell than when trying to read (Caravolas, 2004; Caravolas et al., 2012; Cossu, Gugliotta, &

[1] The teaching of the systematic relationship between letters and their corresponding sounds, and how to decode words by blending the individual sounds together; or how to spell by segmenting words into individual sounds, is traditionally referred to as Phonics instruction (Carnine et al., 2004).

Marshall, 1995; Landerl & Thaler, 2006). In German, the encoding from phoneme to grapheme involved in spelling is further towards the inconsistent end of the continuum of orthographic consistency (see section 1.6.2 for more information regarding the consistency of the German orthography), whereas the grapheme to phoneme associations involved in reading are more towards the consistent end (Landerl & Thaler, 2006). In addition, contextual or semantic cues that may support reading words in a sentence are not really helpful to the task of spelling words (Ehri, 1997). These asymmetries between learning to read and learning to spell may further be amplified by the fact that children tend to engage less frequently in spelling activities than in reading (Bosman & van Orden, 1997).

1.5 Development of reading and spelling

Research over the last decades has resulted in various type of models of literacy development, e.g. specific stage models (cf. Ehri, 2005; Frith, 1985), dual-route models (cf. Coltheart, 2001, 2005) and connectionist developmental models (cf. Harm & Seidenberg, 1999; Plaut, 2005; Seidenberg, 2005). The primary objective of the current project is not to inform and update theoretical developmental models of literacy and discussing all models in greater detail is beyond the scope and purpose of this work. The current thesis focuses on early literacy development and word-level reading skills. Ehri's (2005) developmental phase model will be presented due to its wide-ranging acceptance and its emphasis on the development of initial word-level skills and less on higher text-level reading comprehension.

Ehri's phases of developing readers are characterized by a particular way children retrieve the phonological and sematic information of a word based on its orthographic form. Ehri's model discriminates between four main phases of word reading development. The first *pre-alphabetic* phase is characterized by little knowledge of the alphabet and the lack of grapheme to phoneme connections as a strategy to read words. Children in this stage do not rely on the alphabetic principle[2] to read words, instead they use visual cues to retrieve the phonological and sematic information of a word (Ehri & Wilce, 1985). For example, children in this phase recognise words by its salient features (e.g. two round eyes in ‘look’) or based on pictorial features that typically co-occur with the words (e.g. colour and shape of environment print). Memorizing numerous words in this first phase is difficult and inefficient as no links between letters and sounds are formed and responses rely on visual clues.

Children transition to the *partial alphabetic* phase with a growing awareness of grapheme to phoneme mappings. They begin to make use of letter names or sounds to recognise words. However, children have not yet learned to form connections between all of the sounds and letters of a word. Rather children rely on certain letters, mostly the first and the final letter of a word as clues to recognize words. This strategy may lead to ambiguities between similarly spelled words with the same boundary letters such as spoon and skin. According to Ehri (2005), children only enter the *alphabetic* phase when they have understood the alphabetic principle and acquired proficient phoneme bending and segmenting skills. This means children are able to consistently recode all the constituent graphemes in a word into their corresponding phonemes and blend them. At the same time, the spelling of a word also becomes connected to its pronunciation in children’s memory, setting the groundwork for more effective and accurate sight word reading. The last stage of Ehri’s four phase model goes beyond the ability to individually map sounds to graphemes. Children transition to the *consolidated alphabetic* phase when they become more familiar with recurring letter patterns and start to memorize larger chunks of spellings (e.g. spelling patterns of rimes, syllables or multi-grapheme morphemes). Initial support for the larger chunk theory came from Juel (1983), who showed that 5th graders in the US read words containing recurring letters patterns more efficiently than words with less common letter patterns. The larger chunks of orthographic spelling become consolidated phonological units themselves which in turn facilitate reading of words with common letter patterns as fewer connections are needed to secure new words.

---

[2] The understanding that sounds in spoken words can be represented consistently by specific written letters or symbols (Beck & Beck, 2013)

For example, reading the word interesting can be reduced to four syllabic chunks (in-te-res-ting) instead of decoding 10 single grapheme to phoneme correspondences.

Frith (1985) posits that reading and writing develop interactively and influence each other in distinct phases of literacy development. Children first acquire insight into the alphabetic code through practice at spelling, which in turn leads to improved reading of the words that follow simple grapheme to phoneme mappings during early stages of literacy development. Subsequently, improved reading of words with more complex orthographic patterns leads to improved spelling of these patterns. Georgiou et al. (2019) found support of this cross-lagged relation between reading and spelling in children attending Grade 1 to Grade 2 by showing that once children master a decoding strategy, the ability to read words accurately facilitated children's ability to spell words.

It needs to be noted that the models discussed should be conceptualised as a theoretical framework for learning to read and spell. Literacy development trajectories across children are greatly idiosyncratic and vary substantially according to orthographic systems, educational and cultural environments (Beech, 2005).

1.6 External factors influencing literacy acquisition

It is now well established that external factors, such as the social and cultural context in which children grow up, or the orthographic consistency of the language used to introduce literacy have substantial influences on literacy development (Hoff, 2006; Hoff & Tian, 2005; Ziegler et al., 2010). The following section will present a short synthesis on the main effects of social economic status (SES) and orthographic consistency of the language of literacy instruction on literacy development. Each factor will be briefly discussed with reference to the context of Luxembourg.

1.6.1 Socio-economic factors influencing literacy acquisition

Research suggests that children from low SES backgrounds tend to be exposed to a different oral language (OL) input in terms of quantity and quality (Hoff, 2006, 2013). It has been suggested that lower OL skills may function as a significant mediator of SES related differences in learning to read (Durham, Farkas, Hammer, Bruce Tomblin, & Catts, 2007; Forget-Dubois & Dionne, 2009; Hoff, 2013). Research has documented that children from a disadvantaged social and economic background show lower performances on reading tests and demonstrate lower levels of interest in literacy in general (Aikens & Barbarin, 2008; Bowey, 1995; Duncan & Seymour, 2000). Schiff and Lotem (2011), demonstrated that children from a low SES background demonstrate lower PA skills compared to children from more affluent backgrounds and that those initial discrepancies became more pronounced over time. In some cases children from low SES backgrounds hardly demonstrate any progress at all from second to sixth grade (Schiff & Lotem, 2011). This widening of the achievement gap between children from a high and low SES background over school years has been referred to as the Matthew effect in Education (Journals et al., 2011; Stanovich, 1986; Walberg & Tsai, 1983).

In addition to lower OL skills, variations in early literacy skills between children from a lower and higher SES background are often associated to differences in the home literacy environment. For instance, children growing up in a disadvantaged home literacy environment are less exposed to developmentally stimulating materials (e.g. books) and are also less likely to have family members read to them (Bradley, Corwyn, Burchinal, McAdoo, & Garcia Coll, 2001; Burgess, Hecht, & Lonigan, 2002; Hofslundsengen, Gustafsson, & Hagtvet, 2018; Niklas & Schneider, 2013; Weigel, Martin, & Bennett, 2006). Studies have suggested that early literacy skills develop prior to formal instruction and that it was actively getting involved in home literacy activities with a focus on print (e.g. using letters, using books and picture dictionaries, learning letters sounds, etc.) that were best associated with the development of early literacy skills (Levy, Gong, Hessels, Evans, & Jared, 2006; Lukie, Skwarchuk, LeFevre, & Sowinski, 2014; Scarborough & Parker, 2003). In comparison to caregivers from higher SES backgrounds, caregivers from lower SES backgrounds are less likely to model literacy activities at home that expose children to literacy usage (Baroody & Diamond, 2012; Burgess et al., 2002; Levy et al., 2006; Lukie et al., 2014; Scarborough & Parker, 2003).

Findings from the National Education Report 2018 in Luxembourg show that children from disadvantaged backgrounds perform worse on oral language and reading measures in Grade 3 (age 9-10). In total, 54% of the children from socially disadvantaged backgrounds do not meet the required reading standard at Grade 3 (age 9). In contrast, only 18% of their non-disadvantaged peers do not meet the required reading standards (Sonnleitner et al., 2018).

1.6.2 Orthographic differences influencing literacy acquisition

In an alphabetic orthography, an optimal coding script would consist of a perfect one-to-one relationship between phonemes and graphemes, as well as between graphemes and phonemes. This would eliminate ambiguities in the decoding and encoding of phonemes and graphemes. Yet, such an ideal one-to-one mapping has not evolved in any alphabetic language and inconsistencies and irregularities are inherent to any alphabetic script (Treiman & Kessler, 2005).

A wealth of research suggests that, consistency in the relationship between letters and sounds positively correlates with the difficulty that children face to learn to read and write (Caravolas et al., 2013; Duncan, Colé, Seymour, & Magnan, 2006; Frost, 2005; Lallier, Valdois, Lassus-Sangosse, Prado, & Kandel, 2014; Seymour, Aro, & Erskine, 2003; Treiman & Kessler, 2013; Treiman, 1993; Wimmer & Landerl, 1997; Wimmer & Goswami, 1994). Seymour et al. (2003) examined possible effects of orthographic complexity on learning to read across 14 different alphabetic orthographies and concluded that grapheme to phoneme recoding skills are learned more rapidly by children acquiring literacy in consistent- than in inconsistent orthographies. Recent work has further provided support for different development patterns of reading elicited by varying orthographic complexity. Caravolas and colleagues (2013) compared specific growth trajectories of reading development in inconsistent (English) and consistent orthographies (Czech and Spanish). Their analyses of developmental patterns suggest a slow, but steady growth trajectory in reading skills in the inconsistent language English. In contrast, beginning readers in the more consistent languages Czech and Spanish show a rapid initial increase in early reading skills, which is followed by a noticeable deceleration once children have fully grasped the alphabetic principle.

Research on early literacy predictors (i.e. PA, LSK, rapid automatized naming) examined whether orthographic complexity also influences early predictors of reading. The classic predictors of variations in reading development seem to be highly stable predictors of reading and spelling across languages and rather uninfluenced by orthographic differences (Caravolas et al., 2012; Caravolas, Volín, & Hulme, 2005). However, there has been some discussion regarding the predictive importance of PA as studies in more consistent orthographies either assigned a limited role to PA (Landerl et al., 2019; Papadopoulos & Georgiou, 2009) or have failed to identify PA as predictor of reading (Aarnoutse et al., 2005; Fricke et al., 2015; Georgiou et al., 2012; Silvén, Poskiparta, Niemi, & Voeten, 2007). This pattern of prediction has been explained by the fact that, in consistent orthographies, the highly regular correspondence grapheme to phoneme can be grasped with just a few months of formal reading instruction and will put less demands on children PA skills then learning to read in a more inconsistent orthography (Goswami, 2002; Landerl et al., 2013; Ziegler et al., 2010)***.*** Given the influence of orthographic complexity on learning to read, it seems quintessential to take orthographic complexity into account in any research on reading and spelling skills.

Although, an official orthographic system exists for Luxembourgish it is not the language of literacy instruction in Luxembourg. Children are alphabetized in German in Grade 1. Although, Luxemburgish is regarded as a language of its own, in its origin Luxembourgish has to be considered as a Central Franconian dialect that shares many linguistic features with Standard German (Gilles & Trouvain, 2013). The orthographic consistency of Standard German is contingent on the direction of the sound-letter transformation. In the direction of reading, German demonstrates high grapheme to phoneme correspondences and is located further towards the consistent end of the spectrum of orthographic continuum. Considering the phoneme to grapheme correspondences involved in spelling, the German orthography is considered less consistent (Bergmann & Wimmer, 2008).

STUDY ONE - FOSTERING LITERACY DEVELOPMENT IN LINGUISTICALLY DIVERSE LEARNERS: A QUASI-EXPERIMENTAL INTERVENTION STUDY

CHAPTER II - Literature review: Intervention study

Learning to read and spell can be a challenging process and children who experience reading difficulties are not uncommon (Peterson & Pennington, 2012). Children who exhibit risk factors that increase the probability of experiencing reading difficulties are generally referred to as children at-risk of reading difficulties (Carroll & Snowling, 2004). Risk factors can be “child internal” or “child external”. For example, specific language impairments (Bishop & Snowling, 2004) or deficits in PA (Ramus, 2001; Saksida et al., 2016) are considered child internal risk factors of reading difficulties, whereas growing up in a low SES background or speaking a minority-language are considered child external risk factors of reading difficulties (Aikens & Barbarin, 2008; Duncan & Seymour, 2003; Hoff, 2013; Hofslundsengen et al., 2018; Leseman & Jong, 1998).

Encouraging work comes from early intervention studies showing that children at-risk of experiencing reading difficulties can be successfully supported in their literacy development (Bowyer-Crane, Fricke, Schaefer, Lervåg, & Hulme, 2017; Fricke, Bowyer-Crane, Haley, Hulme, & Snowling, 2013; Richards-Tutor, Baker, Gersten, Baker, & Smith, 2016; Van Tuijl, Leseman, & Rispens, 2001). It is also widely accepted that early provisions of growth-promoting experiences early in life are much more effective than remediation of later difficulties (cf. Knudsen, Heckman, Cameron, & Shonkoff, 2006). For example, children at-risk of educational difficulties that have been provided early support are less likely to repeat school grades, or are less likely to require additional support in future years (Knudsen et al., 2006).

The language of instruction in preschool in Luxembourg is Luxembourgish, but children are introduced to literacy in German in Grade 1 (MENFP, 2011). In Luxembourg only half of the population has the Luxembourgish citizenship and Luxembourg's school population is one of the most culturally and linguistically heterogeneous in Europe (OECD, 2010). In the academic year 2015-16 of the current study, 54% of the primary school population had the Luxembourgish nationality, 23% were Portuguese, 5 % French, 4% came from former Yugoslavia, 2% were of Belgian and 2% were German (MENJE, 2017). This heterogeneous school population imposes distinct challenges on teaching beginning reading in Luxembourg, and indeed literacy-achievement has been a long-standing concern in Luxembourg.

The PISA 2012 study highlighted how far Luxembourg's children are trailing in reading scores in comparison to their OECD peers, ranking only 30th place in PISA's international tables (OECD, 2012). National standardized student assessments in Luxembourg also reveal that over 40% of Luxembourg's nine-year-olds do not meet the national reading standards (Hoffmann et al., 2018; Martin & Brunner, 2012). The situation is particularly alarming for the Portuguese-speaking community. Seventy-one per cent (71%) of the Portuguese-speaking children struggle with reading comprehension in Grade 3 (compared to 20% of the L1 Luxembourgish- and L1 German-speaking children).

For 98% of the school population in Luxembourg, the language of literacy instruction German is an L2 (MENJE, 2017). As a result, the large majority of the children face the double challenge of discovering the alphabetic principles while at the same time acquiring a new language. Engel de Abreu, Hornung and Martin (2015) posited that this double challenge may lead to “cognitive overload” in Grade 1. Many L2 learners need to spend a lot of cognitive resources on learning German, which might restrict cognitive space that could otherwise have been used to acquire reading and writing skills. Give this possible cognitive overload, it seems crucial that children in Luxembourg are optimally prepared in preschool for the task of reading in Grade 1.

It has been abundantly shown that well developed PA and LSK are essential components of literacy acquisition across alphabetic languages and that a strong foundational in those early literacy skills facilitates literacy development (Caravolas et al., 2012; Landerl et al., 2013; Ziegler et al., 2010). Rigorously conducted scientific experimental studies have confirmed that an early training of PA and LSK is an effective method to support children's early word reading skills in L1 learners (Bowyer-Crane et al., 2008; Bus & van IJzendoorn, 1999; NRP, 2000) and in German L1 learners (Fischer & Pfost, 2015; Schneider, Roth, & Ennemoser, 2000).

However, questions remain about the effectiveness of early literacy support for L2 learners. Murphy and Unthiah (2015) reviewed intervention studies from English-speaking countries that aimed at improving L2 English language and /or literacy skills (between 2000 and 2014). The authors concluded that the scarcity on the effectiveness of early literacy support outside the US is alarming. The current study aims to address this gap by adding to the knowledge base of L2 literacy development. The current intervention study explores the efficacy of a newly developed early literacy programme for improving early literacy skills in linguistically diverse learners from Luxembourg who are learning to read in German.

2.1 Foundational skills of reading and spelling

Based on the framework of the SVR, efficient reading comprehension builds upon the interaction of listening comprehension and proficient decoding skills (Gough & Tunmer, 1986; Hoover & Gough, 1990). To become a proficient decoder, children need to learn to identify words accurately and reliably (Trieman, 2000). In early stages of reading development, the most efficient way to do this in alphabetic languages is by learning how the letters in printed words relate to sounds in spoken words. The understanding of the so called "alphabetic principle" is conceived as the most efficient way to decipher written words on their first encounter (Ehri, 2014). Once children manage to apply the alphabetic principle to read words, the frequent encounter of recurrent spelling patterns or entire words allows them to advance to more efficient orthographic reading or reading by sight (Ehri, 2014) (see section 1.5 on the development of reading and spelling). The two foundational skills on which learning to decode rests are PA and LSK (Trieman, 2000). PA refers to the understanding that spoken language is constituted of smaller components which can be manipulated independently from meaning (Chard & Dickson, 1999; Stahl & Murray, 1994). LSK is children's knowledge about the letters and the corresponding sounds in a specific language (Piasta & Wagner, 2010). The second study in the thesis (Study 2) explores the role of those two foundational literacy skills in learning to read and spell and provides an extensive literature review on their development and their role for L1 and L2 literacy development (see sections 5.1 & 5.2 for PA, and sections 5.3 & 5.4 for LSK).

Based on the SVR, decoding is necessary but not sufficient for proficient reading. The second important foundation of reading comprehension is listening comprehension (Gough & Tunmer, 1986). This means, in order to create meaning, oral language skills are necessary (Kendeou, van den Broek, White, & Lynch, 2009). OL skills such as grammar and syntactic knowledge, narrative skills and particularly vocabulary size have all been connected to literacy skills (Duff, Reen, Plunkett, & Nation, 2015; Lervåg, Hulme, & Melby-Lervåg, 2017; Muter et al., 2004; Scarborough, 2001; Van Viersen et al., 2018). A more in-depth discussion of the importance of OL in learning of early reading and spelling development in L1 and L2 learners is provided again in Study 2 (sections 5.5 & 5.6 for more information on the relationship between OL and reading and spelling).

With regards to Study 1, it is important to note that it is well established that LSK, PA and OL lay the foundation for literacy skills and that a lack in those foundation skills can be a major cause of reading difficulties (Brown, 2014; Lervåg, Bråten, & Hulme, 2009; Moll et al., 2016; Trieman, 2000). Deficits in LSK have been related to later literacy difficulties (Hammill, 2004; Storch & Whitehurst, 2002) and an overwhelming body of research has shown that phonological deficits are a main cause of word decoding difficulties (Boets et al., 2010; Kudo, Lussier, & Swanson, 2015; Law, Vandermosten, Ghesquière, & Wouters, 2017; Melby-Lervåg, Lyster, & Hulme, 2012; Muter et al., 2004; Saksida et al., 2016; Vellutino, Fletcher, Snowling, & Scanlon, 2004). Deficits in phonological skills seem to present before literacy instruction begins and may limit the ability to establish and sound out the mappings between letter strings and sounds (Kim, Petscher, Foorman, & Zhou, 2010; Lerner & Lonigan, 2016; Ramus, 2001; Snowling, Gallagher, & Frith, 2003). In addition, low OL skills represent an important foundation skill that has been shown to influence emerging reading skills, and particularly early reading comprehension development (Hulme et al., 2015; Lervåg et al., 2017; NICHD Early Child Care Research Network, 2005).

2.2 Foundational skills of reading and spelling in L2 learners

As mentioned above, L2 learners often present lower oral language skills in the instructional language than L1 learners (August & Shanahan, 2010; Kieffer & Vukovic, 2013; Lervåg & Aukrust, 2010; Mayo & Leseman, 2008; Scheele et al., 2010), which have been linked to lower L2 reading and L2 spelling performances (Hoffmann et al., 2018; Jean & Geva, 2009; Lervåg et al., 2017; Melby-Lervåg & Lervåg, 2014; Raudszus, Segers, & Verhoeven, 2018) (see section 5.6). In addition, research has shown that L2 learners tend to struggle more with the acquisition of proficient PA skills and LSK than L1 learners (Hoff, 2013; Páez, Tabors, & López, 2007). For example, Muter and Diethelm (2001) compared performances on PA and LSK tasks between L1 learners and L2 learners in Switzerland. The authors tested the children in preschool (age 5) and one year later in Grade 1. Results showed that in preschool, L1 learners outperformed the L2 learners on measures of receptive vocabulary, LSK and rhyme production. In Grade 1, differences in vocabulary and LSK were maintained, but the two groups showed comparable performance on PA measures. A difference of moderate effect size was also observed in reading performances between the L1 and the L2 learners. Weber, Marx and Schneider (2007) compared performances on early literacy skills between L1 German speakers to L2 speakers in Germany in preschool (age 5-6). The authors found that the L1 speakers consistently outperformed the L2 speakers on various measures of PA. No differences in LSK were observed, however, LSK was low for the two groups. Verhoeven (2000) compared early literacy skills between Dutch L1 speaker and Dutch L2 speakers in the Netherlands. In Grade 1 (age 6-7), the L2 learners did not struggle with Dutch word blending, but they presented deficits in L2 phoneme segmentation and L2 LSK. Concerning L2 learners in Luxembourg, the national educational report (2018) showed that 78% of the L1 Luxembourgish and L1 German speakers have acquired the highest performance level of early literacy skills in Grade 1 (age 6-7), whereas only 54% of the Portuguese-speaking children reach the highest performance level (Hoffmann et al., 2018). These studies conducted in various contexts draw a common conclusion: L2 learners lack behind their L1 peers in foundational literacy skills.

There are a few studies showing a more positive picture by reporting comparable performances on early literacy measures between L1 and L2 learners. Lipka and Siegel (2007) found that L2 learners in Canada showed equal levels of LSK as their English-speaking L1 peers in Kindergarten (age 5-6). Kieffer and Vukovic (2013) explored differences in LSK and PA skill between L2 learners from ethnically diverse backgrounds and L1-English speaking children in the US, and found comparable performances of the two groups on the early literacy tasks in Grade 1. Other studies also reported equal performances LSK and PA between L1 and L2 children (Goodrich & Lonigan, 2017; Lesaux, Rupp, & Siegel, 2007). The authors give possible explanations for the comparable performances of L1 and L2 learners, e.g. higher SES of the L2 learners (in Canada the bilingual schools are seen as prestigious schools) or a different educational culture with a greater emphasis on emergent literacy in some school contexts (Lipka & Siegel, 2007).

However, other factors such as attendance rates of preschool education, cultural differences in home literacy environment and the amount exposure to the school language prior to school entry also influences early literacy skills in L2 learners (c.f. Hammer, Jia, & Uchikoshi, 2011).

Overall, a bulk of research has suggested that lower vocabulary puts L2 learners at a greater risk of experiencing reading comprehension difficulties in contrast to their L1 peers. Despite counterevidence from a limited amount of studies, the majority of studies have demonstrated that L2 learners in mainstream schools lag behind their L1 peers in the development of L2 PA and L2 LSK. Given the importance of these early foundational literacy skills and the long-term negative consequences that deficits in those skills can have (Hoffmann et al., 2018), it is considered a priority to find ways to provide optimal support to L2 learners as early as possible (Murphy & Unthiah, 2015; Tunmer & Hoover, 2019).

2.3 Supporting early literacy skills

Word-level reading difficulties in early stages of reading development are often explained by a phonological deficit (Melby-Lervåg et al., 2012), that also impedes their development of the understanding of the letter-sound mappings (Foulin, 2005). By implication, supporting children in developing proficient PA skills and in understanding the links between sounds and letters should be particularly effective to prepare kids for literacy acquisition. Indeed, research has shown that training in those two foundational skills is an effective way of supporting children in the task of learning to read and write (Bowyer-Crane et al., 2008; Bus & van Ijzendoorn, 1999; Fischer & Pfost, 2015; Hatcher et al., 1994; Schneider et al., 2000; Suggate, 2016). The current sections reviews intervention studies that aimed to support children's early literacy/decoding skills (i.e. PA and LSK) in L1 and L2 learners. The focus of the review is on studies that had a preventative approach to reading difficulties (Nation, 2019), which means the reviewed studies were conducted before children were formally introduced to literacy, or in the first year of formal literacy instruction.

2.3.1 Supporting phonological awareness in L1 learners

This section reviews studies that delivered interventions targeting only PA skills (without training in LSK). In a landmark study over three decades ago, Bradley & Bryant (1983) demonstrated that a specific training in categorizing sounds (40 sessions over two years) gave pupils in the UK (age 4-5) a head start of three to four months in reading and spelling skills. Their results showed that PA could be successfully trained and that fostering PA generalises to measures of reading and spelling. A second seminal study in the field of early literacy interventions was conducted by Lundberg et al. (1988), who explored the effects of daily meta-linguistic phonological activities in format of games delivered in group sessions of 15-20 children (age 6) in Denmark. The first two months of the intervention focused on the rhyme and the syllable level, and phonemic level was introduced in the third month out of the eight-month intervention period. Their results showed that the PA training programme administered by the teachers to preschool children led to statistically significant improvements in reading and spelling in Grade 1.

Children in Luxembourg are introduced to literacy in German and the effectiveness of PA training in children acquiring literacy in German has also been examined. Schneider, Küspert, Roth, Visé, and Marx (1997) adapted the Danish PA programme from the Lundberg et al. (1988) to German and evaluated its effectiveness in German preschools (age 5-6). The intervention was delivered by instructed classroom teachers over six months. Schneider and colleagues (1997) confirmed that the training successfully trained PA in comparison to a control group who followed the regular preschool programme in Germany. In line with the intervention effects from Lundberg at al. (1988), Schneider et al. (1997) also found that intervention effects on PA in preschool generalised to measures of reading and spelling in Grade 1 and even in Grade 2. Yet two other noteworthy findings emerged from their study. Firstly, only nine out of the 21 intervention teachers did administered the intervention consistently to the end. The authors conjectured that teachers fell behind because of their many other concurrent obligations. However, this meant that many teachers did not deliver the most difficult and most important segments of the training who should be covered towards the end of the intervention, i.e. phoneme identification and segmentation. Notably, the children who were delivered the training consistently and in its entirety showed larger intervention effects, particularly larger longer term effects, than children who were administered the training consistently. A second interesting find was that the training was more effective for measures of spelling than of reading. According to Schneider at al. (1997), this was unsurprising as it is consistent with the position that the success of the PA intervention program may vary as a function of the consistency of the orthography involved in spelling and in reading (see section 1.6.2).

Similar positive intervention effects emerged in another intervention study evaluating the effectiveness of PA training in German preschools (age 5-6) with at-risk children (Schneider et al., 2000). Schneider et al. (2000) screened over 700 children on PA skills, and the lowest performing 208 children were categorized as at-risk of experiencing reading difficulties. Children were delivered a shorted version of the Lundberg et al. (2000) intervention (i.e. daily sessions over five months instead over eight months) in small groups (5-8 pupils) in session of 15-20 minutes. The intervention effects of the PA training were compared to two other at-risk groups, who were administered either a LSK alone-, or a combined PA and LSK training. The study design also included a passive control group of non at-risk children. Results revealed that immediately post-intervention, the group that trained PA alone, showed the highest performances on PA measures. However, by the time of the follow-up in Grade 1, the other three groups (i.e. LSK alone, PA combined with LSK, and the non at-risk control group) had caught up and showed comparable performances on measures of PA. With regards to measures of reading and spelling in Grade 1, the PA alone group and the LSK alone group scored significantly lower than the group trained on PA combined with LSK and the non at-risk control group. The stronger maintained effects and the larger effects on reading and spelling measures of the combined training of PA and LSK clearly suggests that a combined training of the two skills is more effective than a training focusing on either of those two skills.

To the best of my knowledge, to date only one study has examined the effectiveness of PA training in the context of Luxembourg, i.e. a field experiment by Bodé and Content (2011). The authors adapted the intervention programme from Schneider et al. (1997) to Luxembourgish (Bodé, 2004). The intervention was delivered by mainstream classroom teachers to all children in 20 preschool classes (age 5-6), with minimal external supervision by the research team. Intervention effects were compared to a passive control group of 21 preschool classrooms. The training consisted of daily PA session of 10 minutes over 20 weeks. The control group followed the regular preschool curriculum. Immediately post-intervention, statistically significant intervention effects were observed on multiple PA tasks in Luxembourgish (with Cohen *d*'s ranging from .43 to .67). However, gains did neither generalize to LSK immediately post-intervention in preschool nor to spelling in German in Grade 1. Notably, similarly to Schneider et al. (2000), Bodé and Content (2011) also observed that only half of the teachers fully applied the intervention programme until the last stages.

A few meta-report on the effectiveness of early literacy intervention haven been produced over the last decades. A first landmark meta-review was conducted by Bus and van Ijzendoorn in 1999. The metric used to compare effect sizes across studies were Cohen's *d* effect sizes. Their results indicated an overall very large training effect of PA training on PA measures in preschool (d = 1.26) and an overall moderate effect of PA training on measures of reading (d = 0.32). However, most of the reviewed studies were conducted in the US. Especially relevant for the current study is a meta-review by Fischer and Pfost (2015), who only included intervention studies (N = 19) conducted with German-speaking children up to and including Grade 2 in their meta-review. The authors found a positive meta-intervention effect of PA interventions on measures of PA (d = 0.36), LSK (d = 0.26), decoding (d = 0.18), reading comprehension (d = 0.26) and spelling (d = 0.26). For studies including a delayed follow-up over a year after the intervention, small but significant intervention effects of PA interventions were still observable on measures of PA (d = 0.27), LSK (d = 0.20) and spelling (d = 0.19). No significant long-term effects could be observed for measures of decoding (d = 0.03) and reading comprehension (d = 0.16). The intervention effect sizes on the efficacy of PA training in German writing system seem substantially smaller than effects reported from early studies in English-speaking countries. However, as discussed in section 1.6.2, it has been shown that the higher consistency of the German orthography allows easier access to the phonemic structure in German. In contrast, English is considered a highly inconsistent orthography and hence, additional support in PA may yield larger effect sizes in children learning to read and write in English (Goswami, 2002; Landerl et al., 2013; Ziegler et al., 2010).

Interestingly, Bus and van IJzendoorn (1999) showed that PA training was more effective, with regards to PA outcome measures, if adminisered in small groups (*d* = 1.15) than individual training (*d* = 0.53). However, this modulating effect for the size of the training group was not observed in outcomes measures of reading. A second meta-report by the NRP (2000) also found that PA training yields larger effects on PA measures if administered in small groups (*d* = 1.38), as compared to individual tutoring (*d* = 0.60) and classroom-based instruction (*d* = 0.67). The NRP also reported a greater effect for reading measures for children taught in small groups of two to seven children (*d* = 0.81), as compared to classroom- and individual teaching (*d* = 0.30 and *d* = 0.45). The authors argued that the most plausible explanation for this finding is that PA activites in small groups enhance attention and social motivation, and create observational learning opportunities (Ehri et al., 2001).

2.3.2 Supporting phonological awareness in L2 learners

As mentioned above, research on the effectiveness of early literacy support for L2 learners is sparse. Most intervention studies aiming to support L2 learners to learn to read and spell have been conducted after L2 learners have been formally introduced to literacy and included explicit training in reading or spelling (cf. Lovett et al., 2008; O'Connor, Bocian, Beebe-Frankenberger, & Linklater, 2010; Oxley & de Cat, 2019; Richards-Tutor et al., 2016; Solari & Gerber, 2008; Vadasy & Sanders, 2010; Vaughn, Cirino, et al., 2006; Vaughn, Mathes, et al., 2006). To the best of my knowledge, no studies examined the effectiveness of interventions training exclusively PA in L2 learners before children had been formally introduced to literacy. However, a few studies looked at the effectiveness of combined PA and LSK training for L2 learners and will be discussed below (see section 2.3.4).

2.3.3 Supporting letter-sound knowledge in L1 learners

Becoming acquainted with letters is a major landmark in alphabetic literacy acquisition (Foulin, 2005). However, to my knowledge, no rigours intervention studies have been conducted focussing on the effectiveness of interventions targeting only LSK. Intervention studies always delivered a combined training of LSK with PA and are discussed below (section 2.3.5). However, one study included a control group who were administered an intervention only targeting LSK. Schneider et al. (2000) compared the effectiveness of three types of preschool interventions (i.e. PA alone, LSK alone, and PA combined with LSK) for at-risk children (age 5-6) in Germany. The group averages of the three tasks were compared to each other and to a non at-risk control group following the regular preschool programme in Germany. Results indicated that the group who was only trained in LSK showed the highest raw mean scores in LSK immediately postintervention, even higher than the non at-risk control group. However, the difference to the other groups was not statistically significant. Yet the LSK alone group performed significantly lower than the other intervention groups on measures of PA immediately post-intervention. In Grade 1, the LSK alone group still scored lower than the other groups on measures of PA and the difference was approaching statistical significance ($p = .070$). With regards to measures of reading and spelling in Grade 1, and in Grade 2, the LSK training group showed comparable performances to the group who was only trained in PA, but lower performances than the group who received a combined training of LSK and PA, or the non at-risk control group.

2.3.4 Supporting letter-sound knowledge in L2 learners

To my knowledge, no rigorous study has been conducted training exclusively LSK (without PA) in an L2 population. All studies training LSK in L2 learners also included training of PA and are described below (see section 2.3.6).

2.3.5 Supporting phonological awareness and letter-sound knowledge in L1 learners

As mentioned above, Schneider et al. (2000) compared the effectiveness of three training groups (i.e. PA alone, LSK alone, and PA combined with LSK) in children identified as at-risk of reading difficulties in Germany (age 5-6). At post-test, children who received a combined training of PA and LSK performed lower than the PA alone group on measures of PA, but higher than the LSK alone group and an untrained non at-risk control group. No significant differences between the groups were observed for LSK. However, notably, in Grade 2, the PA alone and the LSK alone at-risk groups still performed significantly lower than the untrained non at-risk control group on measures of word reading, reading comprehension and spelling. Yet no differences on the literacy measures could be observed between the at-risk group that was administered a PA combined with LSK training and the untrained non at-risk group. Thus, the study provides strong evidence that PA training combined with LSK shows larger transfer effects to measures of reading and spelling than only training PA or LSK.

Bus and van IJzendoorn's (1999) meta-review confirmed that a purely PA training (d = 1.19) showed smaller effects on PA outcome measures than a training combining PA with LSK (d = 1.75). With regards to word reading outcome measures, a purely PA training showed smaller effects (d = 0.18) than a training of PA involving letters (d = 0.66). Interestingly, a second meta-review by the NRP (2000) reported similar results and concluded that, PA trainings including letters were not necessarily more effective than trainings that excluded teaching letters on measures of PA (d = 0.89 with letters vs. d = 0.82 without letters), but showed a greater transfer to reading measures at post-test (d = 0.67 vs. d = 0.38) and at delayed follow-up (d = 0.59 vs. d = 0.36). Similar results were reported by a meta-review of PA intervention studies in German (Fischer & Pfost, 2015). Fischer and Pfost (2015) found that PA trainings that did not included the training of LSK showed larger effect sizes (d = 0.61) on measures of PA than studies that did include LSK training (d = 0.25). However, preschool PA interventions without the training of LSK presented lower meta-effect size on measures of reading and spelling (d = 0.19) than interventions including LSK training (d = 0.31).

In all, the review of the literature draws a conclusive picture. PA training and LSK training can be effectively administered to children before they formally learn to read and write, and while PA training without letters can effectively support PA in pre-schoolers, it is the combined training of PA with LSK that shows the largest transfer effects to measures of reading and spelling.

2.3.6 Supporting phonological awareness and letter-sound knowledge in L2 learners

As discussed above (see section 2.3.2 & 2.3.4), there is limited research on how to support L2 learners in literacy development. Intervention studies aiming to support L2 learners in learning to read and spell have mostly been conducted after L2 leaners have been formally introduced to literacy, which means they included explicit training in reading or spelling (Lovett et al., 2008; O'Connor et al., 2010; Oxley & de Cat, 2019; Richards-Tutor et al., 2016; Solari & Gerber, 2008; Vadasy & Sanders, 2010; Vaughn, Cirino, et al., 2006; Vaughn, Mathes, et al., 2006). Only very few studies looked at the effectiveness of interventions targeting PA with LSK in L2 learners before children had been formally introduced to literacy.

Stuart (1999) aimed to support literacy development of L2 learners in inner-city schools in London in Grade 1 (age 5). Out of a sample of 112 pupils, only 16 (14%) children spoke English as an L1. Stuart (1999) explored the effectiveness of PA combined with LSK training in an active control design. The intervention in the active control group consisted of a “Big Book” intervention, focusing on drawing children's attention to written words in texts, talking about letters in words, introducing letter-sounds, and encouraging children to notice and read words in the classroom environment. The interventions were administered by mainstream teachers on a classroom basis over 12 weeks with daily session of one hour. Results at post-test showed that children from the PA combined with LSK group outperformed children from the “Big Book” group on phoneme segmentation skills and LSK, and the differences remained significant in a delayed follow-up one year later. Concerning measures of reading and spelling, positive intervention effects in favour of the group who received the PA and LSK were observed immediately postintervention, as well as one year later in Grade 2.

A second notable study on early literacy skills in L2 learners was conducted in Germany. Weber, Marx, and Schneider (2007) explored the effectiveness of an PA with LSK intervention in German preschools for children (age 5-6) with a migration background. The training was administered in small groups of four to eight children and ran over 20 weeks. However, the study did not include an untrained control group as most teachers in German preschools already administered such early literacy training, and the teachers did not want to withhold it from their pupils. The intervention results on the L2 learners were compared to the intervention results of a control group of L1 learners who were administered the same intervention programme. Immediately post-intervention, the authors found that the intervention led to a comparable progress on PA measures from pre- to post-test for the L1 and the L2 group. However, the L1 learners still outperformed the L2 children on measures of phonemic awareness at the end of preschool, and also on literacy measures in German at the end of Grade 1. Notably, on a non-word spelling tasks developed by the authors, the L2 learners showed equal performance to the L1 learners. Although the authors admit that only tentative conclusions can be drawn from their study due to the trained control group, the authors posit that combined training of PA with LSK should be effective for L2 leaners.

Although the early literacy intervention was not administered in an L2, noteworthy evidence for the effectiveness of PA combined with LSK training in multilingual settings comes from Wawire and Kim (2018). In a multicultural cosmopolitan region in Kenya, children speak an ethnic home language and Kiswahili, and learn to read and write in this ethnic language. English is taught as a subject in Grade 1. Wawire and Kim (2018) used an RCT study design to investigate whether effects from an eight week training in PA and LSK in Kiswahili[3] in Grade1, transferred to PA and LSK in English. The training was administered in addition to their formal literacy instruction in Grade 1. Children from the intervention group received support in PA and LSK in groups of three pupils in sessions of 20 minutes three times per week. The control condition followed the standard Grade 1 curriculum in Kenya. The training in Kiswahili showed positive training effects on PA and LSK in Kiswahili, but the intervention group also outperformed the control group on measures of PA and LSK in English. Interestingly, path analyses revealed distinctively divergent patterns for different outcome measures. For instance, when PA was assessed by the same type of task in Kiswahili and English (i.e. a blending measure), the training effect was completely mediated by performance in Kiswahili. This provides direct causal evidence for a cross-language transfer of

[3] Kiswahili has an alphabetic orthography and is anchored in the Roman alphabet with a slightly different set of letters and graphemes. It is considered a highly consistent orthography.

PA. In addition, when PA was assessed with different types of tasks in the two languages (an oddity task in Kiswahili and a detection task in English), improvement in the Kiswahili oddity task predicted the performance on the English detection task. In all, Wawire and Kim (2018) provided evidence for PA transfer effects between languages and support for PA being a language independent meta-cognitive skill serving multiple languages. However, conversely to what theory would insinuate, trainings effects of PA training did not generalise to performance on reading tasks, neither in Kiswahili nor in English. The authors explained this lack of an effect on reading skills by the short duration (only 8 weeks) and the low intensity of the intervention (sessions of 20 minutes, three times per week).

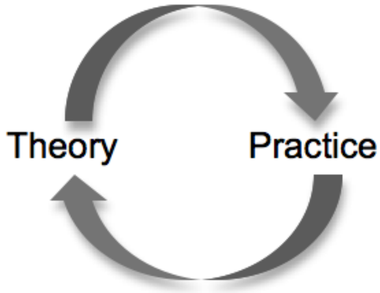
2.4 Summary on effective early literacy interventions

Intervention studies have shown that PA and LSK can be trained before formal literacy instruction in L1 and L2 learners, and that preschool support of PA and LSK generalises to literary measures in Grade 1 (Bus & van Ijzendoorn, 1999; National Reading Panel, 2000; Schneider et al., 2000). Notably, preschool support combing PA and LSK has shown larger transfer effects to measures of reading and spelling than training focusing on only PA or only LSK. Particularly the training of the level of the smallest linguistic unit of PA, i.e. phonemic awareness, has shown to be effective in supporting the development of literacy (Ehri et al., 2001; Fischer & Pfost, 2015). Awareness of the individual phonemes has been directly linked to literacy as it enables children to recognize that words are made up of individual sounds that are represented by letters (see section 5.2 for a more specific explanation on PA and its role in learning to read). Wawire and Kim (2018) also argue that the metalinguistic awareness about how graphemes represent phonemes is likely to increase a cross-linguistic transfer of training early literacy skills. Despite sounds of letters being orthographic specific, the recognition of the alphabetic principle is a metalinguistic awareness and applicable in all alphabetic languages. No training consists of a magic solution that eradicates all later reading difficulties, but so far no other early training approach has shown to be as promising in preventing early word reading difficulties than the training of PA and LSK.

2.5 Importance of evidence-based interventions

Before a particular preventative or remedial approach should be implemented at scale, it needs to be ensured that it is well-founded and evidence-based (Duff & Clarke, 2011). Ideally, programmes used to teach children in classrooms or to support children with difficulties should be both embedded in a theoretical model and tested empirically (Asmussen, Brims, & Mcbride, 2019; Carroll, Bowyer-Crane, Duff, Hulme, & Snowling, 2011). Yet theoretical models and intervention studies are intertwined, as effective intervention studies always provide supporting evidence for theoretical models. Carroll et al. (2011) refer to this reciprocal process as the "virtuous circle" between theory and practice (see figure 2.1).

Figure 2.1: *The virtuous circle between theory and practice*
(adapted from Carroll, et al., 2011, figure 2.1, p. 18)



It is not possible to know to what extent an intervention can be effective until it is evaluated in practice. However, many teaching and intervention programmes to date are administered in schools and real world settings, despite the lack of rigorous evidence of their effectiveness (Strong et al., 2011)***.*** For example, Strong et al. (2011) estimated that, during the first 10 years since its launch, over 570,000 children in more than 3,700 schools in the US were trained on a commercialized product to promote children's reading and oral language skills, without any peer-reviewed evidence supporting its effectiveness. Clearly, for intervention programme to be considered and endorsed as a suitable training for children in schools, it needs to be tested rigorously (Asmussen, Brims, & Mcbride, 2019).

Traditionally, the most rigorous evaluation of the effectiveness of an intervention is provided by randomized control trials (RCTs), which are seen as the gold standard for evidence in medical sciences (Carroll et al., 2011). The *Early Intervention*

Foundation (Asmussen, Brims, & Mcbride, 2019) recent published a report describing a 10 step guide to the evaluation and quality insurance of early interventions. The authors of the report view RCTs and quasi-experimental designs as the most robust method of attributing causality to an intervention model.

2.6 The current study

RCTs are considered the gold standard in intervention research. However, RCTs are seen critical by some people working in the field of education (Biesta, 2007) and are also completely absent from educational research in Luxembourg. Thus, the current study adopts a quasi-experimental design to explore the efficacy of a newly developed intervention targeting early literacy skills in Luxembourgish for a sample of linguistically diverse children growing up in Luxembourg. The intervention is delivered by regular classroom teacher to the children of Year 2 of preschool and targets PA and LSK. Details on the early literacy intervention are described in section 3.9. The control group follows the regular preschool programme in Luxembourg.

It will be explored whether the intervention successfully trained early literacy skills in Luxembourgish and whether intervention effects in Luxembourgish generalise to reading and spelling measures in German in Grade 1. The efficacy of the intervention for all children in the classrooms will be explored. However, the *Early Intervention Foundation* (EIF) (2019) states in their report *10 steps for evaluation success*, that post-hoc subgroup analysis of the efficacy of the intervention are a promising approach to gain a better understanding for whom the intervention worked best. The EIF emphasises that such subgroup analysis can be fundamentally *post-hoc*, meaning that they are performed after the viewing of the data and are no true tests of the original evaluation of efficacy. However, the authors argue that findings from such a sub-analysis can usefully inform future research (Asmussen, Brims, & Mcbride, 2019). Hence, the efficacy of the intervention will also be explored for a subgroup of children identified as having low oral language proficiency in the instruction language. The question whether especially children with low oral language skills in the language of instruction (i.e.

Luxembourgish) benefitted from the intervention was considered important for two reasons. Firstly, these children are considered a large at-risk population in the context of Luxembourg (Hoffmann et al., 2018), and secondly, it is possible that children with limited proficiency in the school language in preschool might not be able to fully engage in the intervention activities, which in turn might reinforce their disadvantage further.

The specific research questions associated with the evaluation of the early literacy intervention are as follows:

1) Can a structured training targeting early literacy skills in preschool (targeting PA and LSK) improve children's early literacy skills in Luxembourgish in children from Luxembourg, many of whose first language is not Luxembourgish?

2) Will any effects of the early literacy training in Luxembourgish be maintained nine months after the training has finished in Grade 1?

3) Will the early literacy training in Luxembourgish show any transfer effects to children's PA skills in German in Grade 1?

4) Will the early literacy training in Luxembourgish show any transfer effects to children's responsiveness to literacy instruction in German in Grade 1?

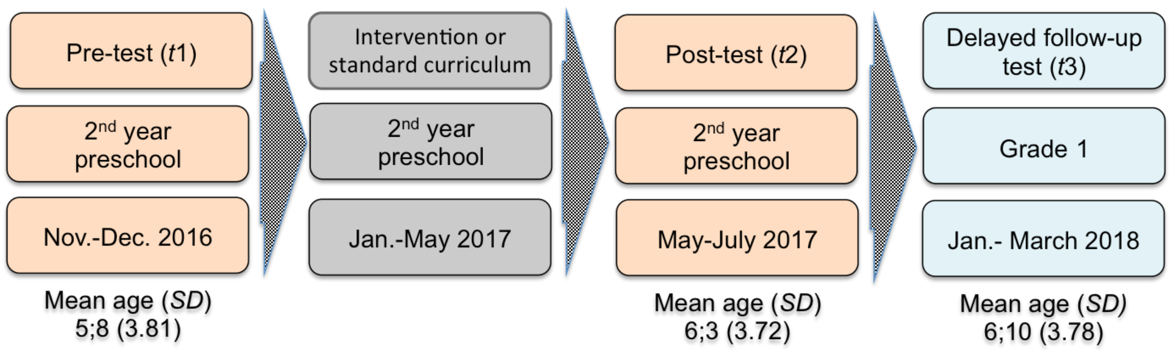
Research question for the *post-hoc* analysis for the sample with low oral language proficiency in Luxembourgish:

5) To what extent is the early literacy intervention in Luxembourgish beneficial for children with low oral language proficiency in the language of preschool instruction?

Chapter III - Method: Intervention study

To assess the efficacy of the classroom-based early literacy intervention, 189 children from 28 preschool classes were followed over the course of 14 months from Year 2 of preschool to Grade 1 of primary school. The study was set up as quasi-experiment with a matched group design and children were assessed on three different occasions: before the intervention in preschool (*t*1), immediately after the intervention in preschool (*t*2) and nine months after the end of the intervention in Grade11 (*t*3). The control schools followed the standard preschool curriculum in Luxembourg. Figure 3.1 shows an overview of the study design.

Figure 3.1: *Timeline of study showing the three testing points, the educational phase of the children, mean ages and SDs (in months)*



Eight schools and 28 classrooms participated in the study. To minimize risk of contamination and John Henry effects, the decision was taken against the allocation of classrooms within a school to either the intervention or the control group (Rhoads, 2011; Saretsky, 1972). Instead, entire schools were assigned to either the intervention or the control condition. The intervention was delivered over twelve weeks between the 30th of January 2017 and the 15th of May 2017 to all the children who were in their second year of preschool (Cycle 1.2). Ethical approval for the study was obtained from the Ethics Review Panel (ERP) of the University of Luxembourg. The study was approved by the Ministry of Education in Luxembourg (*Ministère de l'Éducation nationale, de l'Enfance et de la Jeunesse,* MENJE), the schools' administrative districts directors, the preschool coordinators and the classroom teachers. The study was also registered with the National Commission for Data Protection in Luxembourg (*Commission Nationale pour la Protection des Données,* CNPD). Informed caregiver consent was obtained for the testing of the children for all the study phases. As the intervention was approved by the Ministry of Education, the regional school authorities and the teachers of the preschools, no caregiver consent for taking part in the intervention was required for the delivery of the intervention.

3.1 Participants

A priori power analysis using the G*power3 computer programme (Faul, Erdfelder, Lang, & Buchner, 2007) was conducted to determine the required sample size. The estimate of the population effect size was taken from a recent meta-review on the effectiveness of early literacy intervention studies in German (Fischer & Pfost, 2015). The meta-analysis indicated a small overall intervention effect (Cohen's *d* for 0.39) for phonemic awareness. The power analysis (two tailed) for the differences between two independent groups indicated that a sample of 172 children is needed to detect a small effect size (d = 0.39) with 95% power with using a *t*-test with alpha at 0.05. Hence, the aim was to recruit a sample of at least 172 children.

Participants were recruited from public preschools in Luxembourg. Nine schools from three school districts (region 12, 13 and 14) were identified based on the Luxembourg Ministry of Education's public school database. The aim was to select schools that would lead to a sample which is broadly representative of the wider pupil population in Luxembourg. Only small to medium sized public preschools from rural areas (i.e. village schools) in the Central and the North regions of Luxembourg were targeted. Preschools were selected to be relatively similar in terms of infrastructure, available resources, teaching methods, percentage of L2 learners and class sizes. All preschools followed the same national curriculum (*plan d'études*, MENFP, 2011). Another selection criterion was that schools did not participate in other research projects or specialized programmes. The respective district directors were contacted by email and the project was presented to the district director in a meeting of 90 minutes. The classroom teachers were contacted by email and the project was orally presented in teacher meetings of 90 minutes at each school. Teachers had to consent to participate in the study without knowing whether their school would be allocated to the intervention or control condition to ensure equivalent teacher motivation to carry out the intervention across the groups. This led to the drop out of one school who only agreed to participate as a control school. Teachers also needed to authorize the research team to observe four individual intervention sessions without prior notification in case they intended to participate. All children in second year of preschool of the selected schools were invited to take part in the study (*N* = 201). In total, 201 informed consent forms were sent out. Twelve caregivers did not give consent for their child to participate in the testing. The final sample consisted of 189 children. The sample included 104 boys (55%) and the age ranged from 5;2 – 6;9 (mean 5;8) at pre-test, 5;9 – 7;4 (mean 6;3) at post-test, and 6;4 – 7;11 (mean 6;10) in Grade 1.

3.1.1 Group allocation

Allocation to the intervention and control condition was done on the school-level. Steps were taken to ensure that the groups would be as similar as possible on important confounding variables. Based on data from the Ministry of Education schools were paired on school size, number of L2 learners and SES. School pairs were created that minimise differences on these possible confounding variables. Out of each pair, one school was allocated to either the intervention or the control group. See table 3.1 for an overview of the school pairings.

Table 3.1: *Overview of the demographic data of the preschool pairings based on the data from the Ministry of Education (number of children per school, number of L2 learners (%) and SES*

	Intervention schools			Control schools		
School pairs	Size[1]	L2 learners (%)	SES[2]	Size[1]	L2 learners (%)	SES[2]
1	30	17 (57%)	5	38	16 (42%)	6
2	23	16 (70%)	4	22	13 (59%)	4
3	14	9 (64%)	3	14	5 (36%)	5
4	26	11(42%)	4	24	17 (70%)	5

Note. [1]Number of children in the first year of preschool the year before the intervention was administered. These children were in the second year of preschool when the intervention was administered.
[2]Average ISEI of Municipality on 6 point Likert scale:
1 = 35 ≤ 40; 2 = 40 ≤ 45; 3 = 45 ≤ 50; 4 = 50 ≤ 55; 5 = 55 ≤ 60; 6 = 60 ≤ 65.

3.1.2 Background variables of participants

Background questionnaires were sent out to the main caregivers of the 189 children for whom consent was obtained (see appendix A). The questionnaire was developed for the purpose of this study and was available in five languages (i.e. Luxembourgish, German, French, English and Portuguese). It sought information on the following areas:

- Home language- and migration background,
- Information on SES, e.g. number of books at home, parental education and occupation of parents.

In total, 177 out of 189 background questionnaires (94%) were returned.
On the basis of this questionnaire SES, and language learner status was determined: Responses to an open-ended occupational questions were coded to four-digit ISCO codes, which was transformed into the International Socio-Economic Index of occupational status (ISEI-2008, Ganzeboom, De Graaf, & Treiman, 1992). The higher ISEI score of either parent or the only available parent's ISEI score was used as dependent variable. The average ISEI-08 score of 54.4 (*SD* = 22.4) and ranged from 14 (agricultural farmers, cleaners) to 89 (medical doctors, managers). Children were classified as L2-learners if they predominantly spoke another language than Luxembourgish at home. Table 3.2 presents information on the linguistic background of the children of the intervention and the control group based on the returned care-giver questionnaires.

Table 3.2: *Overview of the languages spoken at home by the children in the intervention and control group*

Intervention Group (*n* = 83)		Control Group (*n* = 94)	
Language	*n* (%)	Language	*n* (%)
Luxembourgish	45 (54%)	Luxembourgish	60 (63%)
Portuguese	18 (22%)	Portuguese	15 (16%)
French	9 (11%)	French	6 (6%)
Créole	3 (4%)	German	3 (3%)
Polish	2 (2%)	Bosnian	2 (2%)
English	1 (1%)	Polish	2 (2%)
Spanish	1 (1%)	Russian	1 (1%)
Italian	1 (1%)	Italian	1 (1%)
Serbo-Croatian	1 (1%)	Serbo-Croatian	1 (1%)
German	1 (1%)	Slovakian	1 (1%)
Wolafs	1 (1%)	Chinese	1 (1%)
		Créole	1 (1%)

Luxembourgish was the dominant language spoken by the children in the intervention and the control group (54% and 63%), followed by Portuguese (22% and 16%) and French (11% and 6%). This is broadly representative of the wider pupil population in Luxembourg. According to national school data, Luxembourgish-speaking students represent 56%, Portuguese-speaking students represent 19,9% and the French-speaking students represent 6,9% of the student population in Luxembourg preschool (MENFP, 2015). In total, 15 different languages were spoken by the sample in the home context.

3.2 Testing procedure

Each child was tested individually in a quiet area outside the classroom in the school. At pre- and post-test, children were tested in two sessions of approximately 25 minutes each. All pre-test (*t*1) and post-test (*t*2) measures were administered in Luxembourgish. At delayed follow-up (*t*3), children were tested individually in three sessions of approximately 30 minutes. The first session at delayed follow-up tapped oral language and early literacy skills in Luxembourgish. Testing sessions two and three tapped oral language, early literacy and literacy skills in German. Children never completed more than one testing session on the same school day to provide optimal performance on all tasks.

Assessments were administered by the author with the help of two trained research assistants at *t*1 and *t*2, and with the help of three trained research assistants at *t*3[4]. Different research assistants were recruited for each assessment phase. All research assistants were native Luxembourgish speakers, fluent in German and had prior experience in working with children. They were either students in the psychology bachelor degree at the University of Luxembourg and/or had a background in teaching children with special educational needs, social care work, or speech and language therapy. Prior to the assessment phase, all research assistants were given a three-hour training session by the author, followed by individual practice in the administration of the assessments without children. In addition, a testing-out session in a preschool that was not part of the project was arranged prior to each assessment phase. Consent for the testing-out training was obtained from the schools' administrative districts directors, the teachers and the parents of the children. The extensive training was done to ensure that research assistants were familiar with the testing procedures and testing was administered homogenously across the assistants. On their first testing session at each assessment phase, the research assistants were observed by the author and given final corrective feedback in administration of the tasks. All participating children received a sticker after each testing session and a diploma after each assessment phase.

3.3 Measures

Children were administered a comprehensive test battery tapping into oral language, early literacy skills, cognitive measures, literacy skills and arithmetic skills. The aim was to use standardized rather than experimenter-developed measures as the former tend to yield less biased effect sizes (Coolican, 2009; Innocenti et al., 2014). However, no standardized tests in Luxembourgish exist, hence, all the tests in Luxembourgish needed to be newly developed or adapted from existing German or English tests. Standardized tests were used to assess PA and literacy skills in German in Grade 1. As no norms on these measures exist

[4] The author tested 88 out of 189 (47%) children at t1; 65 out of 185 children (35%) at t2;. 51 out of the 172 (30%) children.

(the norming sample are L1 German speakers), raw scores will be used in all the analyses.

Measures are divided into primary outcome, secondary outcome and control measures. Primary outcome measures tapped the early literacy skills directly targeted by the intervention, i.e. PA in Luxembourgish and LSK. Secondary outcome measures tapped into domains that were indirectly targeted by the intervention, i.e. PA in German and literacy skills. Control measures focused on skills that show a relation to early literacy acquisition, but that are not expected to be influenced by the intervention, i.e. non-verbal reasoning, number naming and vocabulary. Control measures were important in the examination of potential confounding factors. A number naming task was included as a control measure to further check for specificity of effects. Table 3.3 presents an overview of individual tasks tapping primary and secondary outcome measures, and control measures for each testing point.

Table 3.3: *Overview of primary, secondary and control measures for each testing point (t1, t2, t3)*

	Pre-test (*t*1)	Post-test (*t*2)	Delayed follow-up (*t*3)
PRIMARY OUTCOMES MEASURES			
PA Luxembourgish			
Syllable segmentation	x	x	---
Rhyme detection	x	x	---
Onset–rhyme blending	x	x	---
Onset identification	x	x	---
Phoneme blending	x	x	x
Onset manipulation	x	x	x
Phoneme segmentation	----	x	x
Letter-sound knowledge	x	x	x
SECONDARY OUTCOME MEASURES			
PA German			
Phoneme blending	---	---	x
Onset manipulation	---	---	x
Phoneme segmentation	---	---	x
Literacy skills			
Basic[1] word reading	x	x	x
Basic[1] non-word reading	x	x	x
Word reading	---	---	x
Non-word reading	---	---	x
Reading comprehension	---	---	x
Spelling	---	---	x
CONTROL MEASURES			
Receptive vocabulary			
Luxembourgish	x	x	x
German	---	---	x
Number naming	x	x	---
Non-verbal reasoning	x	---	---

Note. x = administered, --- = not administered.
[1]Basic reading refers to the ability to read words or non-words that are identical in German and Luxembourgish, with only regular grapheme to phoneme correspondences ranging from two to five graphemes in length.

3.4 Pilot study

The test battery was piloted on 12 children (age 5-6) in Year 2 of preschool. Participants were recruited from two classrooms in one school. Consent for the piloting was obtained from the schools' districts directors, the teachers and the parents of the children. Testing procedures were the same as for the main study. Testing took place in a quiet area outside the classroom in the school. Each child was tested individually in two sessions of approximately 25-30 minutes each on different school days. The pilot study confirmed that the time allocated to each test was adequate, that most task instructions were clear and that the majority of the measures were ready for full-scale implementation. Cronbach's alphas for the pilot data were above .75 for all instruments, yet caution needs to be paid to the interpretation of the internal consistency measure due to the small sample size ($N = 12$).

The pilot study led to minor modifications of materials or procedures of some measures. The standard procedure of the German *Test für Phonologische Bewusstheitsfähigkeiten* (TPB) [Test for phonological awareness skills] (Fricke & Schaefer, 2008) foresees that pictures of test items of the input tasks (i.e. task that can be completed nonverbally by pointing) are not named by the examiner. The test developers specified this procedure to avoid that the child receives auditory input, which would restrict the reliance on the children's own lexical representations when completing the phonological input tasks (Schaefer et al., 2009). Only the pictures of the training items are named by the examiner in the original TPB (Fricke & Schaefer, 2008). However, the pilot study showed that this procedure, developed in the context of L1 German-speaking children, was too challenging for children in Luxembourg with limited lexical knowledge in German. Children did either not point at any picture or asked the examiner to name the pictures. This increased the time required to complete the tasks substantially and led to a noticeable decrease in motivation on the following tests. Hence, the decision was taken to not only name the pictures during the training, but to also name the pictures for the main test items.

Number knowledge in the pilot was assessed via a number naming task comprising 11 numbers between 1-20 and the numbers 50 and 100. The pilot revealed ceiling effects on this task (possible max score = 13, *M* = 11.2, *SD* = 1.9). Out of 12 children, eight (67%) scored the two highest possible scores on this task. The number naming task was therefore increased in difficulty by expanding the range of numbers to 16 numbers between 1-101.

The pilot confirmed floor effects on the onset manipulation task (*M* = 1.8, *SD* = 0.0). Ten out of twelve children (83%) scored zero. A similar pattern was observed for the basic word reading and basic non-word reading tasks with ten out of twelve children (83%) scoring zero in each task (basic word reading: *M* = 1.7, *SD* = 2.8; basic non-word reading: *M* = 1.9, *SD* = 0.7). These effects were anticipated as children had not been formally introduced to literacy yet by the time of the piloting. These tasks were, however, retained in the final test battery as they served as a baseline to measure developmental progress. The tasks included early discontinuation criteria, and therefore did not result in frustration.

3.5 Primary outcome measures

3.5.1 Phonological awareness in Luxembourgish

A new PA assessment tool was developed for the purpose of the present study. Task development was based on the German '*Test für Phonologische Bewusstheitsfähigkeiten'* (TPB) [Test for phonological awareness skills] (Fricke & Schaefer, 2008). The new assessment contains seven subtests that cover four linguistic units (i.e. syllable, rhyme, onset-rhyme and phoneme) and different levels of explicitness (i.e. identification, segmentation, blending and manipulation). Five subtests required spoken answers from the children, viz. output tasks, and two subtests could be completed nonverbally by pointing, viz. input tasks. However, completing an input test verbally was also scored as correct. An overview of the seven subtests can be found in table 3.4.

Table 3.4: *Overview of the linguistic units and levels of explicitness tapped by the subtests of the PA assessment battery in Luxembourgish for each testing point*

				time points		
Test	Linguistic unit size	Level of explicitness	Type of task*	*t*1	*t*2	*t*3
1	Syllable	Segmentation	Output	✓	✓	
2	Rhyme	Identification	Input	✓	✓	
3	Onset-rhyme	Blending	Output	✓	✓	
4	Onset identification	Identification	Input	✓	✓	
5	Phoneme	Blending	Output	✓	✓	✓
6	Onset / phoneme	Manipulation	Output	✓	✓	✓
7	Phoneme	Segmentation	Output		✓	✓

*output tasks required a spoken answer and input tasks could be completed nonverbally by pointing

Whenever possible, the tests included drawn coloured pictures to reduce working memory load and to increase motivation (Schaefer et al., 2009). The test was computerized and the test procedure for each subtest was as follows: the child completed three training items with corrective feedback by the experimenter. At least one training item had to be scored as correct in order for the main test items to be administered. Each subtest consisted of 12 test items with a repetition of the task instructions half way through the test. No discontinuation criterion was used during the administration of the test items. Apart from general praise, the experimenter was not allowed to give corrective feedback during the twelve test items. A correct answer received a score of one and the total maximum score for each PA subtest was 12.

Syllable segmentation (t1 - t2). Children were shown a picture on a computer screen. They were asked to name the picture and then they had to segment the depicted noun into its constituent syllables. If a child was unable to name the picture, the experimenter named the depicted word once. Stimuli containing double consonants were avoided due to ambiguous syllable segmentation. The length of the stimuli varied from two to four syllables, each represented by four items. Item examples of the syllable segmentation test: <Lee-der> /leː-dɐ/ (ladder), <Te-le-fon> /tə-lə-fõː/ (telephone).

Rhyme identification (t1 - t2). This task required children to look at a stimulus picture at the top of the computer screen and select the correct rhyming word amongst three possible drawings underneath by pointing. The three answer choices consisted of the correct item, a semantic distractor (e.g. hyponym, hypernym) and a phonological distractor. The phonological distractor was matched phonologically and in vowel length to the stimulus (max. two phonemes were different to the stimulus). The stimulus word and the answers were named once by the experimenter. See table 3.5 for item examples of the rhyme identification subtest.

Table 3.5*: Item examples of the rhyme identification subtest*

stimulus	target	phonol. distractor	semantic distractor
Dësch */dəʃ/* (table)	Fësch /fəʃ/ (fish)	Posch /poʃ/ (bag)	Schaf */ʃaːf/* (cupboard)
Hond */tsɑnt/* (dog)	Mond */mont/* (mouth)	Zant */tsɑnt/* (tooth)	Kaz */kaːts/* (cat)
Hues /huəs/ (rabbit)	Nues */nuəs/* (nose)	Haus */hæːʊs/* (house)	Kou */kəʊ/* (cow)

Onset-rhyme blending (t1- t2). The experimenter said a noun with a pause of one second in between the onset and the rhyme unit. The child had to blend the onset and the rhyme together and pronounce the target word. The test items consisted of only monosyllabic words of which the rhyme contained at least two phonemes. Four stimuli onsets were a fricative / nasal / liquid, a plosive or consonant cluster.

Item examples of the onset-rhyme blending test: <F-ouss> /fəʊs/ (foot), <T-ass> /ta:s/ (mug), <Br-oud> /Brəʊt/ (Bread).

Onset identification (t1 - t2). Children were presented with a stimulus picture at the top of the page on the computer screen and had to select the word beginning with the same onset amongst three possible answer stimuli underneath by pointing. The possible answer choices consisted of the correct item, a phonological distractor and a semantic distractor. The two matching words shared either, a single consonant (e.g. Luucht /luːχt/ (lamp) - Léiw /lɜɪf/ (lion)), a two letter consonant cluster (Traap /tʀaːp/ (stairs) - Tromm /tʀom/ (drum)), or the first consonant of the two consonant onset clusters (Broud /bʁoːt/ (bread – Blaat /blɑːt/ (leaf), each represented by four items. The onset of the phonological distractor differed from the target in either place or manner of articulation (voicing was not considered). To avoid confusion with the rhyme task, the target and the distracters never rhymed with the target. The semantic distractor was semantically related to the stimulus word (e.g. hyponym, hypernym). The stimulus word and the three possible answers were named once by the experimenter. Item examples are presented in table 3.6.

Table 3.6: *Item examples of the onset identification subtest*

stimulus	target	phonol. distractor	semantic distractor
Luucht /luːχt/ (lamp)	Léiw /lɜɪf/ (Lion)	Nues /nuəs/ (nose)	Käerz /kɛːəts/ (candle)
Floss /flos/ (riwer)	Fläsch /flæʃ/ (bottle)	Schleek /ʃleːk/ (snail)	Séi /zɜɪ/ (lake)
Traap /tʀaːp/ (stairs)	Tromm /tʀom/ (drum)	Kroun /kʀəʊn/ (crown)	Haus /hæːʊs/ (house)
Broud /bʁoːt/(bread)	Blaat /blɑːt/ (leaf)	Glas /glɑːs/ (glass)	Kuch /ˈkʊːx/ (cake)

Phoneme blending (t1 - t3). The experimenter said a noun with a pause of one second in between each phoneme. The child had to blend the phonemes and pronounce the target word. The test increased in difficulty by increasing the length of the nouns from two up to five phonemes. Each phonemic length was represented by three stimuli. Consonant clusters in onset position were not included, only a consonant-vowel word structure was adopted for the items. Item examples of the phoneme blending test: <S-ee> /z/ + /eː/ (saw); <T-u-t> /t/ + /uː/ + /t/ (bag), < P-i-r-a-t> /p/ + /i/ + /r/ + /aː/ + /t/ (pirate).

Onset manipulation (t1 - t3). Children had to delete (a) specified initial sound(s) and pronounce the resulting non-word. The children saw a picture on the computer screen. The experimenter named the noun once and subsequently pronounced the initial element to be deleted. The element to be deleted was either a single consonant, a consonant cluster, or the first consonant of a two consonant cluster, each represented by four items. Item examples of the onset manipulation test: C <(N)ues> /(n)uəs/ (nose), CC <(Bl)umm> /(bl)um/ (flower), CC <(T)raap> /(t)ʀaːp/ (stair).

Phoneme segmenting (t2 - t3). Children had to segment nouns into their constituent phonemes. They were presented with a picture on a computer screen. The experimenter named the picture once before the children had to phonemically segment it. The length of the stimuli varied from two to five phonemes, each represented by three stimuli. Item examples of the phoneme segmenting test: <Wo> /v/ + /oː/ (scale), <Nol> /n/ + /oː/ + /l/ (nail), <Mama> /m/ + /ɑ/ + /m/ + /ɑ/ (mummy), <Pilot> /p/ + /i/ + /l/ + /oː/ + /t/ (pilot).

3.5.2 Letter-sound knowledge

Letter-sound knowledge (t1- t3). LSK task was a bespoke measure developed for the purpose of this project. The LSK task evaluated children’s knowledge of 20 letters of the alphabet. The selected 20 letter-sounds are identical in Luxembourgish and in German. Children were presented with monographs on a computer screen and were asked to say the sound or the name of each letter. The uppercase and the lowercase version of each letter were simultaneously presented in the font developed by the Ministry of Education of Luxembourg (i.e. *Drockschrëft*). This font is used in most school books in the early years in Luxembourg. Task administration was in Luxembourgish and children were expected to respond in Luxembourgish. In the rare case of a child responding correctly in French (one of the official school languages in Luxembourg) this was also scored as correct. The total maximum score on the test was 20.

3.6 Secondary outcome measures

3.6.1 Phonological awareness in German

Children completed the phoneme blending and the onset manipulation subtests of the '*Test für Phonologische Bewusstheitsfähigkeiten'* (TPB) [Test for phonological awareness skills] (Fricke & Schaefer, 2008). Administration procedures of the original TPB were followed, with one exception: In line with the testing procedures for the PA tests in Luxembourgish, the experimenter did not only name the pictures of the training items, but s/he also named test items on the onset manipulation subtest. In addition to the two tasks of the TBP (Fricke & Schaefer, 2008), one phoneme segmentation task in German was specifically developed for the purpose of this study. The structure of the phoneme segmentation task followed the structure of its Luxembourgish equivalent (see section 3.5.1). The three subtests in German were output tasks (i.e. required spoken answers from the children). An overview of the PA tasks in German can be found in table 3.7

Table 3.7*: Overview of the linguistic unit and level of explicitness of the PA tests in German in Grade 1*

				time points		
Test	Linguistic unit	Level of explicitness	Type of task*	t1	t2	t3
1	Phoneme (TPB)	Blending	Output			✓
2	Onset / phoneme (TPB)	Manipulation	Output			✓
3	Phoneme	Segmentation	Output			✓

*output tasks required a spoken answer

Phoneme blending (t3). The experimenter said nouns with a pause of one second in between each phoneme. The child had to blend the phonemes and pronounce the target word. The test increases in difficulty by increasing the length of the words from two up to five phonemes. Each phonemic length was assessed by three items. Consonant clusters in onset position were not included, only a consonant-vowel word structure is adopted. A correct answer received a score of one and the total maximum score was twelve. Item examples of the phoneme blending test: <Z-eh> /ts/ + /eː/ (toe); <D-o-s-e> /d/ + /oː/ + /z/ + /ə/ (can), < L-a-s-t-e-r> /l/ + /a/ + /s/ + /tː/ + /ɐ/ (lorry).

Onset manipulation (t3). Children had to delete (a) specified initial sound(s) and pronounce the resulting non-word. The child saw the picture of a word on the computer screen. The examiner named the picture once and pronounced the initial element to be deleted. The element to be deleted could be a single consonant, a consonant cluster, or the first consonant of a two consonant cluster, each represented by four items. A correct answer received a score of one and the total maximum score was twelve. The correct response would always be a non-word. Item examples of the onset elision test: C <(L)ampe> /(l)ampə/ (lamp), CC <(Fl)öte> /(fl)øːtə/ (flute), CC <(T)reppe> /(t)rɛpə/ (stair).

Phoneme segmenting (t3). Children had to segment words into their constituent phonemes. The child was presented with a picture on a computer screen and the examiner named the picture once. The length of the stimuli varied from two to five phonemes, each represented by three stimuli. A correct answer received a score of one and the total maximum score was twelve. Item examples of the phoneme segmenting test: <Schuh> /ʃ/ + /uː/ (shoe), <Rock> /r/+ /ɔ/ + /k/ (skirt), <Kamel> /k/+/aˑ/+/m/+/eː/+/l/ (camel).

3.6.2 Literacy measures

Basic word reading (t1-t3). Twelve words were chosen that are orthographically identical in German and in Luxembourgish. All the words had regular grapheme to phoneme correspondences. The task increased in difficulty by increasing the length of the stimuli from three to five letters, each represented by four items. The twelve words were written in the font of the Ministry of Education (i.e. *Drockschrëft*). The test contained two test sheets with six words on each sheet. The words were uncovered one by one by the experimenter. If the child was able to read one out of the first two words, the step-wise uncovering was continued for the remaining four items on sheet one. If the child could not read one out of the first two words, the remaining four words on the sheet were uncovered all at once and the child was asked whether s/he could read any of those words. If the child could not read any of the words on sheet one, the task was discontinued. If the child could read at least one word on sheet one, the second test sheet was administered. The task administration procedure of test sheet two was identical to sheet one. To be considered as read correctly, it was not sufficient to sound out the word but blending of the sounds was required. Words had to be pronounced according to German/ Luxembourgish sound-letter correspondence rules. A correct answer was scored as one and the total maximum score on the test was 12. Item examples of the basix word reading test: <Bus> /bʊs/ (bus), <Lama> /laːma/ (llama), <Insel> /ɪnzəl/ (island).

Basic non-word reading (t1-t3). Children had to read twelve non-words. Children were told that the list consisted of words that did not really exist (fantasy-words), but could nevertheless be read out loud. The twelve non-words were similar to real words, but did not form a word in none of the four most common languages spoken in Luxembourg, i.e. Luxembourgish, German, French and Portuguese. Only letters introduced in the early literacy intervention were used to build the non-words and a balanced selection of letters was used. Each vowel was represented at least two times and not more than three times, and no consonant was used more than two times. The length of the stimuli varied from two to four letters, each represented by four items. Task administration was identical to the administering of the basic word reading task described above. To be considered as read correctly, it was not sufficient to sound out the non-words, but blending of the sound was required. Words had to be pronounced according to German/ Luxembourgish sound-letter correspondence rules (e.g. the French pronunciation of <Lum> /lym/, was not scored as correct). A correct answer was scored as one point and the total maximum score was 12. Item examples of the singe non-word reading test: <Fi> /fɪ/, <Lum> /lʊm/, <Naro> /naro/.

In addition to the basic-reading tasks above, three standardized reading tasks in German were administered in Grade 1: the word reading and the non-word reading subtests of the *Salzburger Lese- und Rechtschreibtests II* [Salzburg reading and spelling test II]) (SLRT-II, Moll & Landerl, 2010), and the word reading comprehension test of *the Ein Leseverständnistest für Erst- bis Sechstklässler* [reading comprehension test for grades 1–6] (ELFE 1-6, Lenhard & Schneider, 2006). The SLRT-II is a standardized diagnostic test for developmental reading difficulties in German-speaking 1st to 4th graders (approximately six to eleven year-olds). In addition, one standardized German spelling task was administered in

Grade 1 the *Hamburger Schreibprobe* [Hamburg writing sample] (HSP, May, 2002). The HSP assesses orthographic knowledge and primary spelling strategies. The HSP +1 version was used which is appropriate for the middle of Grade 1 in Germany (i.e. children aged 6 to 7 years).

Word and non-word reading (t3). Children completed the word reading and the non-word reading subtest of the *Salzburger Lese- und Rechtschreibtests* [Salzburg reading and spelling test]) (SLRT-II Moll & Landerl, 2010). The test was administered according to the manual. Children had to read out loud as many words, or non-words, as possible within one minute from a sheet of paper containing either 72 words or 72 non-words. Each word, or non-word, read correctly according to German sound-letter correspondence rules was credited one point. The total maximum score was 72 for each subtest.

Reading comprehension (t3). The reading comprehension word subtest from the *Ein Leseverständnistest für Erst- bis Sechstklässler* [reading comprehension test for grades 1–6] (ELFE 1-6 Lenhard & Schneider, 2006) was administered individually. The subtest consists of 72 pictures with four printed words next to each picture. The children are asked to identify the correct word that matches the picture. The test stops after three minutes. Correct answers receive a score of one and the total possible maximum score on the test is 72.

Spelling (t3). Children's completed the *Hamburger Schreibprobe +1* (HSP+1) [Hamburg Writing Sample] (HSP+1, May, 2002). Children had to write four words and a short sentence. The test was scored as prescribed in the test manual. The number of correctly spelled graphemes (maximum = 40) was calculated.

3.7 Control Measures

3.7.1 Vocabulary

Vocabulary in Luxembourgish (t1). Children's receptive vocabulary in Luxembourgish at pre-test was assessed with the Luxembourgish version of the Cross Linguistic Lexical Task of the *LITMUL* battery (*LITMUL-CLT, Language Impairment Testing in Multilingual Setting - Cross Linguistic Lexical Task*, Simonsen & Haman, 2016; Luxembourgish version, Engel de Abreu, 2016). This test had been developed for preschool children in Luxembourg. It contains 40 frequent and early-acquired words. Children are asked to match a spoken word to a picture out of a choice of four. Correct answers are coded as one and the total maximum score was 40.

Vocabulary in Luxembourgish (t1-t3). Children completed an adapted version of the Peabody Picture Vocabulary Test (PPVT-4,Dunn & Dunn, 2007) at each assessment phase. A predetermined fixed set of 40 items (i.e. uneven items from sets two to nine) were administered to all children. Children had to identify a target picture out of choice of four to match a spoken word. Correct answers were coded as one and the total maximum score was 40.

Vocabulary in German (t3). Children completed the standardized published German version of the Peabody Picture Vocabulary Test (PPVT - 4, Dunn & Dunn, 2007, German version by Lenhard, Lenhard, Segere & Suggate). The experimenter says a word and the child had to point to the correct picture out of choice of four. The test consists of 19 sets that increase in terms of item difficulty. Each set comprises 12 test items. The PPVT was administered according to the manual. All children started at set four. If a child made zero or one error in set four testing continued forwards. If a child made more than one error in set four, testing continued backwards. To establish the baseline set, this backward procedure was repeated until children made zero or one error in a set. If the child made more than one error in set one, set one was considered the baseline set and testing was continued forwards by administering the items of set five. Testing was discontinued when a child made a total of eight errors within a set.

3.7.2 Non-verbal reasoning

Matrix reasoning (t1). Children completed the matrix reasoning subtest of the Wechsler Preschool and Primary Scale of Intelligence (WPPSI III - Wechsler, 2002). The child needed to complete matrices by finding the missing piece among four or five possible drawings. Matrices and response options progressively increased in difficulty. The test consisted of three training items and 29 test items. Testing stops after four consecutive mistakes. An accurate response received a score of one and the maximum possible score was 29.

3.7.3 Number knowledge

Number naming (t1-t2). Children were presented with a sequence of 16 numbers on a sheet of paper (4 9 17 20 25 39 47 50 55 61 68 72 77 84 93 100). They were asked to name each number in Luxembourgish. A correct answer received a score of one. The maximum score was 16.

3.8 Reliability of the instruments

To assess internal consistency of the different measures, Cronbach's alphas were computed for the entire sample for all three testing points. For the three timed literacy measures (i.e. the word- and non-word reading tasks of the SLRT II and the reading comprehension task of the ELFE 1-6), Cronbach's alpha was not calculated due to its sensitivity to the number of items completed (Streiner, 2003). For those measures the reported reliability coefficients were taken from the respective test manuals. The reliability coefficients for all the measures ranged from acceptable (> .70) to excellent (> .90) (Nunnally & Bernstein, 1994). Reliability coefficients of the scores on all the measures for the different assessment phases are presented in table 3.8.

Table 3.8: *Internal consistency for all measures at t1, t2 and t3*

	t1	t2	t3
	N = 189	*N* = 185	*N* = 172
PA Luxembourgish			
Syllable segmentation	.77[1]	.79[1]	---
Rhyme detection	.90[1]	.89[1]	---
Onset–rhyme blending	.91[1]	.91[1]	---
Onset detection	.79[1]	.83[1]	---
Phoneme blending	.92[1]	.90[1]	.81[1]
Onset manipulation	.93[1]	.93[1]	.91[1]
Phoneme segmentation	---	.91[1]	.87[1]
PA German			
Phoneme blending TPB	---	---	.85[1]
Onset manipulation TPB	---	---	.91[1]
Phoneme segmentation	---	---	.84[1]
LSK	.95[1]	.95[1]	.79[1]
Literacy			
Basic word reading	.96[1]	.94[1]	.83[1]
Basic non-word reading	.95[1]	.94[1]	.76[1]
Word reading SLRT II	---	---	.98[2]
Non-word reading SLRT II	---	---	.96[2]
Reading comprehension ELFE 1-6	---	---	.95[3]
Spelling HSP+1	---	---	.93[1]
Receptive vocabulary			
Luxembourgish PPVT	.86[1]	.82[1]	.79[1]
Luxembourgish CLT	.79[1]	---	---
German PPVT	---	---	.98[1]
Number naming	.93[1]	.94[1]	
Matrices WPPSI	.80[1]	---	---

Note. --- not administered.

TPB, Test für Phonologische Bewusstheitsfähigkeiten [Test for phonological awareness skills]; SLRT II, Salzburger Lese- und Rechtschreibtests [Salzburg reading and orthography Test]; ELFE 1– 6, Ein Leseverständnistest für Erst-bis Sechstklässler [ELFE 1– 6: A reading comprehension test for students in Grades 1 through 6]; HSP+1, Hamburger Schreib-Probe [Hamburg spelling-task]; PPVT, Peabody Picture Vocabulary Test; CLT, Cross-Linguistic Lexical Task; Matrices WPPSI, Wechsler Preschool and Primary Scale of Intelligence.

Reliability: [1]Cronbach's alpha based on collected data; [2]Parallel forms reliability of original standardized test for children in Grade 2 in Germany; [3]Cronbach's alpha of original standardized test for children in Grade 2 in Germany.

3.9 Intervention programme

Preschool children in the intervention group received the *LALA - Lauter lëschteg Lauter* [LALA many funny sounds] programme. The central aim of the LALA programme was to improve children's PA and LSK, to promote print and book awareness, and to encourage parent involvement in their child's literacy development. A large number of existing intervention programmes and other resources from Luxembourg, Germany, Switzerland, the UK and the US had been reviewed (e.g. *Phonologesch Bewosstheet am Spillschoulsalter* [Phonological Awareness at preschool age] (Bodé, 2004); national preschool curriculum in Luxembourg (MENFP, 2011); *Hören, Lauschen Lernen* [Hearing, Listening, Learning] (Küspert & Schneider, 2008); *Die Alphas* [The Alphas] (Huguenin & Dubois, 2013); *Jolly Phonics* (Lloyd, 1998); *Get Ready4Learning* (Bowyer-Crane et al., 2008); *Sound Linkage* (Hatcher, Duff, & Hulme, 2000); *Nuffield Early Language Intervention* (Fricke, Bowyer-Crane, Haley, Hulme, & Snowling, 2013); and *Wilson Foundations Program* (Wilson, 2005).

The programme follows established principles and approaches that have been shown to be effective in the instruction of L2 learners and linguistically heterogeneous school populations (August & Shanahan, 2005; Richards-Tutor et al., 2016). For instance, every session followed an identic systematic and explicit structure to build up instructional routines. Children's first languages were included on a regular basis and the programme also made frequent use of non-words. Moreover, different difficulty levels of the activities were developed to adapt the activities to skill-level of the children. Meanings of basic words were clarified verbally and with visual support. Close attention was paid to develop culturally appropriate materials and the programme provided extensive opportunities for practice, revision, clarifying meanings of basic words and modulating language demands (August et al., 2005; Richards-Tutor et al., 2016). The LALA programme used a playful approach embedded in a structured and repetitive framework and language rich context. The storyline of the intervention programme revolved around a macaw from Brazil (viz. LALA) who visited the children in the classroom to learn the "funny" sounds of Luxembourgish. LALA was embodied by a hand puppet manipulated by the teachers to boost motivation in children and to increase the playing character of the activities.

3.9.1 Intervention structure

The intervention ran over 12 weeks and contained a total of 48 structured intervention sessions of approximately 25 minutes each (total intervention time: 20h). The classroom teachers administered four intervention sessions per week. The first three sessions of each week introduced new content which was consistently consolidated in the fourth session of the week. The first nine units (≈ week 1 to 2) of the programme consisted of PA activities at the syllable, rhyme and onset-rhyme level. Units 10 to 48 (≈ week 3 to 12) specifically trained phonemic awareness skills by introducing or consolidating the linkage between phonemes and their letters. The phonemic awareness activities included phoneme identification (week 2 to 12), phoneme blending (week 4 to 12) and phoneme segmentation tasks (week 9 to 12). Two consolidation weeks (week 6 and week 12), during which no new letters were introduced, were incorporated in the programme to provide opportunities to further consolidate previously learned content.

All 48 intervention sessions followed the same general structure. Each session started with a short introduction, followed by three core activities (A1, A2, A3) and a closing activity. In the introduction, LALA's rules[5] were repeated and previously introduced sounds were briefly revised. The three core activities followed the structure as follows:

[5] The three LALA rules were adapted from the listening rules of the „Get Ready4Learning" (Bowyer-Crane et al., 2008) and consisted of:
Good Looking – look at the person who is talking
Good Listening – use your ears to listen to what the person is saying
Good Sitting – sit still when you are listening

A1: PA activity consolidating sounds that had already been introduced,

A2: Introduction of a new sound and letter with the support of a story and a short jingle,

A3: Interacting with the new letter (e.g. drawing it in sand, form letters with their body).

Each session was closed with either a review of the activities that had been covered in the session, or a short revision of the newly introduced sound of the session.

3.9.2 Supporting phonological awareness

The PA training of the intervention programme was predicated on the developmental nature of PA from larger (syllables) to small linguistic units (phonemes), and from more implicit phonological tasks (blending) to more explicit task (segmentation -see section 5.1 for more details on the developmental nature of PA) (Anthony, Lonigan, Driscoll, Phillips, & Burgess, 2003; Stackhouse & Wells, 1997). The four first intervention sessions (week 1) introduced the children to the concept of syllables and rhymes. Activities included the blending of syllables to form real words and the forming of rhyme pairs. Intervention sessions five to nine (≈ week 3) focused on the level of the onset-rhyme level. A prototypical onset-rhyme activity consisted of the teacher saying a noun with a pause of one second between the onset and the rhyme unit, and the children had to blend the onset and the rhyme to form the target word.

After session nine, individual letter-sounds were sequentially introduced and all of the remaining intervention sessions (sessions 10 to 48, ≈ week 3 to 12) specifically trained phonemic awareness. Initial sounds identification activities (e.g. finding matching pairs of flashcards with the same initial sound) were administered from week 2 to week 12. Phoneme blending was introduced in week 4 and generally consisted of LALA (the hand puppet) trying to speak. LALA said a noun in Luxembourgish in “parrot language”, i.e. with a pause of one second in between each phoneme, and the children had to blend the phonemes and pronounce the target word. Phoneme segmentation was only introduced in week 9 and consisted of the children trying to imitate “parrot language”. In total 23 sounds were introduced in weeks 2 to 12, but the intervention systematically revised previously introduced sounds. Each sound with its corresponding letter was repeated in at least seven different phoneme awareness activities.

In addition, PA was fostered using short educational jingles. A short jingle that featured the target sound was developed for each sound. In combination with a corresponding movement, the song served as an auditory mnemonic aid. Songs were short (approximately 30 seconds) and were introduced with support of visual material to explain the lyrics of the song. In each song, the lyrics involved a repetition of the targeted letter-sound (e.g. k-k-k k-k-k) and children were encouraged to do a movement corresponding to the letter when the letter-repetition section came. Using singing and movements as a complementary approach to teaching has been shown to increase motivation by making activities more fun and engaging (Rinne, Gregory, Yarmolinskaya, & Hardiman, 2011; Viladot et al., 2018; Walton, 2014).

3.9.3 Supporting letter-sound knowledge

In total 23 different sounds and their corresponding letters or letter combinations were introduced at a rate of three new sounds per week (one new letter-sound per unit). Two consolidation weeks (week 6 and week 12) were incorporated in the programme to provide opportunities to further consolidate previously learned letters. In week six, eight sounds and their corresponding letters were revised, and in week 12, all 23 sounds and their letters were consolidated. The programme included four vowels, 16 consonants and three letter-combinations.

Letters were introduced in the font developed by the Ministry of Education of Luxembourg (i.e. Drockschrëft) on individual flashcards. Four main guidelines were considered when creating the sequence of introduction of the sounds and their letters (Beck & Beck, 2013; Carnine, Silbert, Kame'enui, & Tarver, 2004):

- frequent and more salient sounds were introduced first,
- only one sound per letter was introduced,
- auditory similar sounds or visually similar letters were separated by the introduction of at least two other dissimilar sounds or letters,
- sounds represented by two or three letter graphemes (three in total) are introduced last to avoid confusion.

The lower and uppercase versions of the letters were introduced simultaneously as this parallels the standard teaching approach in Luxembourg. All sounds were introduced with an accompanying movement and a flashcard. Children were told the two-three letter graphemes form a "team" of letters that make a special sound together. An overview of the introduced letters in the order of introduction is presented in table 3.9.

Table 3.9: *Overview of the 23 sounds with the corresponding letter(s) and the mnemonic word, in the order of introduction*

sound	letter	Mnemonic word Luxembourgish	German translation	English translation
[l]	/l/	Libell	Libelle	dragonfly
[aː]	/a/	Adler	Adler	eagle
[m]	/m/	Maus	Maus	mouse
[iː]	/i/	Igel	Igel	hedgehog
[t]	/t/	Tëntefësch	Tintenfisch	squid
[z]	/s/	See	Säge	saw
[oː]	/o/	Ozean	Ozean	ocean
[b]	/b/	Bier	Bär	bear
[ʀ]	/r/	Riserad	Riesenrad	ferris wheel
[eː]	/e/	Efeu	Efeu	ivy
[g]	/g/	Gorilla	Gorilla	gorilla
[f]	/f/	Fanger	Finger	finger
[uː]	/u/	Ufo	Ufo	ufo
[h]	/h/	Hex	Hexe	witch
[n]	/n/	Nilpäerd	Nilpferd	hippo
[k]	/k/	Kinnek	König	king
[d]	/d/	Deckel	Deckel	lid
[v]	/w/	Wäschmaschinn	Waschmaschine	washing machine
[ʦ]	/z/	Zauberer	Zauberer	wizard
[p]	/p/	Pirat	Pirat	pirate
	Di- & trigraphs			
[ɑʊ]	/au/	Auto	Auto	car
[ɑɪ]	/ei/	Eidechs	Eidechse	lizard
[ʃ]	/sch/	Schéier	Schere	scissors

Each sound and corresponding letter was associated to a mnemonic word and image that begins with the target sound. For instance, the sound and letter /m/ was associated with the mnemonic word "Maus" [mouse]. The mnemonic words were chosen to be similar in Luxembourgish and German (words that start with the same sound and letter in Luxembourgish as well as in German) and that could be depicted in a drawing. In addition, each sound had an accompanying movement. For example, the mnemonic word for the sound /l/ was “Libell” [dragonfly] and the specific movement for this sound was to spread the arms and pretend to fly.

3.9.4 Promoting print and book awareness

To raise print and book awareness and to engage in a meaning oriented literacy activity, regular book reading activities were incorporate during the intervention (week 3-12). Developmentally appropriate Luxembourgish short stories were designed for each letter-sound. Each story incorporate frequent words with the corresponding letter-sound to further consolidate sound-letter linkage. Teachers were provided a single big book (A1 size) containing the 23 stories. During the reading activity, teachers laid out the big book in front of the children and read the text out loud. Teachers pointed to the print as its read, drawing children’s attention to the written words in the text. The reading of the story was followed by a brief discussion of the storyline. The story activity always concluded with the same task: Each child had to look for the corresponding letter amongst the words of the story text and cover it with a small token. Once all children had placed their token, the teachers picked up the tokens one by one and read out lout the word, followed by brief discussion of the position of the targeted sound in the word (i.e. initial, middle or final position).

3.9.5 Encouraging parent involvement

To encourage caregiver involvement after the end of the intervention, intervention materials and resources from the LALA programme were made available to caregivers at the end of Year 2 of preschool. Each child received the story book (A4-size), the songbook (A4-size), a CD with the songs and 23 flashcards of the 23 trained letter-sounds (A6-size). In addition, a parent guide was developed (in Luxembourgish, French, and Portuguese) with easily understandable practices and specific strategies that can be implemented from home to support children’s literacy development, irrespective of the home language.

3.9.6 Delivery of intervention

The intervention was delivered by the classroom teacher to all their Year 2 preschool children. Teachers were given a detailed manual containing the objectives of each of the 48 intervention session, an overview of the material needed and a detailed description with suggestive instructions for each activity. All the material was provided to the teachers. Teachers were trained by the research team (sees section 4.1.2) and solicited to adhere to the manual as closely as possible, while at the same time adapt the level of difficulty of the activities and the scope of the scaffolding to the needs of the children.

An overview of the intervention structure is presented in the appendices (see appendix B).

3.10 Measures on fidelity of implementation of intervention

To monitor the degree to which the intervention was administered as intended, various fidelity measures were implemented:

- teachers received specific training in the administration of the intervention prior to start of the programme to ensure clarity of the aims and teaching methods,
- each teacher was observed four times by the author delivering an intervention unit. A checklist was devised and completed by the author assessing the quality of implementation and engagement of the children,
- teachers completed written self-reports for each session, including attendance rates of the children,
- members of the research team (PEdA and CW) held monthly school tutorials with teachers (three in total across the intervention),
- a focus group was held with all teachers at the end of the intervention to gather qualitative feedback on the teachers' views on the intervention,
- teachers were handed out an anonymous questionnaire at the end of the intervention allowing to express their views on the intervention,
- an anonymous caregiver's questionnaire was handed out to the caregivers four months after the intervention (see appendix C), seeking information about the home use of the material (i.e. parent kit, see section 3.9.5) and children's responsiveness to the intervention,
- children were individually asked whether they enjoyed the programme at the immediate post-test.

CHAPTER IV - Results and Discussion: Intervention study

This chapter addresses the research questions related to the efficacy of the early literacy intervention for the entire sample. The chapter is split into four sections: (1) fidelity of intervention implementation; (2) background variables and participants' performance at pre-test; (3) effects of the intervention for the primary and secondary outcome measures with pre- to post-test, and pre- to delayed follow-up test comparisons, (4) discussion of the findings in relation to the research questions.

The specific research questions addressed in this chapter are as follows:

1) Can a structured intervention targeting early literacy skills in preschool improve children's early literacy skills in Luxembourgish in comparison to an untrained control group in children from Luxembourg, many of whose first language is not Luxembourgish?
2) Will any effects of the early literacy intervention in Luxembourgish be maintained nine months after the intervention has finished in Grade 1?
3) Will the early literacy intervention in Luxembourgish have any intervention effects on children's PA skills in German in Grade 1?
4) Will the early literacy intervention in Luxembourgish have any intervention effects on children's responsiveness to literacy instruction in German in Grade 1?

Data from 89 children, who received a twelve week classroom-based early literacy intervention in Luxembourgish in preschool, are longitudinally compared to data from 100 children in the control group who followed the standard national preschool curriculum. All children (N = 189) were assessed over three time points: prior to intervention in preschool (pre-test, $t1$), immediately after the intervention in preschool (post-test, $t2$) and nine months after the end of the intervention in Grade 1 of primary (delayed follow-up test, $t3$).

4.1 Fidelity of intervention implementation

In intervention research, fidelity is understood as both adherence to the intervention model and quality of implementation. Traditionally, fidelity is described across five dimensions: adherence to the programme, dosage of intervention, quality of intervention delivery, participant responsiveness to the intervention and programme differentiation (Dusenbury, Brannigan, Falco, & Hansen, 2003; Furtak et al., 2008; Wolery, 2011). Measures assessing fidelity in the current study were described in section 3.10.

4.1.1 Adherence to the intervention

To ensure that the intervention was administered as intended, teachers were provided with a detailed intervention manual and all the intervention material needed to administer the intervention. The manual included a comprehensive session-by-session guide, explicitly listing the aims and the materials needed for each activity of each session and describing each activity with precise instruction examples. In addition, three tutorials (one per month) were held in each school to offer teachers the opportunity to clarify any questions the might have concerning the intervention activities in the coming weeks. All teachers delivering the intervention attended all three tutorials.

Adherence to the programme was also ensured by four unnotified on-site observations per teacher by the researcher (CW) spread over the duration of the intervention. Each on-site visit included the researcher observing the teacher delivering an intervention session and the completing of a session observation sheet by the researcher. The adherence to the manual was rated on a five-point Likert scale with the manual instructions as a reference point (1 = several aspects missing/not satisfactory, 2 = some aspects missing/not satisfactory, 3 = according to manual, 4 = according to manual with good use of resources/questions/techniques to support early literacy skills, 5 = according to manual with very good use of resources/ questions/ techniques). On average, teachers achieved an implementation quality rating of 3.12 (*SD* = 0.47; range = 2.5 - 3.8).

4.1.2 Quality of intervention implementation

To ensure quality of implementation, all 14 teachers who administered the intervention were trained by the research team (CW and PEdA) during one day (8h) approximately two months prior to the start of the intervention.
This training was followed-up with a small-group refresher training session of two hours within each school in the week prior to the start of the intervention.
Training centred around describing the programme, its rationale, intervention procedure and activities, the importance of using rich language, how to effectively use puppetry as a teaching tool, and strategies on how to support L2 learners. The training also included a description of the background of the study, the programme- and session structure, the rationale for intervention procedures and the type of activities. It also incorporated an introduction into key instructional approaches used in the intervention, such as the importance of using rich language, the effective use of the hand-puppet as a teaching tool and strategies on how to support L2 learners.

Teachers were offered to join a WhatsApp for their respective school with the researcher for immediate troubleshooting. All teachers agreed to join the WhatsApp. All books and flashcards were produced in collaboration with local artists and a graphic designer and professionally edited and printed. This was done to ensure that the material is appealing to both teachers and children.

4.1.3 Dosage of the intervention

Preschool in Luxembourg is compulsory and the intervention was administered on a classroom-basis. During the 12 intervention weeks the early literacy intervention was delivered as part of the standard classroom activities by the class teacher. All 14 teachers administered all 48 intervention sessions. Children only missed an intervention session if they did not attend school (e.g. due to sick leave). No individual catch-up sessions were held if a child or more children had missed a session or multiple sessions. Teachers completed a self-report for each of the 48 intervention sessions, including the date of the session and attendance rates of children. On average children attended 46.24 out of the 48 (96%) intervention sessions (*SD* = 2.53, range= 36-48).

4.1.4 Participant responsiveness to the intervention

To monitor the children's responsiveness to the intervention, their reactions and engagement was rated by the researcher during the onsite observations on a five point Likert scale (1 = poor responsiveness, 2 = below expected responsiveness, 3 = expected responsiveness, 4 = above expected responsiveness and 5 = extraordinary responsiveness). On average, children achieved an responsiveness to the intervention rating of 3.01 (*SD* = 0.47, range = 1.75 - 3.75). To further evaluate participant responsiveness, children were asked, as part of the post-testing, to indicate their satisfaction with the programme on a rating scale from one to three (1 = sad smiley, 2 = neutral smiley and 3 = happy smiley). Out of the 85 children at post-test, 81 children (95%) gave the intervention a happy smiley, three children (4%) responded with a neutral smiley and one child (1%) gave it a sad smiley.

Data on teacher's responsiveness to the intervention was also monitored via an anonymous questionnaire handed out to the teachers at the end of the intervention and was returned to the research team in a sealed envelope. The questionnaire contained two questions in relation to satisfaction with the programme. Firstly, a dichotomous question (1 = yes, 2 = no) on whether the teachers enjoyed working with the programme. All the teachers (100%) indicated that they enjoyed working with the intervention programme. The second question was a 4 point rating scale on whether they liked the intervention programme (1 = did not like it, 2 = like it a little bit, 3 = liked it, 4 = liked it very much). Hundred percent of the teachers indicated that they liked the intervention programme very much. Eleven out of the 14 intervention teachers participated in the final focus group immediately post-intervention. Qualitative data generated from the focus group using thematic analysis (Charmaz, 2006) indicated that all eleven teachers felt that all their children enjoyed working with the programme and that they had benefitted from the intervention. Teachers reported to have noticed that children who tend to struggle with the language of instruction also managed to follow the programme well and made substantial progress.

Feedback on the intervention programme from the caregivers was gathered via an anonymous questionnaire handed out at the delayed-follow up to all children from the intervention group (*n* = 72). Teachers gave the questionnaires to the children to take home with them. Caregivers returned the questionnaires in a sealed envelope via their child to classroom teacher and the researcher (CW) picked up the sealed envelopes from the teachers. Out of the 72 questionnaires, 57 (79%) were returned which can be considered a very good response rate (Mangione, 1995; Morton, Bandara, Robinson, & Atatoa Carr, 2012). Caregivers were asked the dichotomous questions (1 = yes, 2 = no) whether they thought their children enjoyed receiving the intervention and whether they thought the intervention had a learning effect. Out of the 57 returned questionnaires, 52 caregivers (91%) indicated that the children enjoyed receiving the programme and 54 caregivers (95%) indicated that the intervention had a learning effect.

4.1.5 Programme differentiation

Programme differentiation consists of the degree to which the main components of the early literacy intervention are distinguishable from the standard preschool school curriculum in Luxembourg. The content of the intervention programme was compared to the national curriculum for public preschools that is uniform across public schools in Luxembourg. Study content and levels of competences to be acquired are fixed in the Grand-ducal Regulation of 11 August, 2011 (MENFP, 2011). With reference to the acquisition of early literacy skills, at the end of Year 2 of preschool, children are expected to be able to: identify rhymes and initial sounds and segment words; differentiate different written signs; handle a book, discover the social use of writing; discover their first name among other names; recognize well-known pictograms; follow the course of events in an easy text that is read to them (MENFP, 2011). It further states that preschool teaching is foremost a social experience and that it should focus on a global, holistic approach through immersing children in stimulating contexts, not on the explicit teaching of skills (MENFP, 2011). Thus, the pedagogical approach to teachings of the national curriculum is fundamentally different to the highly structured, explicit and systematic training of early literacy skills of the intervention evaluated in the current study.

4.2 Preliminary data considerations

This section gives an overview of the procedures that were used to screen the data prior to analysis and addresses floor and ceiling effects. All data analyses were conducted using IBM SPSS Statistics Base 24 (IBM Corp, 2016) and IBM SPSS Amos 25 (Arbuckle, 2017). Throughout the thesis, the juxtaposed results of children from the intervention and the control group were colour-coded according to the following convention:

green for children in the intervention group

orange for children in the control group

4.3 Data screening

The data were screened via visual inspection of histograms with normality curves and frequency tables to identify data-entry errors, missing data or implausible values, distributional characteristics and outliers. Normality of the data was verified using measures of standardized skew. Although the parametric methods used in the study (ANOVAS, ANCOVAS) are fairly robust against violations to the normality assumption for sample sizes over 45 participants (Blanca, Alarcón, Arnau, Bono, & Bendayan, 2017; Ziegler, Beyer, Schmider, Danay, & Bühner, 2010), severe departures from normality and extreme outliers can be problematic as they could lead to inaccurate estimates of parameters (Tabachnick & Fidell, 2012). It has been suggested that for a sample size between 50 and 300 participants, standardized skew or kurtosis values below 3.29 are acceptable (Kim, 2013). Individual variables with standardized skew values above 3.29 were therefore transformed prior to the inferential statistics using a square root or a log transformation for normality (Tabachnick & Fidell, 2012). Square root transformations were used for LSK at post-test, the phonological awareness measures in German at delayed follow-up and the SLRT reading tasks at delayed follow-up. The number naming task at post-test and the spelling task at delayed follow-up were log-transformed. Measures that showed a ceiling or floor effect were not transformed because non-normal data due to such effects cannot be meaningfully transformed. In the current thesis, a task was considered at ceiling when at least 50% of the sample scored the two highest possible scores on a task.

Floor effects were defined as the reverse response pattern with at least 50% of the sample scoring the two lowest possible scores on a task. Inferential statistics on measures with a floor or ceiling effect need to be interpreted very cautiously and hence, are highlighted by a subscripted $_C$ (ceiling) or subscripted $_F$ (floor) throughout the thesis. No outliers were removed from the data set. An overview of the variables that were transformed and the type of transformation that was applied is presented in the appendices (see appendix D).

4.4 Data analysis procedures

In the evaluation of the efficacy of the intervention, intention to treat (ITT) principles were followed. This means that all children (N = 189) whose schools were allocated to the intervention and the control group were included in the analysis at all time points, irrespective of the number of intervention sessions children received. This approach is considered a rather conservative approach to the analysis of treatment effects because children with low compliance rates are included in the analysis of treatment effects (Gupta, 2011). However, as discussed above (see section 4.1.3), the attendance rates of the participants were very high.

In total 17 out of the 189 children (9%) were lost by the time of the delayed follow-up test: eight children (4%) from the intervention group and nine children (5%) from the control group. A few participants moved schools (n = 13), were retained in preschool (n = 2), started home-schooling (n = 1) or left the country (n = 1). The children who moved schools during the study could not be followed-up because no study consent had been obtained for the new school districts. Bennett (2001) argues that missing data greater than 10% may introduce a statistical bias, however, traditionally, the pattern of missing data is considered more important than the amount of missing data (Tabachnick & Fidell, 2012). Little's MCAR test showed that the missing data on the outcome measures can be considered as missing completely at random, $\chi^2(75) = 94.72$, $p = .062$ (Little, 1988). Due to the low attrition rate from pre-test to delayed follow-up (< 10%), a similar attrition rate for the intervention group (4%) and the control group (5%) and the data missing completely at random, no further steps were needed to account for missing data. To evaluate the efficacy of the intervention, separate analyses of variance /covariance (ANOVAS/ANCOVAs) were run on each outcome measure (or each outcome component when individual measures were combined, as will be described later, see section 4.6) at post-test and delayed follow-up, controlling for scores at pre-test. This analysis thus answers the question whether the post-test scores, adjusted for pre-test scores, differ between the intervention and control groups. For secondary outcome measures or components that did not include baseline scores at pre-test, analyses of variance (ANOVAs) on the outcome measures were performed. For the measures that departed from normality, the analyses were performed on the transformed data. With the exception of spelling at the delayed follow-up test, the homogeneity of variance assumption was not violated. For the spelling measure at delayed follow-up, a Welch's *F*-ratio correction (Welch, 1951) was applied as a robust method to control for the Type 1 error rate. The assumption of homogeneity of regression slopes was checked by initially adding an interaction term. With the exception of one outcome variable (i.e. LSK at post-test), the interaction term was not significant and subsequently dropped. As the assumption of homogeneity of regression slopes for the ANCOVA was violated for LSK at post-test, the interaction term between the group allocation and the covariate (i.e. LSK at pre-test) was included in the model.

Increasing awareness of limitations of focusing on significance levels has led to stronger emphasis on reporting effect sizes (Ferguson, 2009; Maxwell, Kelley, & Rausch, 2008; Wilkinson, 1999). The predominant criticisms of significance testing are high sensitivity to sample size and failure to provide information about the practical importance of statistical relationships. Effect sizes provide a better estimate of the magnitude of an effect between variables and they are more resistant to influences of sample size. Thus, effect sizes are reported with all inferential analyses. For all measures for which baseline data were available, Cohen's *d* effect sizes were computed based on differences in progress between the intervention and the control group from pre-test to post-test and from pre-test to follow-up test, divided by the initial pooled standard deviation. As measures with floor effects at pre-test imply an excessively small standard deviation artificially inflating Cohen's *d* values, effect sizes for measures with floor effects at pre-test were divided by the pooled standard deviation at post-test. If baseline data were not available, effect sizes were computed based on differences in means divided by the pooled standard deviation at post-test or follow-up test respectively. A positive Cohen's *d* value represents an intervention effect in favour of the intervention group, whereas a negative Cohen's *d* means that the control group outperformed the intervention group. Traditionally, effect size values of Cohen's *d* are considered small (0.20 - 0.49), medium (0.50 - 0.79) or large (≥ .80) (Cohen, 1988). Yet, more recently the range has been expanded by Sawilowsky (2009) by a very large (1.20 - 1.99) and a huge Cohen's *d* (≥ 2.00). It has been postulated that in education contexts, Cohen's *d* values of equal or larger than 0.25 can be considered of educational importance (Network Promising Practices, 2007; What Works Clearing House, 2007).

4.5 Examination of possible study confounds

In a quasi-experimental intervention design, an essential quality indicator is that known confounding factors are controlled prior to the start of the intervention (Innocenti et al., 2014). Schools were therefore matched on SES, number of students and Luxembourgish language learners (see section 3.1.1).

In the following sections, firstly, differences between the intervention and the control group on background and control variables that have shown to be related to early literacy achievement, i.e. gender, age, SES, language learner status and NVR were investigated. Secondly, due to the intimate link between vocabulary and early literacy skills, it was examined whether differences in vocabulary in Luxembourgish or in German between the intervention and the control group could be detected. Thirdly, despite best efforts to match control/intervention groups, baseline differences emerged on some of the pre-test outcomes measures and are presented.

4.5.1 Group differences on background and control variables

Information on important background and control variables (i.e. gender, age, SES, number of Luxembourgish language learners, NVR) between the intervention and control group are presented in table 4.1 and 4.2. Frequencies with percentages and Pearson's chi-squared significant levels for group differences are reported for gender and number of Luxembourgish language learners in table 4.1.

Table 4.1: *Number (%) of male students and Luxembourgish language learners in the intervention and in the control group at pre-test, with chi-square significance levels and Cramer's V effect sizes for group differences*

	Frequency (%)				
	Intervention group (*n* = 89)	Control group (*n* = 100)	χ^2	*p*	*V*
Gender (male)	49 (55)	56 (56)	0.02	.896	-0.01
Luxembourgish language learners	41 (46)	38 (38)	1.20	.262	0.08

No significant differences were observed for gender or the number of Luxembourgish language learners between the intervention and the control group. Means for age, SES and NVR for the two groups, with *t*-test significance values and effect sizes are reported in table 4.2.

Table 4.2: *Mean raw scores (SD) for age (t1, t2, t3), SES (t1) and NVR (t1) for children of the intervention and control group, with t-test significance levels and Cohen's d effect sizes for group differences*

	Mean (*SD*)				
	Intervention group (*n* = 89)	Control group (*n* = 100)	*T*	*p*	*d*
Age (months)					
*t*1	67.85 (3.60)	68.27 (3.99)	-0.75	.455	-0.11
*t*2	74.09 (3.44)	75.06 (3.90)	-1.79	.078	-0.26
*t*3	81.31 (4.24)	82.35 (3.99)	-1.66	.099	-0.25
SES (ISEI[1])	53.44 (23.02)	55.31 (21.91)	-0.55	.582	-0.08
NVR (matrices[2])	14.87 (3.63)	14.43 (4.65)	0.71	.478	0.11

Note. [1]ISEI, International Socio-Economic Index of occupational status; [2]Matrices WPPSI-III, Wechsler Preschool and Primary Scale of Intelligence.

No significant differences between the intervention and the control group could be observed for any of the background variables.

4.5.2 Group differences on vocabulary knowledge

Mean raw scores and *SD*s for the vocabulary measures in Luxembourgish and in German are reported in table 4.3, with significance levels for group differences and effect sizes for each time point.

Table 4.3: *Mean raw scores (SD) for vocabulary measures in Luxembourgish (t1, t2, t3) and German (t3) for children of the intervention and control group, with t-test significance levels and Cohen's d effect sizes for group differences*

	Mean (*SD*)				
	Intervention group (*n* = 89)	Control group (*n* = 100)	*T*	*p*	*d*
Vocabulary Lux. CLT (40)					
*t*1	33.10 (4.49)	33.94 (4.48)	-1.28	.201	-0.19
Vocabulary Lux. PPVT (40)					
*t*1	29.15 (6.44)	29.15 (6.40)	-0.00	.997	0.00
*t*2	31.56 (5.04)	31.22 (5.33)	0.45	.654	0.07
*t*3	33.20 (4.27)	33.89 (4.23)	-1.05	.288	-0.19
Vocabulary Ger. PPVT (228)					
*t*3	94.91 (27.33)	104.99 (29.20)	-2.34	.021	-0.36

Note. (), maximum raw scores; CLT, Cross-Linguistic Lexical Task; PPVT, Peabody Picture Vocabulary Test. Cohen's *d*: differences in means divided by pooled *SD*.

The raw mean scores on the vocabulary CLT task in Luxembourgish of the two groups were very similar at pre-test and did not differ significantly. The mean raw scores on the PPVT in Luxembourgish between the two groups were also highly similar across all three times points. No significant group differences could be observed at any time point. The German vocabulary measure was only administered in Grade 1 as the children had not been introduced to German in preschool. The mean raw scores on the PPVT in German showed that the control group outperformed the intervention group on average by 10 items, which was a statistically significantly difference with a small effect size.

4.5.3 Group differences on outcome measures at pre-test

Mean raw scores with *SD*s for outcome measures at pre-test for the intervention and the control group are reported in table 4.4. An independent *t*-test was conducted between the groups to test for statistical significance of differences. The syllable segmentation task presented a ceiling effect and a floor effect was observed for the phoneme blending, onset manipulation and basic reading tasks. Thus, results on these tasks need to be interpreted cautiously.

Table 4.4: *Mean raw scores (SD) for early literacy measures at pre-test for children of the intervention and control group, with significance levels and effect sizes for group differences*

Pre-test outcome measures	Mean (*SD*) Intervention group (*n* = 89)	 Control group (*n* = 100)	*T*	*p*	*d*
PA Luxembourgish					
Syllable segmentation (12)	9.92C (1.98)	9.54C (2,76)	1.08	.282	0.16
Rhyme detection (12)	5.72 (4.11)	7.40 (3.91)	-2.88	.004	-0.42
Onset–rhyme blending (12)	3.70 (3.73)	5.00 (4.10)	-2.27	.024	-0.33
Onset identification (12)	4.26 (3.07)	5.30 (4.25)	-2.29	.025	-0.33
Phoneme blending (12)	0.65F (1.50)	1.86F (3.25)	-2.26	.002	-0.48
Onset manipulation (12)	0.36F (1.50)	0.50F (1.85)	-3.22	.570	-0.48
Letter-sound knowledge (20)	5.84 (5.76)	7.85 (6.78)	-2.18	.031	-0.32
Basic reading					
Basic word reading (12)	0.19F (1.30)	0.57F (2.14)	-1.45	.148	-0.21
Basic non-word reading (12)	0.27F (1.38)	0.74F (2.33)	-1.66	.089	-0.25

Note. (), maximum raw scores; Cohen's *d*: differences in means divided by pooled *SD*.
C = ceiling effect; F = floor effect.

With regard to the PA measures, scores on the syllable segmentation task showed a ceiling effect for the control group, limiting the conclusions that can be drawn from the *t*-test. However, for the remaining PA tasks, results indicated consistent significant (*p*s = .002 to .025) pre-test differences in favour of the control group, apart from onset manipulation (*p* = .570). Effect sizes of the difference between the two groups were small (*d*s = -0.33 to -0.48).

Performance on pre-test measures showed that at children of the control group knew on average two more letters than the children of the intervention group. This difference was statistically significant with a small effect size. As expected given that children had not yet been formally introduced to reading by the time of the pre-test in preschool, the basic reading measures also showed clear floor effects at pre-test. Group differences were nonsignificant for basic word reading and basic non-word reading. Yet the results on the basic reading measures need again be interpreted cautiously due to the mentioned floor effects.

4.6 Data preparation

Principal components analyses (PCA) were conducted when multiple measures tapped the same underlying ability. The computation of theoretical motivated component procedure aimed to deflate the type 1 error risk by reducing the number of performed analyses, to enhance reliability of the measures, and to ease the reporting of results (Tabachnick & Fidell, 2012).

As mentioned in the literature review of Study 2 (see section 5.2), PA is conceptualised as operating on a larger linguistic level (syllables, rhyme and onset-rhyme) and on a smaller linguistic level (phonemes). Particularly, the small-unit PA measures (i.e. on the level of the phoneme) are important in the development of learning to read and write (Caravolas et al., 2005; Muter et al., 2004). While, larger-unit PA skills were also included in the first weeks of the intervention, 48 intervention sessions, 39 sessions (81%) specifically target the phonemic level. Thus, it was important to evaluated intervention effects for separable large-unit and small-unit aspects of PA. Given these theoretical points, the decision was taken to create the following PA components:

- a *large linguistic unit PA component* in Luxembourgish (*t*1-*t*2) including the following individual measures: rhyme identification, onset-rhyme blending and onset identification. The syllable segmentation task already showed a large ceiling effect at pre-test and only correlated weakly with the other large-unit PA measures at pre-test, *r*'s < .25, and at post-test, *r*'s < .10. Thus, syllable segmentation was not included into the large-unit PA component;

- a *small linguistic unit PA component* in Luxembourgish (*t*1-*t*2-*t*3) including the following individual measures: phoneme blending, onset manipulation and phoneme segmentation;
- a *small linguistic unit PA component* in German (*t*3) including the following individual measures: phoneme blending, onset manipulation, phoneme segmentation.

In addition, to the PA measures, individual measures of early reading skills were also combined using separate PCAs. The following reading components were created:

- a *basic reading component* including the basic word reading and basic non-word reading measures (*t*1-*t*2-*t*3)
- an SLRT reading component in German (*t*3) including the SLRT word reading and SLRT non-word reading measures

In all instances, the PCA returned a one-factor solution with an Eigenvalue greater than one. Component loadings were all above .60 and explained variance ranged from 63%– 97%. An overview of the factorability statistics for each component at each time point are presented in the appendices (see appendix E).

Normality of the components was checked using the same criteria as for the individual measures (i.e. standard skew below 3.29, sees section 4.3). The small-unit PA component and the SLRT reading component showed a standardized skew greater than 3.29. To reduce skewness, square root data transformations were applied to all the individual variables prior to performing the PCA. The two components based on the square rood transformed data presented a standardized skew below the 3.29 threshold (see appendix D).

4.7 Efficacy of the early literacy intervention for the entire sample

The examination of the efficacy of the early literacy intervention followed the structure of the research questions presented above. Firstly, with reference to research questions 1 and 2, results addressing the overall efficacy of the intervention on primary outcomes immediately post-intervention and at a nine-months delayed follow-up are presented. Secondly, with reference to research question 3 and 4, results examined intervention effects on secondary outcome measures.

4.7.1 Effects on primary outcome measures

Descriptive and inferential statistics for the three primary outcome measures, i.e. the small- and large-unit PA components in Luxembourgish and LSK, immediately post intervention and at delayed follow-up are reported in table 4.5.

Table 4.5: *Mean raw scores (SD), with ANCOVA significance values and Cohen's d effect sizes for intervention effects at pre-test (t1), at post-test (2) and at delayed follow-up test (t3) for primary outcome measures*

	Mean (*SD*)			
	Intervention group	Control group	*d**	ANCOVA results
Syllable segmentation Lux. (12)				
*t*1	9.92 (1.98)	9.54_C (2.76)		
*t*2	9.20 (2.87)	10.12_C (2.19)	-0.54[1]	*F*(1,182) = 8.63, *p* = .004, *d* = -0.44
Rhyme Identification Lux. (12)				
*t*1	5.72 (4.11)	7.40 (3.91)		
*t*2	8.35 (3.46)	8.98 (3.29)	0.26[1]	
Onset-rhyme blending Lux.(12)				
*t*1	3.70 (3.73)	5.00 (4.10)		
*t*2	9.19 (3.48)	6.69 (3.98)	0.97[1]	
Onset identification Lux.(12)				
*t*1	4.26 (3.07)	5.30 (4.25)		
*t*2	7.45 (3.52)	5.99 (3.34)	0.79[1]	
PA large-unit component Lux.				
*t*2 – *t*1				*F*(1,182) = 53.40, *p* < .001, *d* = 1.09
Phoneme blending Lux.(12)				
*t*1	0.65_F (1.50)	1.86_F (3.25)		
*t*2	4.64 (3.58)	2.77_F (3.29)	0.90[2]	
*t*3	8.96 (2.71)	7.89 (3.24)	0.76[2]	
Onset manipulation Lux. (12)				
*t*1	0.36_F (1.50)	0.50_F (1.85)		
*t*2	2.06_F (3.34)	1.30_F (2.85)	0.29[2]	
*t*3	5.25 (4.26)	4.86 (4.11)	0.13[2]	
Phoneme segmentation Lux. (12)				
*t*2	3.72 (3.79)	1.65_F (2.71)	0.63[3]	
*t*3	8.33 (3.39)	7.11 (3.67)	0.35[3]	
PA small-unit component Lux.				
*t*2 – *t*1				*F*(1,182) = 36.62, *p* < .001, *d* = 0.90
*t*3 – *t*1				*F*(1,169) = 9.58, *p* = .002, *d* = 0.48.
Letter-sound knowledge (20)				
*t*1	5.84 (5.76)	7.85 (6.78)		
*t*2	15.53 (5.47)	9.90 (6.59)	1.21[1]	*F*(1,182) = 145.86, *p* < .001, *d* = 1.78
*t*3	19.67_C (0.91)	19.17_C (1.85)	0.40[1]	*F*(1,169) = 9.96, *p* = .002, *d* = 0.49

Note. (), Maximum raw scores.

*Cohen's *d*: [1]difference in progress between groups divided by pooled initial *SD*, [2]difference in progress divided by pooled *SD* at post-test / follow-up (pre-test scores were at floor), [3]difference in means divided by pooled *SD* at post-test / follow-up (pre-test scores were not available).

$_C$ = ceiling effect; $_F$ = floor effect.

Large-unit PA Luxembourgish

As mentioned above, syllable segmentation was not included in the large-unit PA component and is reported individually. The syllable segmentation task was the only PA measure that showed a ceiling effect at pre-test. The ANCOVA indicated that children from the control group significantly outperformed children from the intervention group at post-test, $F(1,182) = 8.63$, $p = .004$, $d = -0.44$. However, this finding needed to be interpreted cautiously due to the ceiling effect.

With regards to the large-unit PA measures, the individual Cohen's *d*s on differences in progress showed that the children from the intervention group had made larger progress from pre-test to post-test than children from the control group (individual Cohen's *d*s ranging from 0.26 - 0.97). The ANCOVA on the large-unit PA component showed that the children from the intervention group significantly outperformed the children from the control group at immediate post-test with a large effect size, $F(1,182) = 53.40$, $p < .001$, $d = 1.09$.

Small-unit PA Luxembourgish

The two groups showed floor effects on the small-unit PA measures at pre-test. However, while children from the control group still presented floor effects at post-test, children from the intervention group were able to complete the phoneme blending and phoneme segmentation tasks. Only scores on the onset manipulation task remained at floor for the children from the intervention group at post-test. The individual Cohen's *d*s on the differences in progress from pre-to post-test ranged from 0.29 to 0.90 for the individual small-unit PA measures. The largest intervention effect could be observed for the phoneme blending tasks ($d = 0.90$). The ANCOVA on the small-unit PA Luxembourgish component indicated that the children from the intervention group significantly outperformed children from the control group with a large effect size, $F(1,182) = 36.62$, $p < .001$, $d = 0.90$. However, the parameter estimates need to be interpreted with caution due to the floor effects from the control group on all small-unit PA measures at post-test.

At delayed follow-up in Grade 1, the mean raw scores of the individual small-unit PA measures at delayed follow-up indicated that children had made substantial progress in small-unit phoneme abilities from post-test to delayed follow-up. The floor effects of the control group at post-test had disappeared by the time of the delayed follow-up. The onset manipulation task showed the lowest mean and, arguably, required the cognitively most demanding phonological manipulations out of the three small-unit PA tasks. The individual Cohen's *d* values for differences in progress from pre-test to follow-up ranged from 0.13 to 0.76. The largest retained effect could be observed for the blending tasks ($d = 0.76$). The ANCOVA on the small-unit PA component at delayed follow-up indicated a significant effect in favour of the intervention group, $F(1,169) = 9.58$, $p = .002$, $d = 0.48$. The two ANCOVAs performed at post-test and delayed follow-up on the small-unit PA components remained significant with Bonferroni corrected p-values ($p = .050/2= .025$) to account for multiple comparisons.

Letter-sound knowledge

Descriptive results for LSK at pre-test indicated that the children from the intervention group ($M = 5.84$, $SD = 5.76$) and children from the control group ($M = 7.85$, $SD = 6.78$) knew a few letters already. Children in the control group knew on average two more letters at pre-test than children in the intervention group and this difference was significant (see section 4.5.3). The mean raw scores indicated that immediately post-intervention, the children from the intervention group knew on average approximately six letters more than the children from the control group. The Cohen's *d* on differences in progress indicated very large difference in progress in favour of the intervention group ($d = 1.21$). The ANCOVA on LSK indicated a significant intervention effect, $F(1,182) = 145.86$, $p < .001$, $d = 1.78$[6]. By the time of the delayed follow-up test in Grade 1, children had been introduced to formal teaching of letters for over five months and as expected LSK shows clear ceiling effects in both groups. Thus, the Cohen's *d* for difference in progress of 0.40 and the significant ANCOVA need to be interpreted very cautiously, $F(1,169) = 9.96$, $p = .002$, $d = 0.49$.

[6] The assumption of homogeneity of regression slopes for the ANCOVA was violated for LSK at post-test, the interaction term between the group allocation and the covariate (i.e. LSK at pre-test) was included in the model.

4.7.2 Effects on phonological awareness in German in Grade 1

With reference to the third research question, it was analysed whether the intervention showed any transfer effects to PA in German in Grade1. Raw means scores, tests of differences and effect sizes for the secondary outcome measures were reported in table 4.6. As lexical knowledge can effect performance on PA tasks in German, tests of differences were computed twice: once without (ANOVA) and once with control (ANCOVA) for vocabulary in German.

Table 4.6: *Mean raw scores (SD), with ANCOVA / ANOVA significance values and Cohen's d effect sizes for intervention effects at delayed follow-up test (t3) for PA in German*

	Mean (*SD*)			
	Intervention group	Control group	*d**	ANCOVA / ANOVA results
Phoneme blending Ger. TPB (12)				
*t*3	9.83 (2.44)	9.12 (3.09)	0.25	
Onset manipulation Ger. TBP				
*t*3	6.20 (3.84)	5.98 (4.26)	0.05	
Phoneme segmentation Ger. (12)				
*t*3	8.99 (2.73)	7.65 (3.12)	0.46	
PA small-unit component Ger. (12)				
*t*3				$F(1,170) = 3.95$, $p = .048$, $d = 0.31$[3]
Controlling for vocabulary in German at *t*3				$F(1,169) = 10.96$, $p = .001$, $d = 0.51$[4]

Note. (), maximum raw scores.
TBP, Test für Phonologische Bewusstheitsfähigkeiten [Test for phonological awareness skills].
*Cohen's *d*: difference in means between groups divided by pooled *SD.*
[3] ANOVA, [4] ANCOVA.

PA German

The raw mean scores indicated that the children of the two groups were able to perform PA operations on phoneme level in German after having been introduced to German for only roughly five months. Similar to the small-unit PA tasks in Luxembourgish, children show higher performances on the phoneme blending and phoneme segmentation tasks than on the onset manipulation task. The individual Cohen's *d*s on differences in means ranged from 0.05 for onset manipulation to 0.46 for phoneme segmenting. The ANOVA showed that the children of the intervention group significantly outperformed the children of the control group on the PA component in German with a small effect size, $F(1,170) = 3.95$, $p = .048$, *d*

= 0.31. Controlling for vocabulary in German, the ANCOVA revealed a significant difference with a moderate effect size, $F(1,169) = 10.96$, $p = .001$, $d = 0.51$.

4.7.3 Effects on literacy measures

With reference to the fourth research question, it was analysed whether the intervention showed any facilitative effects to measures of reading and spelling. Descriptive statistics, tests of differences and effect sizes for the secondary outcome measures were reported in table 4.7.

Table 4.7: *Mean raw scores (SD), with ANCOVA / ANOVA significance values and effect sizes for intervention effects at pre-test (t1), at post-test (2) and at delayed follow-up test (t3) for literacy measures*

	Mean (*SD*)			
	Intervention group	Control group	*d**	ANCOVA / ANOVA results
Basic real words (12)				
t1	0.19_F (1.30)	0.57_F (2.14)		
t2	2.22_F (3.50)	1.03_F (2.59)	0.51[1]	
t3	11.06_C (1.77)	10.79_C (2.14)	0.33[1]	
Basic non-words (12)				
t1	0.27_F (1.38)	0.74_F (2.33)		
t2	3.39 (3.94)	1.31_F (2.97)	0.73[1]	
t3	10.14_C (2.05)	10.01_C(2.45)	0.26[1]	
Basic reading component (12)				
t2 – t1				$F(1,182) = 28.58$, $p < .001$, $d = 0.80$[3]
t3 – t1				$F(1,169) = 0.85$, $p = .359$, $d = 0.14$[3]
Reading real words SLRT-II (72)				
t3	8.62 (9.39)	9.47 (9.73)	-0.09[2]	
Reading non-words SLRT-II (72)				
t3	13.75 (8.05)	14.97 (7.98)	-0.15[2]	
SLRT reading component				
t3				$F(1,170) = 0.69$, $p = .406$, $d = -0.13$[4]
Controlling for vocabulary in German at t3				$F(1,169) = 0.02$, $p = .890$, $d = 0.00$[3]
Reading comprehension ELFE 1-6 (72)				
t3	12.73 (5.02)	11.86 (5.61)	0.16[2]	$F(1,170) = 1.29$, $p = .258$, $d = 0.18$[4]
Controlling for vocabulary in German at t3				$F(1,169) = 4.01$, $p = .047$, $d = 0.31$[3]
Spelling HSP-graphemes (40)				
t3	34.25 (3.24)	32.37 (7.50)	0.32[2]	$F(1,170) = 2.39$, $p = .124$, $d = 0.24$[5]
Controlling for vocabulary in German				$F(1,169) = 6.65$, $p = .011$, $d = 0.40$[3]

Note. (), maximum raw scores.

SLRT II, Salzburger Lese- und Rechtschreibtests [Salzburg reading and orthography Test]; ELFE 1– 6, Ein Leseverständnistest für Erst-bis Sechstklässler [ELFE 1– 6: A reading comprehension test for students in Grades 1 through; HSP, Hamburger Schreib-Probe [Hamburg spelling-task].

*Cohen's *d*: [1]difference in progress between groups divided by pooled *SD* at post-test / follow-up (pre-test scores were at floor); [2]difference in means divided by pooled *SD* at follow-up.

[3] ANCOVA, [4] ANOVA, [5] Welch's ANOVA

$_C$ = ceiling effect; $_F$ = floor effect.

Basic reading

The basic word reading task and the basic non-word reading task presented floor effects at pre-test, indicating that the children in the two groups were not able to read at the beginning of second year of preschool. Immediately post-intervention, the control group still showed floor effects on the two basic reading tasks. However, only the basic word reading measure showed a floor effect for the intervention group. The children from the intervention group were able to complete the non-word reading task and read more than 3 out of the 12 non-words on average. The individual Cohen's *d* on differences in progress ranged from 0.51 for word reading to 0.73 for non-word reading. The ANCOVA showed that the children from the intervention group significantly outperformed the children from the control group, $F(1,182) = 28.58$, $p < .001$, $d = 0.80$. However, due to the floor effects listed above, the p-value and the effect sizes need to be interpreted very cautiously.

At delayed follow-up, mean raw scores on the reading basic words and the reading basic non-words showed a ceiling effect. This indicates that the children rapidly acquire basic word reading skills within approximately five months of formal literacy instruction. The individual Cohen's *d*s on differences in progress ranged from 0.26 for word reading to 0.33 for non-word reading, but the ANCOVA did not show a difference in performances between the two groups at delayed follow-up, $F(1,169) = 0.85$, $p = .359$, $d = 0.14$. However, the parameter estimates at delayed follow-up need to be interpreted cautiously due to the ceiling effects.

The ANCOVA performed at post-test on the basic reading component remained significant with Bonferroni corrected *p*-values ($p = .050/2= .025$) to account for multiple comparisons.

Word- and non-word reading SLRT

Mean raw scores on the German SLRT reading tasks showed that the children from the two groups were able to read German words and German non-words after five months of formal literacy instruction. The means also revealed that on both the word and the non-word reading task of the SLRT, the children from the control group seemed to be able to read on average one more word then the children from the intervention group. The Cohen's *d*s on the individual differences in means ranged from -0.09 for the non-word reading SLRT task to -0.15 for the word reading SLRT task. The ANOVA on the SLRT reading component indicated nonsignificant differences between the two groups, $F(1,170) = 0.69$, $p = .406$, $d = -0.13$. As lexical knowledge can effect performance on reading, an ANCOVA on the

SLRT reading component controlling for German vocabulary was conducted. The results indicated no significant differences between the two groups, $F(1,169) = 0.02$, $p = .890$, $d = 0.00$.

Reading comprehension ELFE

On the German reading comprehension measure ELFE 1-6, the data showed a higher mean raw score for the children from the intervention group than for children from the control group. The Cohen's *d* on differences in means suggest an effect size of 0.16 in favour of the intervention group. The between group ANOVA was nonsignificant, $F(1,170) = 1.29$, $p = .258$, $d = 0.18$. However, when controlling for vocabulary knowledge in German, the ANCOVA indicated a statistical significant intervention effect with a small effect size in favour of the intervention group, $F(1,169) = 4.01$, $p = .047$, $d = 0.31$.

Spelling HSP

Results on the German spelling task reveal that the children from the intervention group achieved a higher mean than the children from the control group. The Cohen's *d* on differences in means suggest an effect sized of 0.32 in favour of the intervention group. Notably, there was a considerable difference in the *SD* between the groups. The *SD* for the children of the control group ($SD = 7.50$) was more than twice as large as the *SD* for the intervention group ($SD = 3.24$). The Levene's test of homogeneity of variance indicated that this difference in variances was significant, $F(1,170) = 4.38$, $p = .038$. A Welch's *F*-ratio (Welch, 1951) correction was applied as a robust method to control for the Type 1 error rate if the assumption of homogeneity of variances is violated. The ANOVA with the Welch corrected *F* did not indicated significant differences between the two groups, $F(1,170) = 2.39$, $p = .124$, $d = 0.24$. However, when controlling for vocabulary knowledge in German, the ANCOVA indicated a statistical significant intervention effect with a small effect size, $F(1,169) = 6.65$, $p = .011$, $d = 0.40$.

4.7.4 Specificity of intervention effects

To control for the specificity of the intervention effects, group differences in number naming were analysed and reported in table 4.8.

***Table 4.8**: Mean raw scores (SD), ANCOVA significant values and Cohen's d effect size for group differences at pre-test (t1) and at post-test (t2) for number naming*

	Mean (*SD*)			
	Intervention group	Control group	*d**	ANCOVA results
Number naming (16)				
*t*1	2.79 (3.13)	3.55 (4.05)		
*t*2	4.53 (4.26)	5.93 (4.88)	-0.18	*F*(1,182) = 1.70, *p* = .194, *d* = -0.19

Note. (), maximum raw scores.
*Cohen's *d*: difference in progress between groups divided by pooled initial *SD*

Number naming

The raw means showed that the control group slightly outperformed the intervention group at pre-test and at post-test. The Cohen's *d* in differences in progress between the two groups was -0.18. ANCOVA did not show a significant difference between the two groups at delayed follow-up, *F*(1,182) = 1.70, *p* = .194, *d* = -0.19.

4.8 Efficacy of the early literacy intervention for children with low oral language proficiency in Luxembourgish

This section addresses the research questions about the efficacy of the intervention for children with low oral language proficiency in Luxembourgish, i.e. the language of instruction. As discussed in the literature review, non-Luxembourgish speakers tend to underperform in the Luxembourg educational system due their low oral language skills.

The specific research question addressed in this section is as follows:

5) To what extent is the early literacy intervention in Luxembourgish beneficial for children with low oral language proficiency in the language of preschool instruction?

The analysis for the children with low OL in the Luxembourgish follows the same analytical approach as the previous analysis for the entire sample, only that the analyses are performed on an at-risk sub-sample.

4.8.1 Identifying children with low oral language proficiency in Luxembourgish

To identify children with low oral language skills in Luxembourgish, a Luxembourgish vocabulary component score was created by performing a PCA on the two vocabulary tasks in Luxembourgish at pre-test (i.e. the PPVT and the CLT). The vocabulary component was created on the entire sample (N = 189) and explained 89% of the variance (KMO = .50, component loadings = .94, Bartlett's p < .001). Children were ranked based on the Luxembourgish vocabulary component and the children performing within the lower tertile (n = 63) were classified as children with LOL.

Children with LOL were equally spread between the intervention and the control group. Out of the 63 children with LOL, 31 children (49%) were in the intervention group and 32 children (51%) were in the control group. Table 4.9 shows the distribution of children with LOL across the preschools as well as experimental groups.

Table 4.9: *Number (%) of children with LOL per preschool, with chi-square test of independence and Cramer's V effect size*

Preschool	*n* / school	Children with LOL	%	χ^2	*p*	*V*
Intervention preschools						
Preschool I	24	7	29			
Preschool II	27	11	40			
Preschool III	26	7	26			
Preschool IV	12	6	50			
n	89	31	36			
Control preschools						
Preschool V	14	3	21			
Preschool VI	22	8	36			
Preschool VII	39	15	38			
Preschool VIII	25	6	24			
n	100	32	30			
				5.26	.628	.17
N	189	63	33			

The percentage of children with LOL ranged from 21-50% per school. There were no significant differences between the number of children with LOL per preschool.

In the following section, background and control variables of the children with LOL are compared to the remaining children of the entire sample not identified as children with low oral language in Luxembourgish (*n* =126, hereinafter referred to as children with no-LOL). Table 4.10 shows the main language spoken at home by the children with LOL and by the no-LOL children.

Table 4.10: *Language backgrounds of children with LOL and their no-LOL peers*

	Children with LOL		Children with no-LOL	
Home language[1]	(*n* = 63)	%	(*n* = 126)	%
Luxembourgish	13	21	97	77
Portuguese	24	38	9	7
French	11	17	6	5
German	1	2	4	3
Other	14	22	10	8

[1]Main language spoken at home as indicated by main caregiver via questionnaire.

While 97 out of the 126 children from no-LOL group (77%) came from a predominantly Luxembourgish speaking home environment, only 13 of the 63 children with LOL (21%) spoke Luxembourgish at home. The most frequent language spoken by the children with LOL was Portuguese (38%). It seems important to note that children from the no-LOL group were predominantly, but not exclusively from a Luxembourgish language home. The reverse pattern was observed for the LOL group, which was predominantly but not exclusively composed of children from a non-Luxembourgish speaking home environment.

The following tables 4.11 to 4.13 provide background characteristics of the children with LOL compared to their no-LOL peers. However, the no-LOL group will not be used as a comparison group for the children with LOL in any of the further analyses. The tables are only for informative purposes. The evaluation of the efficacy of the intervention for children with LOL will focus on children with LOL of the intervention group (*n* = 31) in comparison to children with LOL from the control group (*n* = 32). Table 4.11 compares the mean raw scores for age, SES and non-verbal reasoning between the LOL and the no-LOL group.

Table 4.11: *Means (SD) for age and SES for the LOL and the no-LOL group, with t-test significance values and Cohen's d effect sizes for group differences*

	Mean (*SD*)				
	Children with LOL (*n* = 63)	Children with no-LOL (*n* = 126)	*T*	*p*	*d*
Age (months)					
t1	67.37 (3.88)	68.43 (3.74)	1.82	.070	-0.28
t2	73.97 (3.82)	74.94 (3.63)	1.69	.092	-0.26
t3	81.51 (3.82)	82.02 (4.27)	0.74	.459	-0.13
SES (ISEI[1])	42.28 (20.63)	60.77 (20.66)	5.62	<.001	-0.90

Note. [1]ISEI, International Socio-Economic Index of occupational status;

At no testing point a significant difference for age between the two groups could be detected (all *p*s > .05). However, the children with LOL came from a significantly lower SES.

An overview of the number of boys and L2 learners between the LOL and the no-LOL group is presented in table 4.12.

Table 4.12: *Gender and number (%) of Luxembourgish language learners of the LOL and the no-LOL peer group, with chi-square significance levels and Cramer's V effect sizes for group differences*

	Frequency (%)				
	Children with LOL (*n* = 63)	Children with no LOL (*n* = 126)	χ^2	*p*	*V*
Gender (male)	32 (50.79)	73 (57.94)	0.87	.357	-0.07
Luxembourgish language learners	50 (79.37)	29 (23.02)	54.82	<.001	0.54

The percentage of boys in the LOL group was not statistically different from the percentage of boys in the no-LOL group. However, the LOL group included significantly more L2 learners. This seems unsurprising as the groups were formed on the basis of OL proficiency in Luxembourgish. L2 learners vary in their amount of exposure to Luxembourgish and tend to show relatively lower levels of Luxembourgish language proficiency in comparison to their Luxembourgish L1 speaking peers in early school years (Hoffmann et al., 2018).

Performances on all vocabulary measures at each time point between the children with LOL and with no-LOL were examined and results are reported in table 4.13.

Table 4.13: *Means (SD) for vocabulary in Luxembourgish and German for children with LOL compared to their no-LOL peers, with t-test significance levels and Cohen's d effect sizes for group differences*

	Mean (*SD*)				
	Children with LOL (n = 63)	Children with no-LOL (n = 126)	*T*	*p*	*d*
Vocabulary Lux. CLT (40)					
*t*1	28.49 (3.58)	36.07 (2.17)	18.07	<.001	-2,56
Vocabulary Lux. PPVT (40)					
*t*1	21.92 (4.24)	32.75 (3.67)	18.06	<.001	-2.73
*t*2	26.00 (3.99)	34.09 (3.24)	14.79	<.001	-2.23
*t*3	29.00 (3.50)	35.60 (2.70)	13.46	<.001	-2.11
Vocabulary Ger. PPVT (228)					
*t*3	75.02 (16.10)	111.48 (25.81)	9.49	<.001	-1.70

Note. (), maximum raw scores; CLT, Cross-Linguistic Lexical Task; PPVT, Peabody Picture Vocabulary Test.

Table 4.13 shows that the children in the LOL group were outperformed by the no-LOL group on vocabulary measures in Luxembourgish and in German at all testing points (all *p*s < .05). This was unsurprising as children with LOL were selected based on low oral language skills in Luxembourgish at pre-test. The size of the differences for the PPVT in Luxembourgish revealed a tendency to get smaller from pre-test to delayed follow-up. However, even at the delayed follow-up, the difference in Luxembourgish vocabulary between the children with LOL at pre-test and their no-LOL peers was still huge (d = -2.11).

4.9 Examination of possible study confounds

Background variables and measures of vocabulary were examined with regard to potential differences between the intervention and control group for the children with LOL.

4.9.1 Group differences on background and control variables

Information on important background and control variables (i.e. gender, age, SES, number of Luxembourgish language learners, NVR) between the LOL intervention and the LOL control group are presented in tables 4.14 and 4.15. Frequencies with percentages and Pearson's Chi-square significant levels for group differences are reported for gender and number of Luxembourgish language learners in table 4.14.

Table 4.14: *Number (%) of male students and Luxembourgish language learners in the intervention and in the control group with LOL, with t-test significance levels and Cohen's d effect sizes for group differences*

	Frequency (%)				
	Intervention children with LOL (n = 31)	Control children with LOL (n = 32)	*T*	*p*	*d*
Gender (male%)	14 (45)	18 (56)	0.78	.454	-.11
L2 learners	5 (16)	8 (25)	0.76	.536	-.11

There were no statistical differences between the intervention group with LOL and the control group with LOL (all *ps* > .05) on gender and the number of L2 learners.

Mean raw scores (*SD*) for age, SES and NVR, with significant levels for group differences and effect sizes are reported in table 4.15.

Table 4.15: *Mean raw scores (SD) for age, SES and NVR for children of the intervention and control group with LOL at pre-test, with significance levels and effect sizes for group differences*

	Mean (*SD*)				
	Intervention children with LOL (*n* = 31)	Control children with LOL (*n* = 32)	*T*	*p*	*d*
Age (months					
*t*1	67.07 (3.87)	67.66 (3.93)	-0.60	.550	-0.15
*t*2	73.37 (3.72)	74.56 (3.89)	-1.20	.233	-0,31
*t*3	81.14 (3.79)	81.92 (3.88)	-0.74	.465	-0,21
SES (ISEI[1])	41.67 (21.09)	42.90 (20.49)	-0.23	.819	-0.06
NVR (matrices[2])	*13.61 (3.60)*	*12.03 (4.47)*	*1.54*	*.128*	*0.39*

Note. [1]ISEI, International Socio-Economic Index of occupational status; [2]Matrices WPPSI-III, Wechsler Preschool and Primary Scale of Intelligence.

There were no statistically significant differences between the two groups (all *p*s > .05) on age, SES and NVR at the three testing points.

4.9.2 Measures of vocabulary

An overview of the mean raw scores for the different vocabulary measures of the intervention group with LOL and the control group with LOL at the different testing points is presented in table 4.16.

Table 4.16*: Mean raw scores (SD) for intervention and control children with LOL for vocabulary measures at pre-test (t1), immediately postintervention (t2) and delayed follow-up (t3), with test of differences and effect sizes*

	Mean (*SD*)				
	Intervention children with LOL (*n* = 31)	Control children with LOL (*n* = 32)	*T*	*p*	*d*
Vocabulary Lux CLT (40)					
*t*1	28.23 (3.80)	28.75 (3.39)	-0.59	.565	-0.14
Vocabulary Lux. PPVT (40)					
*t*1	22.13 (4.72)	21.78 (3.81)	0.32	.748	0.08
*t*2	26.60 (3.77)	25.44 (4.17)	1.15	.255	0.29
*t*3	28.71 (3.40)	29.32 (3.65)	-0.63	.534	-0.17
Vocabulary Ger. PPVT (228)					
*t*3	73.93 (14.29)	76.24 (18.13)	-0.52	.607	-0.14

Note. (), maximum raw scores; CLT, Cross-Linguistic Lexical Task; PPVT, Peabody Picture Vocabulary Test. Cohen's *d*: differences in means divided by pooled *SD.*

No significant differences in terms of vocabulary in Luxembourgish and in German were observed between the two groups at any time point (all *ps* > .05). Thus, no statistical control for oral language proficiency at any time point was required.

4.9.3 Group differences on outcome measures at pre-test

Mean raw scores with *SD*s for outcome measures at pre-test for the LOL intervention and the LOL control group are reported in table 4.17.

Table 4.17: *Mean raw scores (SD) for outcome measures at pre-test for children with LOL of the intervention and control group, with t-test significance levels and Cohen's d effect sizes for group differences*

	Mean (*SD*)				
Pre-test outcome measures	Intervention children with LOL (*n* = 31)	Control children with LOL (*n* = 32)	*T*	*p*	*d*
PA Luxembourgish					
Syllable segmentation (12)	9.52 (2.00)	8.81 (2,61)	1.20	.235	0.30
Rhyme detection (12)	3.61 (3.35)	4.63 (3.59)	-1.16	.252	-0.29
Onset–rhyme blending (12)	2.32_F (3.27)	2.84_F (3.41)	-0.62	.538	-0.16
Onset identification (12)	3.10 (2.47)	3.03 (2.52)	0.10	.917	0.03
Phoneme blending (12)	0.29_F (0.90)	1.06_F (2.56)	-1.64	.106	-0.42
Onset manipulation (12)	0.00_F (0.00)	0.00_F (0.00)	/	/	/
Letter-sound knowledge (20)	4.26 (4.80)	3.63 (5.03)	0.51	.611	0.13
Basic reading					
Basic word reading (12)	0.00_F (0.00)	0.00_F (0.00)	/	/	/
Basic non-word reading (12)	0.10_F (0.59)	0.06_F (3.54)	0.30	.767	0.07

Note. (), maximum raw scores; Cohen's *d*: differences in means divided by pooled *SD*.
$_C$ = ceiling effect; $_F$ = floor effect.

Floor effects were observed for the onset-rhyme blending, phoneme blending, onset manipulation and basic reading tasks. Floor and ceiling effects lead to biased parameter estimates and results on these tasks need to be interpreted cautiously. No significant differences could be observed on the outcome measures at pre-test (all *ps* > .05).

Summary on possible study confounds. The between group analyses showed that the intervention and control group with LOL did not differ on any of the background or control variables such as, age, SES, NVR and number of L2 Luxembourgish learners. In addition, the children did not significantly differ on measures of vocabulary in Luxembourgish or in German across the three testing points. As for the analysis for the entire sample, the analysis on the German outcome measures were conducted twice, once without controlling for and once with controlling for vocabulary knowledge in German. No significant differences between the intervention and the control group could be observed for the background variables. Thus, no further actions were taken to control for these possible confounds in subsequent between group analyses.

4.10 Efficacy of the intervention for children with LOL

To evaluate the efficacy of the intervention for children with LOL, the same descriptive and inferential statistics are used as for the analysis of the efficacy of the intervention for the entire sample.

4.10.1 Effects on primary outcome measures

Descriptive and inferential statistics for the primary outcome measures, i.e. PA in Luxembourgish and LSK, at the three time points are reported in table 4.18.

Table 4.18: *Mean raw scores (SD), with ANCOVA significance values and Cohen's d effect sizes for intervention effects at pre-test (t1), at post-test (2) and at delayed follow-up test (t3) for primary outcome measures*

	Mean (*SD*)			
	Intervention group	Control group	*d**	ANCOVA results
Syllable segmentation Lux. (12)				
*t*1	9.52 (2.00)	8.81 (2,61)		
*t*2	9.53 (2.19)	9.28 (2.36)	-0.20[1]	*F*(1,59) = .04, *p* = .845, *d* = -0.06
Rhyme Identification Lux. (12)				
*t*1	3.61 (3.35)	4.63 (3.59)		
*t*2	6.07 (3.67)	6.19 (3.79)	0.26[1]	
Onset-rhyme blending Lux.(12)				
*t*1	2.32F (3.27)	2.84F (3.41)		
*t*2	8.20 (3.38)	4.38 (3.97)	1.30[2]	
Onset identification Lux.(12)				
*t*1	3.10 (2.47)	3.03 (2.52)		
*t*2	6.17 (3.74)	4.13 (3.04)	0.79[1]	
PA large-unit component Lux.				
*t*2 – *t*1				*F*(1,59) = 16.03, *p* < .001, *d* = 1.04
Phoneme blending Lux.(12)				
*t*1	0.29F (0.90)	1.06F (2.46)		
*t*2	3.20 (3.23)	1.34F (2.47)	0.92[2]	
*t*3	8.39 (2.47)	6.32 (3.76)	0.89[2]	
Onset manipulation Lux. (12)				
*t*1	0.00F (0.00)	0.00F (0.00)		
*t*2	1.00F (2.21)	0.06F (0.25)	0.60[2]	
*t*3	3.96 (4.07)	3.04F (3.55)	0.24[2]	
Phoneme segmentation Lux. (12)				
*t*2	2.37F (3.03)	0.63F (1.91)	0.69[3]	
*t*3	7.32 (3.86)	5.96 (3.70)	0.36[3]	
PA small-unit component Lux.				
*t*2 – *t*1				*F*(1.59) = 17.54, *p* < .001, *d* = 1.09
*t*3 – *t*1				*F*(1,50) = 5.95, p = .018, *d* = 0.69.
Letter-sound knowledge (20)				
*t*1	4.26 (4.80)	3.63 (5.03)		
*t*2	13.77 (5.81)	6.22 (5.48)	1.41[1]	*F*(1,59) = 44.67, *p* < .001, *d* = 1.74
*t*3	19.43C (1.32)	18.36C (2.72)	0.09[1]	*F*(1,50) = 3.17, *p* = .081, *d* = 0.51

Note. (), Maximum raw scores.

*Cohen's *d*: [1]difference in progress between groups divided by pooled initial *SD*, [2]difference in progress divided by pooled *SD* at post-test / follow-up (pre-test scores were at floor), [3]difference in means divided by pooled *SD* at post-test / follow-up (pre-test scores were not available).

C= ceiling effect; F= floor effect

Large-unit PA Luxembourgish

In contrast to the entire sample, children with LOL did not show ceiling effects on the syllable segmentation task at neither pre-test nor post-test. The Cohen's *d* on differences in progress (d = -0.20) suggests that the control children with LOL progressed more than the intervention children with LOL. The ANCOVA indicated that children from the control group did not significantly outperform the children from the intervention group at post-test, $F(1,59) = .04$, $p = .845$, $d = -0.06$.

With regards to the large-unit PA measures, the individual Cohen's *d*s on differences in progress showed that the children with LOL from the intervention group had made larger progress from pre-test to post-test than children from the control group. The Cohen's *d*s for differences in progress on the individual large-unit PA measures ranged from 0.26 for rhyme identification to 1.30 for onset blending. The ANCOVA on the large-unit PA component showed that the children with LOL from the intervention group significantly outperformed the children with LOL from the control group at immediate post-test with a large effect size, $F(1,59) = 16.03$, $p < .001$, $d = 1.04$.

Small-unit PA Luxembourgish

Consistent with the findings for the entire sample, the two LOL groups showed floor effects on the small-unit PA measures at pre-test. However, while children from the control group still presented floor effects at post-test, children from the intervention group were able to complete the phoneme blending task. Scores on the onset manipulation task remained at floor for the two groups. The Cohen's *d*s on the differences in progress for the individual small-unit PA measures ranged from 0.60 for onset manipulation to 0.92 for phoneme blending. The ANCOVA on the small-unit PA Luxembourgish component at post-test indicated that the children from the intervention significantly outperformed children from the control group with a large effect size, $F(1.59) = 17.54$, $p < .001$, $d = 1.09$. However, the parameter estimates need to be interpreted with caution due to the floor effects listed above.

At delayed follow-up in Grade 1, the mean raw scores of the individual small-unit PA measures at delayed follow-up indicated that children with LOL had made substantial progress in small-unit phoneme abilities from post-test to delayed follow-up. However, a floor effect for the control group on the onset manipulation task at delayed was still observed. The Cohen's *d* values for differences in progress on the individual small-unit PA measures from pre-test to follow-up ranged from 0.24 to 0.89. The largest maintained effect could be observed for the phoneme blending tasks (d = 0.89). The ANCOVA on the small-unit PA component at delayed follow-up indicated a significant effect in favour of the intervention group, $F(1,50) = 5.95$, $p = .018$, $d = 0.69$. The two ANCOVAs performed at post-test and delayed follow-up on the small-unit PA components remained significant with Bonferroni corrected *p*-values (p = .050/2= .025) to account for multiple comparisons.

Letter-sound knowledge

Immediately postintervention, consistent with the finding for the entire sample, the children from the intervention group knew on average approximately six letters more than the children from the control group. The Cohen's *d* on differences in progress in LSK indicated very large difference in progress in favour of the intervention group (d = 1.41). The ANCOVA on LSK indicated a significant intervention effect with a very large effect size, $F(1,59) = 44.67$, $p < .001$, $d = 1.74$. By the time of the delayed follow-up test in Grade 1, children had been introduced to formal teaching of letters for over five months and as expected, LSK showed a ceiling effect for both groups. Thus, the Cohen's *d* for difference in progress on LSK of 0.09 and the estimates of the ANCOVA on LSK at delayed follow-up need to be interpreted with caution, $F(1,50) = 3.17$, $p = .081$, $d = 0.51$. The ANCOVA performed at post-test on LSK remained significant with Bonferroni corrected *p*-values (p = .050/2= .025) to account for multiple comparisons.

4.10.2 Effects on phonological awareness in German in Grade 1

Raw means scores, tests of differences and effect sizes for PA in German were reported in table 4.19. As lexical knowledge can effect performance on PA tasks in German, tests of differences were computed twice: once without (ANOVA) and once with control (ANCOVA) for vocabulary in German.

Table 4.19: *Mean raw scores (SD), with ANOVA significance values and effect sizes for intervention effects at delayed follow-up test (t3) for PA in German*

	Mean (*SD*)			
	Intervention group	Control group	*d**	ANCOVA / ANOVA results
Phoneme blending Ger. TPB (12)				
*t*3	8.96 (2,83)	7.12 (3,63)	0.57	
Onset manipulation Ger. TBP (12)				
*t*3	4.11 (2,87)	3.08 (3,86)	0.30	
Phoneme segmentation Ger. (12)				
*t*3	7.96 (3,19)	5.88 (3,15)	0.66	
PA small-unit component Ger.				
*t*3				*F*(1,51) = 5.57, *p* = .022, *d* = 0.66
Controlling for vocabulary in German at *t*3				*F*(1,50) = 7.67, *p* = .008, *d* = 0.79

Note. (), maximum raw scores.
TBP, Test für Phonologische Bewusstheitsfähigkeiten [Test for phonological awareness skills].
*Cohen's *d*: difference in progress between groups divided by pooled *SD*.

PA German

Similar to the small-unit PA tasks in Luxembourgish, children with LOL show the higher scores on the phoneme blending and phoneme segmentation tasks, than on the onset manipulation task. The Cohen's *d*s on differences in means for the PA measures ranged from 0.30 for onset manipulation to 0.66 for phoneme segmenting. The ANOVA on the PA component in German showed that the children of the intervention group significantly outperformed the children of the control group on the PA component in German with a moderate effect size, *F*(1,51) = 5.57, *p* = .022, *d* = 0.66. Controlling for vocabulary in German, the ANCOVA revealed a significant difference with a moderate effect size, *F*(1,50) = 7.67, *p* = .008, *d* = 0.79.

4.10.3 Effects on literacy measures

Mean raw scores (*SD*), tests of differences (i.e. ANCOVAs or ANOVAs as appropriate) and Cohen's *d* effect sizes for the secondary outcome measures were reported in table 4.20.

Table 4.20: *Mean raw scores (SD), with ANCOVA / ANOVA significance values and effect sizes for intervention effects at pre-test (t1), at post-test (2) and at delayed follow-up test (t3) for literacy measures*

	Mean (*SD*)			
	Intervention group	Control group	*d**	ANCOVA / ANOVA results
Basic real words (12)				
*t*1	0.00$_F$ (0.00)	0.00$_F$ (0.00)		
*t*2	1.00$_F$ (2.23)	0.13$_F$ (0.42)	0.55[1]	
*t*3	10.82$_C$ (1.87)	9.72$_C$ (3.09)	0.43[1]	
Basic non-words (12)				
*t*1	0.10$_F$ (.054)	0.06$_F$ (0.35)		
*t*2	1.73$_F$ (3.25)	0.31$_F$ (1.12)	0.57[1]	
*t*3	9.86$_C$ (2.31)	9.40$_C$ (3.33)	0.15[1]	
Basic reading component (12)				
*t*2 – *t*1				*F*(1,59) = 6.30, *p* = .015, *d* = 0.65[3]
*t*3 – *t*1				*F*(1,50) = 1.12, *p* = .296, *d* = 0.15[3]
Reading real words SLRT-II (72)				
*t*3	6.14 (6.73)	6.16 (6.59)	0.00[2]	
Reading non-words SLRT-II (72)				
*t*3	11.71 (7.63)	12.80 (7.01)	-0.15[2]	
SLRT reading component				
*t*3				*F*(1,51) = 0.06, *p* = .939, *d* = 0.00[4]
Controlling for vocabulary in German at t3				*F*(1,50) = 0.02, *p* = .960, *d* = 0.00[3]
Reading comprehension ELFE 1-6 (72)				
*t*3	11.04 (4.63)	9.40 (3.81)	0.39[2]	*F*(1,51) = 1.30, *p* = .260, *d* = 0.32[4]
Controlling for vocabulary in German at t3				*F*(1,50) = 1.42, *p* = .239, *d* = 0.34[3]
Spelling HSP-graphemes (40)				
*t*3	33.14 (3.88)	28.60 (10.66)	0.57[2]	*F*(1,51) = 4.42, *p* = .040, *d* = 0.59[4]
Controlling for vocabulary in German at t3				*F*(1,50) = 5.38, *p* = .024, *d* = 0.66[3]

Note. (), maximum raw scores. SLRT II, Salzburger Lese- und Rechtschreibtests [Salzburg reading and orthography Test]; ELFE 1– 6, Ein Leseverständnistest für Erst-bis Sechstklässler [ELFE 1– 6: A reading comprehension test for students in Grades 1 through; HSP, Hamburger Schreib-Probe [Hamburg spelling-task].

*Cohen's *d*: [1]difference in progress between groups divided by pooled *SD* at post-test / follow-up (pre-test scores were at floor); [2]difference in means divided by pooled *SD* at follow-up.

[3] ANCOVA; [4] ANOVA; $_C$ = ceiling effect; $_F$ = floor effect.

Basic reading

In line with the findings for the entire sample, the basic word reading task and the basic non-word reading task presented floor effects at pre-test for the children with LOL. At pre-test, no child with LOL in either group could read a real word in the basic word reading task, and only one child out of each group could read three basic non-words. Immediately post-intervention, the two groups still showed floor effects. The Cohen's *d*s on the differences in progress from pre- to post-test range from 0.55 for basic word reading to 0.57 for basic non-word reading. The ANCOVA on the basic reading component showed that the children from the intervention group significantly outperformed the children from the control group at post-test, $F(1,59) = 6.30$, $p = .015$, $d = 0.65$. However, both the effect sizes and the ANCOVA significant values need to be interpreted very cautiously due to the floor effects.

At delayed follow-up, mean raw scores on the reading basic words and the reading basic non-words showed a ceiling effect. This indicates that the children had rapidly acquired basic word reading skills within approximately five months of formal literacy instruction. The Cohen's *d*s on differences in progress from pre-to delayed follow-up test ranged from 0.15 for non-word reading to 0.43 for word reading. The ANCOVA on the basic reading component did not show that the intervention group significantly outperformed the control group at delayed follow-up, $F(1,50) = 1.12$, $p = .296$, $d = 0.15$. However, the parameter estimates at delayed follow-up need to be interpreted cautiously due to the ceiling effects.

The ANCOVA performed at post-test on the basic reading component did remain significant with Bonferroni corrected p-values ($p = .050/2 = .025$) to account for multiple comparisons.

Word- and non-word reading SLRT

Results on the two SLRT reading tasks showed that the children with LOL from the two groups managed to read words and non-words in German after five months of formal literacy instruction. The two groups performed very similarly on the two SLRT tasks and the Cohen's *d* on differences in means suggests negligible effects for SLRT word reading ($d = 0.00$) and SLRT non-word reading ($d = -0.15$).

The ANOVA on the SLRT reading component did not reveal any statistic differences between the groups, $F(1,51) = 0.06$, $p = .939$, $d = 0.00$. As lexical knowledge can effect performance on reading, an ANCOVA on the SLRT reading component controlling for German vocabulary was conducted. The results indicated no significant differences between the two groups, $F(1,50) = 0.02$, $p = .960$, $d = 0.00$.

Reading comprehension ELFE

On the German reading comprehension measure, the data showed a higher mean raw score for the children from the intervention group than for children from the control group. The Cohen's *d* on differences in means suggest an small effect size of 0.39 in favour of the intervention group. The between group ANOVA, however, indicated that the effect was not significant, $F(1,51) = 1.30$, $p = .260$, $d = 0.32$. When controlling for vocabulary knowledge in German, the ANCOVA results indicated nonsignificant differences for reading comprehension, $F(1,50) = 1.42$, $p = .239$, $d = 0.34$.

Spelling HSP

Results on the German spelling task show that the children from the intervention group achieved a higher mean than the children from the control group. The Cohen's *d* on differences in means suggest an effect size of 0.32 in favour of the intervention group. Consistent with data on the entire sample, there was a considerable difference in the *SD* between the groups. The *SD* for the control group ($SD = 10.66$) was almost three times as large as the SD for the intervention group ($SD = 3.88$). Yet the Levene's test did not indicate that the assumption of homogeneity of variances had been violated for the spelling task, $F(1, 51) = 0.07$, $p = .793$.. The ANOVA indicated that the effect was statistically significant, $F(1,51) = 4.42$ $p = .040$, and of medium size ($d = 0.59$). When controlling for vocabulary knowledge in German, the ANCOVA indicated again a statistical significant intervention effect with a medium effect size, $F(1,50) = 5.38$, $p = .024$, $d = 0.66$).

4.10.4 Specificity of intervention effects

To control for the specificity of the intervention effects for the LOL group, differences in number naming were analysed and reported in table 4.21.

Table 4.21: *Mean raw scores (SD), ANCOVA significant values and effect size for group differences at pre-test (t1) and at post-test (t2) for number naming*

	Mean (*SD*)			
	Intervention group	Control group	*d**	ANCOVA results
Number naming (16)				
*t*1	1.45F (1.15)	1.53F (1.52)		
*t*2	2.83 (2.91)	3.53 (3.46)	-0.19[1]	*F*(1,59) = 0.42, *p* = .520, *d* = -0.17

Note. (), maximum raw scores; F = floor effect.

*Cohen's *d*: difference in progress between groups divided by pooled *SD* at post-test (pre-test scores were at floor).

Number naming

A *t*-test showed not statistical group differences at pre-test ($t(61) = -0.23$, $p = .816$, $d = 0.06$), and controlling for the autoregressor, no significant group differences could be detected for the number recognition task at post-test.

4.11 Discussion: Efficacy of the intervention study

The overall goal of Study 1 was to evaluate the efficacy of an intervention programme that targeted early literacy skills. Early intervention programmes are important as research has clearly shown that children who experience difficulties early are more likely to experience considerable difficulties later (Hulme & Snowling, 2009; Knudsen et al., 2006). Literacy development is a concern in Luxembourg because over 40% of Luxembourg's Grade 3 pupils are not meeting expected reading levels, and this figure increases if only L2 learners are considered (Hoffmann et al., 2018). The current study adopted a quasi-experimental study to evaluate the efficacy of newly developed intervention programme aiming to improve early literacy skills in Luxembourgish for a sample of linguistically diverse children growing up in Luxembourg. The study included a pre-test and a post-test in Year 2 of preschool, and a follow-up assessment phase in Grade 1. The group averages of the children from the intervention group were compared to children who followed the regular preschool programme in Luxembourg on primary outcome measures (i.e. PA and LSK in Luxembourgish), secondary outcome measures (i.e. PA in German and literacy measures) and control measures (vocabulary and number naming). In a first section, results on the entire sample are discussed, and a second section considers results for the subgroup of at-risk children identified via LOL proficiency in the instruction language. Notably 79% of these children were L2-learners.

The discussion of the results is based on significance values and the effect sizes (Cohen's *d*s) of the ANOVAs/ANCOVAs. Yet effect sizes on the differences in progress or differences in means between the two groups are also considered. Based on the guidelines from the Network Promising Practices (2007) and What Works Clearing House (2007), a Cohen's *d* equal or above of 0.25 is considered to be substantively important, even though they may not reach statistical significance.

4.11.1 Effects on primary outcome measures at post-test for the entire sample

Primary outcomes consisted of PA and LSK in Luxembourgish. The results at post-test for the entire sample indicated that the intervention group significantly outperformed the control group at immediate post-test on the large-unit PA component, the small-unit PA component and LSK, with effect sizes that were in the large to very large range. Thus the current study replicated a wealth of research showing that explicit and systematic combined training of PA and LSK can effectively promote early literacy skills within a relatively short period of time (Bowyer-Crane et al., 2008; Ehri, Nunes, Stahl, et al., 2001; Fricke, Bowyer-Crane, Haley, Hulme, & Snowling, 2013; Hatcher, Hulme, & Ellis, 1994; Lundberg, 1994; Richards-Tutor et al., 2016; Wawire & Kim, 2018) and extends it to a linguistically diverse pupil population in the preschool context of Luxembourg.

Phonological awareness

A more detailed analysis of the individual Cohen's *d* effect sizes on differences in progress for each PA measure showed that the intervention group showed larger progress than the control group on each PA measure, with the exception for syllable segmentation. However, this counter-current finding was to be expected. Segmenting spoken words into individual sounds was a considerable part of the intervention programme and hence, some intervention children found it difficult to segment words into syllables instead of phonemes immediately postintervention. Even after explicit instruction and repeated corrections during the training items, some children from the intervention group were not able to switch from segmenting phonemes to segmenting syllables. A response pattern that had not been observed at pre-test.

Small positive intervention effects were observed for rhyme identification and onset manipulation immediately post-intervention in favour of the intervention group. Medium intervention effect sizes were detected for phoneme segmentation and onset identification, and large intervention effects could be observed for phoneme blending measures and onset-rhyme blending. This pattern of effect sizes aligns well with the content of the intervention. Onset identification and phoneme blending were the two phonological skills targeted by the majority of the phonological awareness activities. Phonological tasks covering the rhyme unit were only included in the first week of the intervention and phoneme segmentation was only covered in the last four weeks. Activities directly targeting phoneme manipulations were not part of the intervention. Out of the seven individual PA measures at post-test, the largest differences in progress effect could be observed for onset-rhyme blending (d = 0.97), followed by the intervention effect for phoneme blending (d = 0.90). Phonological blending was a large part of the intervention training (mainly due to the activity called “Papageiensprooch” [parrot talk], see section 3.9.2) and the results indicate that it was effective.

Notably, while 10 out of the 12 weeks of the intervention focused exclusively on the training of phonemic awareness and only 2 weeks covered PA training on larger linguistic units (syllables, rhyme and onset-rhyme), the intervention effects on the large-unit PA component (d = 1.09) and the small-unit PA component (d = 0.90) were comparably large immediately postinterventinon. An explanation for the similar effect size despite not being equally covered in the intervention was that the onset identification task was included in the large-unit PA component, despite assessing both larger onset- and initial phoneme identification. Initial phoneme identification was the single most trained skill during the intervention. Hence, it was not unsurprising to see a large intervention effect on the large-unit PA component although small-unit PA skills were predominately trained during the intervention. In addition, although debated, PA has been conceptualized as a unified construct of a single cognitive ability (Anthony & Francis, 2005; Anthony & Lonigan, 2004)***.*** Hence, it could be that the greater sensitivity to the phonemic sound structure of language also led to improved performances on tasks tapping larger linguistic units.

Previous work in Luxembourg by Bodé and Content (2011) examined the efficacy of PA training only (without letter support and mainly focusing on larger PA units) in preschool (age 5-6). The authors evaluated effect of daily PA trainings sessions of ten minutes over 20 weeks. They found small to medium effect sizes on PA measures of various degrees of explicitness and linguistic units. The effect sizes of the PA effects of the current project immediately post-intervention are larger than the effect sizes reported by Bodé and Content (2011). For example, Bodé and content (2011) report a moderate effect size (d = 0.66) for phoneme blending, whereas a strong effect on phoneme blending was observed in the current study (d = 0.90). There are a few possible explanations for the larger effect sizes of the current study. Firstly, Bodé and Content only administered pure PA training without any training in LSK and it is well-established that PA training incorporating LSK is more effective than training without letters on PA outcome measures (Bus & van IJzendoorn, 1999; Fischer & Pfost, 2015; Schneider et al., 2000). Secondly, the PA training in Bodé and Content's (2011) field study focused more on the training of larger-unit PA measures than the intervention programme of the present study. Thus, larger effect sizes on the small-unit measures in the current study than in Bodé and Content's (2011) study are not surprising. Another possible reason for the larger effect sizes in the current study is that Bodé and Content (2011) reported that only about half of the teachers reported that they fully applied their programme until the last stages. In contrast, all 14 intervention teachers administered all 48 training session in the current study.

Letter-sound knowledge

The effect size of both the differences in progress (d = 1.21) and the ANCOVA (d = 1.78) indicated a very large effect for LSK. The very large effect on LSK immediately postintervention exceeds effect sizes found in previous research (Bus & van IJzendoorn, 1999; Fischer & Pfost, 2015; NRP, 2000), which could be predominantly explained by two factors. Firstly, a main reason for the effects size for LSK in the current study is arguably, the limited explicit teaching of LSK by the teachers in the control group. Although the children from the control group also acquired on average two new letters over the period of the intervention, no systematic an explicit teaching of letters is foreseen in preschool according to the standard preschool curriculum in Luxembourg. At the end of Year 2 in preschool in Luxembourg, children are expected to distinguish letters from other graphic symbols and to be able to identify the letters in their name and their friends' names (MENFP, 2011). A second reason for the very large intervention effect for LSK was that the extensive and systematic introduction and revision of the 20 introduced letters was a main component of the early literacy intervention. Each of the last 39 session of the intervention included ample opportunities to revise letter-sounds.

Taken together the results on the primary outcome measures at immediate post-test have shown that the intervention was highly effective in training PA and LSK. The effect sizes are higher than reported in a recent meta-review on the effectiveness of PA and LSK training in German speaking settings (Fischer & Pfost, 2015). However, there are limitations to the comparison of effect sizes across studies, due to large variations in factors impacting the intervention efficiency, e.g. doses of instruction, fidelity monitoring, implementation quality, intervention administrator, educational and cultural context, curriculum followed by control group and at-risk status of the population (Innocenti et al., 2014; NRP, 2000).

4.11.2 Effects on primary outcomes at delayed follow-up for the entire sample

Children were followed from preschool to the middle of Grade 1. By the time of the follow-up Grade 1 children had followed literacy instruction for five months. The large-unit PA measures were dropped at follow-up due to post-tests trends and anticipated ceiling effects. Results on the small-unit PA (Luxembourgish) component in Grade 1 showed that the intervention group still significant outperformed the control group, although the intervention effect sizes had declined from post-test (d = 0.90) to delayed follow-up (d = 0.48). This maintained effect aligns with the significant long term intervention effects on PA measures in the moderate range (ds ranging from 0.48 to 0.79) reported in earlier meta-reviews in English-speaking contexts (Bus & van IJzendoorn, 1999; Ehri et al., 2001). A more recent meta-review on intervention studies in German-speaking contexts reported smaller long term effects on PA measures (ds ranging from of 0.27 to 0.28) (Fischer & Pfost, 2015; Suggate, 2016). However, it needs to be noted that the stability of the intervention effects are contingent on myriad factors. For example, the time interval between the post-test and the delayed follow-up test may be source of heterogeneity in the reported decline between post-test and delayed follow-up. Suggate’s (2016) analysis of long term effects included studies with average time interval from post-test to follow-up of 11.17 months, whereas Fisher and Pfost (2015) considered any testing conducted within one year after the intervention as tapping short term effects. Fisher and Pfost (2015) only included studies with a time from post-test to follow-up of over a year in their analysis of long term effects. The time interval between the post-test and delayed follow-up in the current study was approximately nine months, which could explain why the maintained effects on PA in the current study are larger than reported in some meta-reviews. It is not possible to directly compare the maintained effects on PA and LSK to the only other PA intervention administered in the context of Luxembourg, as Bodé and Content (2011) did not include measures of PA and LSK at follow-up in Grade 1.

At delayed follow-up in Grade 1, children from the intervention and the control group showed a ceiling effect on the LSK task. This is not unusual, as mastery of LSK after few months of literacy instruction in German and in other consistent orthographies have been reported (Aro & Wimmer, 2003; Wimmer & Hummer, 1990; Seymour et al., 2003). Consistent orthographies such as German, show more reliable mappings of graphemes onto phonemes than in more inconsistent orthographies with a more ambiguous mapping of graphemes onto phonemes. A more consistent mapping allows children to quickly learn the grapheme to phoneme correspondences which in turn leads to a faster proficiency in the alphabetic principle (Aro & Wimmer, 2003).

4.11.3 Effects on phonological awareness in German for the entire sample

At delayed follow-up, children were assessed on PA in German to explore whether the interventions effects on PA in Luxembourg transferred to measures of PA in German.

Performance on the phoneme blending, segmenting and manipulation tasks in German showed that children from both the intervention and the control were able to perform PA operations in German after five months of oral and written German instruction in Grade 1. The individual effect sizes on differences in means showed that the intervention group outperformed the control group on all PA measures in German in Grade 1, with effect sizes ranging from 0.05 for onset manipulation to 0.46 for phoneme segmenting. The ANOVA result on the small-unit PA component confirmed a significant intervention effect with a small effect size (d = 0.31). this provides strong evidence for a transfer of the intervention effects in PA in Luxembourgish to PA in German. Controlling for lexical knowledge in German, the intervention effect increased (*d* = 0.51). Thus, the current study suggests that a training of PA skills in Luxembourgish generalises to PA skills in German. It has been suggested that a transfer of linguistic skills is more likely between two languages that are phonologically and orthographically similar (Bialystok et al., 2003; Edele & Stanat, 2016; Rosowsky, 2001). Luxembourgish and German share a high degree of linguistic similarity (Gilles & Trouvain, 2013) and thus, cross-linguistic intervention effects were not unexpected. The current data supports the view that children might apply their Luxembourgish PA knowledge to bootstrap their PA skills in German (Wawire & Kim, 2018). While Wawire and Kim (2018) looked at cross-linguistic intervention effects immediately post-intervention, the current study explored cross-linguistic effects nine months after the intervention had stopped. With regards to the effect sizes on differences in means for the individual PA measures in German, a small effect was observed for phoneme blending (*d* = 0.25) and phoneme segmentation (*d* = 0.46), and a negligible effect for onset manipulation (*d* = 0.05).

4.11.4 Effects on literacy measures at post-test for the entire sample

Although the training did not contain any word decoding activities and children had not been introduced to literacy by the time of the post-test, the results showed that children from the intervention group were able to read on average three words on the basic non-word reading task, whereas the same measure presented a floor effect for the control group at post-test. The individual effect size on the differences in progress for the basic non-word reading task indicated a moderate effect (*d* = 0.73) in favour of the intervention group. This finding is line with a body of research showing that supporting children's PA and LSK leads to improvements in early decoding skills (Bowyer-Crane et al., 2008; Bradley & Bryant, 1983; Brown, 2014; Lonigan, Burgess, & Anthony, 2000; Lundberg, Frost, & Petersen, 1988; Melby-Lervåg et al., 2012; NRP, 2000; Schneider et al., 2000).

The two groups still presented a floor effect on the basic word reading measure at post-test. However, children from the intervention group were able to read on average two words on the basic word reading tasks, whereas the children from the control group were able to read only one word on average. This suggests that the positive skewness due to the floor effects is more pronounced for the children of the control group. A possible explanation for the floor effects in the basic word reading tasks and not on the basic non-word reading tasks could be that the first four items in the basic non-word reading tasks required children to blend non-words composed of only two letters. In contrast, the first four items in the basic word reading task consisted of words composed of three letters. Thus, the increased complexity of the first items of the basic-word reading task may partly explain why basic word reading showed floor effects for the intervention groups, whereas basic non-word reading did not.

The ANCOVA at post-test on the basic decoding component suggested a significant large effect (d = 0.80) in favour of the intervention group. However, the statistical significant differences on the basic word reading component needs to be interpreted with caution due to the floor effects of the control group on the basic word- and basic non-word reading tasks, and the floor effects of the intervention group on the basic word reading tasks. Performance on the basic reading component could be interpreted as tapping the initial steps towards a rudimentary understanding of the alphabetic principle in the children of the intervention group. The study extends previous work by showing that explicit and systematic training of PA and LSK supports children's literacy development in its early stages (Bowyer-Crane et al., 2008; Hatcher, Hulme, & Snowling, 2004; Lonigan, Burgess, & Anthony, 2000; Lundberg et al., 1988; Melby-Lervåg et al., 2012; Muter, Hulme, Snowling, & Stevenson, 2004; Schneider et al., 2000).

4.11.5 Effects on literacy measures at delayed follow-up for the entire sample

A secondary aim of the current investigation was to explore whether intervention supported children's response to literacy development in Grade 1, nine months after the intervention had ended.

After five months of literacy instruction in Grade 1, the basic non-word reading and the basic reading tasks presented ceiling effects. This is not unsurprising, as an initial steep increase in early reading abilities within the first months of formal literacy instruction has been reported in other consistent writing systems (Caravolas, Lervåg, Defior, Seidlová Málková & Hulme, 2013). As mentioned above, a more consistent mapping of graphemes onto phonemes in consistent languages like German allows children to quickly learn the grapheme to phoneme correspondences, which in turn leads to more proficient decoding and encoding skills at early stages (Aro & Wimmer, 2003).

With regards to the German reading measures, the two groups were able to complete the SLRT word reading and the reading comprehension tasks in Grade 1. On the SLRT word reading and the SLRT non-word reading task, the two groups showed comparable performances. The individual effect sizes on the differences in means suggested negligible effect in favour of the control group for the SLRT non-word reading (d = -0.09) and the SLRT word reading (d = -0.15) tasks. The ANCOVA on the differences between the two groups on the SLRT was nonsignificant with a negligible effect size (d = -0.13). Controlling for vocabulary knowledge in German, did not change the ANCOVA results for the SLRT reading component (d = 0.00). This result does not bear resemblance with the general findings from English studies on L1 and L2 learners showing that improvements in PA and LSK lead to better word-level reading outcomes (Bowyer-Crane et al., 2008; Bus & van IJzendoorn, 1999; Stuart, 1999). However, as mentioned above, it has been shown that early decoding skills are acquired easier in German than in English and thus, additional PA training may not generalise to measures of early word reading. Schneider et al. (2000) showed that their PA and LSK training in preschools in German-speaking settings had long term effects on measure of early reading skills. However, their sample only included at-risk children, identified based on low PA in preschool, and it has been shown that at-risk children show longer long-term benefits from early support (Bus & van IJzendoorn, 1999; Ehri et al., 2001; Fischer & Pfost, 2015). With reference to a meta-review conducted in German-speaking settings reviewing studies with at-risk and non at-risk sample, Fischer and Pfost (2015) reported minimal long term meta-effect of PA trainings on measures of early decoding (d = 0.03), which is in line with the current results.

With regards to reading comprehension in Grade 1, a nonsignificant Cohen's *d* of 0.18 was observed without controlling for German vocabulary. However, a significant intervention effect could be observed when controlling for vocabulary in German (*d* = 0.32). This suggests that the delivered intervention in Luxembourgish had a stronger transfer effect on reading comprehension than on the SLRT component. Fischer and Pfost (2015) also showed that PA interventions in German tend to yield stronger transfer effects to measures of reading comprehension (*d* = 0.16) than to measures tapping early decoding skills (*d* = 0.03), which is in line with the current findings. It has been posited that reading comprehension assessments in early stages of literacy development are mediated by decoding skills (Cutting & Scarborough, 2006) and thus, decoding skills are the main limiting factor on reading comprehension (Perfetti, Landi, & Oakhill, 2005). However, it could be that additional support in early literacy skills did not led to faster decoding skills in absolute terms, but that it led to more automatized processes which in turn freed up cognitive resources that could have been employed for reading comprehension (Oakhill, Cain, & Elbro, 2015; Perfetti, 2007).

Concerning spelling in Grade 1, the children from the intervention group outperformed the children from the control group. The effect size on the difference in means was small but educationally important (*d* = 0.32). The ANCOVA revealed a statistically nonsignificant effect with a Cohen's *d* of 0.24. However, controlling for lexical knowledge in German, a statistically significant intervention effect was observed for spelling in Grade 1, with a Cohen's *d* of 0.40. Previous work has suggested that supporting PA and LSK in preschool can have positive effects on measure of spelling in German (Bodé & Content, 2011; Schneider et al., 2000). Out of the literacy skills assessed in the current study, the strongest intervention effect size was observed for spelling, which is in line with the meta-review by Fischer and Pfost (2015). The stronger effect of PA combined with LSK trainings for spelling than for reading measures in learning is again explained by the more inconsistent mappings of graphemes onto phonemes than the more consistent mapping of phonemes onto graphemes in German (Landerl & Thaler, 2006) (see section 1.6.2).

In summary, results on literacy measures showed statistical significant differences for reading comprehension (d = 0.31) and spelling (d = 0.40) when controlling for vocabulary in German. According to the guidelines from the Network Promising Practices (2007) and What Works Clearing House (2007), a Cohen's d above 0.25 can be considered educationally meaningful.

Result on the number naming task did not show any differences between the two groups. This suggests that while the intervention exerted significant effects on early literacy development, it did not have an impact on children's performance on a number naming test indicating that the effects were specific.

4.12 Efficacy of the intervention for children with low oral language skills (LOL) in the language of instruction

In addition to the analysis for the entire sample, intervention results for a specific subgroup of children identified as having LOL proficiency in the instruction language were explored. Children with LOL were identified based on performances in the lowest tertile on receptive vocabulary in Luxembourgish at pre-test (n = 63). Notably 50 out of the 63 (79%) of the LOL children were L2 learners. The two main languages spoken by the L2 learners were Portuguese (48%) and French (22%). The question whether this specific subgroup also benefited from the intervention was deemed important as it could be that children with limited proficiency in the school language might not be able to fully engage in the intervention activities that have a heavy language component, which in turn might reinforce their disadvantage further. The same analytic approach was applied for the analysis of this subgroup of children as for the entire sample. However, given the small sample size (n = 63) and the resulting reduced power to determine statistical significance, measures of effect sizes seemed more appropriate to assess intervention effects than indicators of statistical significance (Ferguson, 2009; Maxwell et al., 2008; Wilkinson, 1999). The same guidelines of educational importance as for the entire sample are considered, i.e. a Cohen's d equal or above 0.25 (Network Promising Practices, 2007; What Works Clearing House, 2007).

Firstly, it needs to be noted that poor performance in the language of instruction appeared to be a valid criterion to identify children at risk of reading failure. The LOL group performed significantly lower on all early literacy measures in comparison to their non-LOL peers. This is in line with previous work on L2 learners showing that L2 learners show lower performances on measures of PA and LSK in preschool or Grade 1 in comparison to their L1 peers (Hoff, 2013; Muter & Diethelm, 2001; Páez et al., 2007; Verhoeven, 2000).

4.12.1 Effects on primary outcome measures for the LOL group

The overall pattern of results on the primary outcome measures (i.e. PA and LSK in Luxembourgish) at the immediate post-test shows that the intervention has also been highly efficient for the LOL group.

The Cohen's *d*s on differences in progress between the two groups on the individual PA measures showed that the intervention progressed more than the control group on all measures of PA. A small effect was observed for rhyme identification (d = 0.26), medium effect sizes were detected for onset identification (d = 0.79), phoneme segmentation (d = 0.69) and phoneme manipulation (d = 0.60), and large intervention effects could be observed for onset-rhyme blending (d = 1.30) and phoneme blending measures (d = 0.89). This pattern of effect sizes aligns well with the efficacy of the intervention for the entire sample. Statistical significant difference in favour of the intervention group, with large effect sizes have been found for the small- and large-unit PA components. The effect sizes of both differences in progress (d = 1.41) and the ANCOVA (d = 1.74) also indicated a very large effect for LSK immediately postintervention.

A comparison of the effect sizes on the primary measures for the entire sample and for the LOL group need to be interpreted with caution because the results of current analysis are not independent of the analysis for the entire sample. However, the results on the primary outcome measures suggest that the training was as effective for the children with LOL as for the entire sample. This is in line with previous research showing that children at-risk can also benefit from the early literacy intervention (Ehri et al., 2001; Stuart, 1999; Weber et al., 2007). However, this is particular encouraging finding for Luxembourg. As mentioned above, 50 out of the 63 children (79%) in the LOL subgroup were L2 learners. So far, not many studies have explored the effectiveness of early literacy training for L2 learners, which is problematic for Luxembourg as it has been shown that L2 learners in Luxembourg represent a large at-risk population (Hoffmann et al., 2018). Given the large percentage of L2 learners in the current subsample, the effect sizes for the children with LOL are encouraging for teachers, as they confirm that teachers can effectively teach foundational literacy skills in Luxembourgish to young L2 learners. At delayed follow-up, in line with the findings from the entire group, LSK also showed ceiling effects in the two LOL groups. However, statistically significant differences with a moderate effect sizes in favour of the intervention group for the small-unit PA component in Luxembourgish were still observable in Grade 1. This confirms previous findings showing that at-risk children also show maintained effects (NRP, 2000; Suggate, 2016).

4.12.2 Effects on phonological awareness in German for the LOL group

A medium intervention effect on the PA component in German could be identified for the LOL children at delayed follow-up. The effect sizes for the LOL group (d = 0.66) was much larger than for the entire sample (d = 0.31). When controlling for vocabulary in German, that effect approached a large effect size for the LOL group (d = 0.79). This suggests that the at-risk children may have benefitted more from the intervention. However, as mentioned above, this comparison of effect sizes needs to be interpreted carefully as the analyses for the entire sample also includes the children with LOL and the results are not independent of each other. However the sub-analyses on the sample with LOL suggests that the additional support in PA skills in Luxembourgish seem to allow children with LOL to bootstrap the PA skills in German (Wawire & Kim, 2018).

4.12.3 Effects on literacy measures for the LOL group

Concerning the performance on the basic word reading and basic non-word reading measures in preschool, the children with LOL showed floor effects on the two tasks at the two time points. This suggests that, in contrast to the findings from the entire group, the additional support in PA and LSK did not generalise to early non-word reading skills immediately post-intervention for the intervention group. The additional support in PA and LSK did not yet allow the children with LOL from the intervention group to apply the alphabetic principle to decode basic non-words. However, in line with the findings for the entire group, at delayed follow-up, the two basic reading tasks also presented a ceiling effect for the children with LOL. This is an encouraging finding as it suggests that the rapid initial increase in literacy skills observed for the entire sample also holds for the LOL groups (Caravolas et al., 2013).

With regards to the German literary measures in Grade 1, the effect size of on the SLRT reading component indicated a negligible intervention effect (*d* = 0.00), whereas a small effect could be detected for reading comprehension (*d* = 0.32) and a moderate intervention effect for spelling (*d* = 0.59). Controlling for vocabulary in German, the effect sizes for reading comprehension (*d* = 0.34) and spelling (*d* = 0. 66) were even larger. With reference to the threshold of an educationally important Cohen's *d* of 0.25, the results for the LOL group can be considered educational important (Network Promising Practices, 2007; What Works Clearing House, 2007).

In summary, in line with a limited amount of research on the support of early literacy skills for L2 learners, the current findings suggests that children with LOL in the language of instruction also benefitted from additional support in PA and LSK in the school language (Stuart, 1999; Weber et al., 2007). Interestingly, while the effects on the primary measures between the entire sample and the sample with LOL are comparable, the effects on the secondary measures in Grade 1 (i.e. PA in German and the reading comprehension and spelling task) seem larger for the children with LOL. Greater benefits of early additional support of early – literacy skills for at risk-children than for non at-risk has been reported in previous work (Hulme & Snowling, 2009; NRP, 2000). Although comparisons between effect sizes for the entire sample and the group with LOL need to be drawn with caution due to the non-independent samples, the current findings provide first support that children with LOL in the language of instruction benefit more from the intervention that children who are proficient in the school language.

4.13 Strengths, weaknesses and future directions

A major strength of the current work was the evaluation of the efficacy of the intervention by using a thoroughly designed quasi-experimental design in a real world context (Coolican, 2009). The study featured many important characteristics of rigorous quasi-experimental studies: a control group, matching procedures to form comparable groups, control for confounders, time-series data (a pre-test, post-test) and a nine months delayed follow-up. Follow-up assessments provide a more nuanced, complex picture of the maintenance of possible intervention effects and can be considered a particular strength of the current study (Innocenti et al., 2014; Von Allmen et al., 2015), as not all early literacy intervention studies include delayed follow-up testing point (cf. Wawire & Kim, 2018).

Another strong point of the current work was the myriad steps taken to ensure fidelity of assessment and implementation. For example, research assistants helping with the testing of the children were extensively trained at each testing point. All assistants participated in theoretical training session of three hours, followed by mock testing on other research assistants. On a second day, a testing-out session in a preschool that was not part of the study was arranged prior to each assessment phase in order to train the research assistants in an real testing setting with a child. This ensured that the testing was administered homogenously across the assistants to reduce experimenter bias (Innocenti et al., 2014). Various measures of intervention implementation fidelity were also administered and confirmed that the teachers administered the intervention as intended by the research team. The implementation quality ratings of the intervention sessions showed that teachers delivered the intervention according to the manual and children demonstrated good responsiveness ratings during the intervention sessions. Children presented very high attendance rates. The narrow ranges and small standard deviations on the implementation quality and responsiveness ratings imply that the intervention was relatively homogenously administered across the 14 classrooms. A tremendous strength of the current study was the high acceptability ratings by teachers and caregivers, which may explain why all 14 intervention teachers administered all 48 intervention sessions. This can be considered a strength of the current work as other studies have reported less consistency in the delivery of the intervention where many teachers did not manage to fully administer all segments of the early literacy intervention (Bodé & Content, 2011; Schneider, Küspert, Roth, Visé, & Marx, 1997).

In addition, extensive efforts were put into the development of an early literacy programme which was specifically tailored to the Luxembourgish context. The intervention included many different components: PA activities, LSK training, shared book reading, musical components, or material for caregivers. This can be considered a strength from an applied perspective as the intervention provided more holistic support (Carroll et al., 2011). However, from a theoretical and experimental perspective, this could be considered a weakness. The different components of the training may have distinctive effects on different aspects of literacy acquisition and it is not possible to identify which components of the intervention were crucial for the effectiveness of the intervention. Follow-on studies exploring the effectiveness only administering certain components of the intervention would help to identify the important elements of the training on distinct outcome measures (similar to the work of Schneider et al., 2000). For example, a study could examine the efficacy of only shared book reading combined with LSK training to identify the importance of the PA component of the training in children learning to read in German in Luxembourg. Another study could explore the efficacy of the intervention without providing caregivers' with the support material to establish the importance of the caregiver participating. Another study could examine the efficacy of only providing caregivers the support material without any additional support in classrooms (cf. Sheridan, Knoche, Kupzyk, Edwards, & Marvin, 2011).

Another asset of the current work is its high ecological validity. Voices of limited applicability of research findings to real world settings in health research have been uttered (Butler, 2008; Green & Glasgow, 2006). However, the current study was conducted in a real world setting and bore a close resemblance to the preschool context as it exists in Luxembourg. The classroom teachers administered the intervention themselves in their normal classroom setting and without much additional resources, any preschool teachers in Luxembourg could hence easily integrate the intervention in their normal teaching.

A fundamental limitation of the current study is the main drawback of quasi-experimental research, i.e. the lack of random allocation (Coolican, 2009). Schools had not been randomly allocated on a unit-basis, and uncontrolled factors such as different teaching methods (e.g. greater emphasis on German vocabulary) by the Grade 1 teachers or cultural and socio-economic factors could have confounded the results (Beck, McKeown, & Kucan, 2013; Hoff, 2006, 2013; Webb, 2007). Random allocation could have minimized the pre-test differences observed on the large linguistic PA units and LSK. Although, pre-test differences on outcomes measures are not uncommon in quasi-experimental design (cf. Lundberg, Frost, & Petersen, 1988; Stuart, 1999) they need to be considered a weakness. Adequate procedures were taken to statistically control for pre-test differences in a quasi-experimental design (Dimitrov & Rumrill, 2003; Fricke et al., 2017; Yeung et al., 2013), however, the impact of a larger margin for improvement and the underlying reasons for the pre-test differences cannot be fully disentangle from the analysis of the efficacy of the intervention. Random allocation could also have minimized the observed group differences in vocabulary knowledge in German in Grade 1. As lexical knowledge in German may influence performance on literacy tasks (Storch & Whitehurst, 2002), German vocabulary may represent a confounding factor for the group comparisons of the literacy measures in Grade 1. In addition, while efforts were made to create equivalent groups, schools were not allocated based on outcomes measures (i.e. PA and LSK in Luxembourgish). The allocation process of the groups could have been optimized by including a screening on the outcome measures of the sample prior to the start of the intervention and subsequent group allocation based on data from the screening (cf. Schneider et al., 2000).

Although a control measure (i.e. number naming) was included in the design to test for the specificity of the intervention effects, the study did not include an active control group to minimize the risk of potential Hawthorne effects (McCarney et al., 2007). In the context of the study, the teachers of the intervention group received a lot of teaching material and the research team (CW) was in regular contact with the intervention teachers. Hence, in comparison to the control teachers, the intervention teachers may have been additionally motivated to administer their early literacy teaching during the 12 weeks of the intervention. The control teachers had not received any material and were not in regular contact with the research team. Thus, results might be partly tainted by the psychological effect of receiving extra attention on the side of the teachers from the intervention group. To account for the additional attention effect, future work could incorporate another active control group with teachers administering for example a math intervention (McCarney et al., 2007).

Another weak point of the study was that, although extensive training of testing assistants was ensured, researchers and testers were not blind to the group allocation of the participants. This implies that a subliminal systematic tester biases could have effected the testing of the children (Kaptchuk, 2001). In addition to the researchers, the school teachers administering the intervention were also not blind for their group status for self-explanatory reasons. However, the control teachers were also aware of the study design and their role as a control group. This could be seen as a weakness as it cannot be excluded that teacher in the control group may had been induced to work harder, even unintentionally, to overcome the disadvantage of being in the control group. (Rhoads, 2011; Saretsky, 1972). Knowing that the children of their school classes were tested on PA and LSK may have led control teachers to focus more on the teaching of those two skills than they would have if they had not participated in the study. To avoid the tainting of results through Hawthorn or John Henry effects, future work could again try to replicate the current findings using an experimental active control design.

Final limitations of Study 1 are related to the post-hoc analysis on the LOL subgroup. Firstly, the analysis on the children with LOL was conducted on a sample that was not independent from the sample for the entire group (Tabachnick & Fidell, 2012). Children with LOL were included in the estimates of the effect sizes for the entire sample and hence, comparison between the effects sizes for the entire group and the LOL control group meant participants were partly compared to their own performance. Future work would need to test the efficacy of the intervention specifically on an independent sample of at-risk children (cf. Schneider et al., 2000; Stuart, 1999). The second limitation is more of a caveat than a limitation. Meaningful comparisons between the efficacy of the intervention for the entire sample and the children with LOL would presuppose that similar input would lead to similar progress in non at-risk children as in at-risk children, which appears to be, *ipso facto,* contradictory to the at-risk status. It seems unjustified to claim that a hypothetical Cohen's *d* of 0.75 for a non at-risk group compared to a hypothetical Cohen's *d* of 0.55 for an at-risk group would represent reliable evidence that the training was less effective for the non-at risk group. Such statements would presuppose that similar increases on the same metric would represent similarly developmental progress in non at-risk children and at-risk children (Protopapas et al., 2016).

Finally, while the sample size for the analysis on the entire sample provided good power (see section 3.1), the analysis on the LOW children clearly did not provide enough statistically power. Although measures of effect sizes were also used to discuss the efficacy of the intervention, a large-scale quasi-experimental or randomized control study including only L2 learners could help to provide more robust evidence on the efficacy of the intervention for at-risk children (cf. Bowyer-Crane et al., 2008; Fricke et al., 2013).

STUDY TWO: EARLY PREDICTORS OF LEARNING TO READ AND SPELL IN A SECOND LANGUAGE – A LONGITUDINAL STUDY WITH LINGUISTICALLY DIVERSE CHILDREN

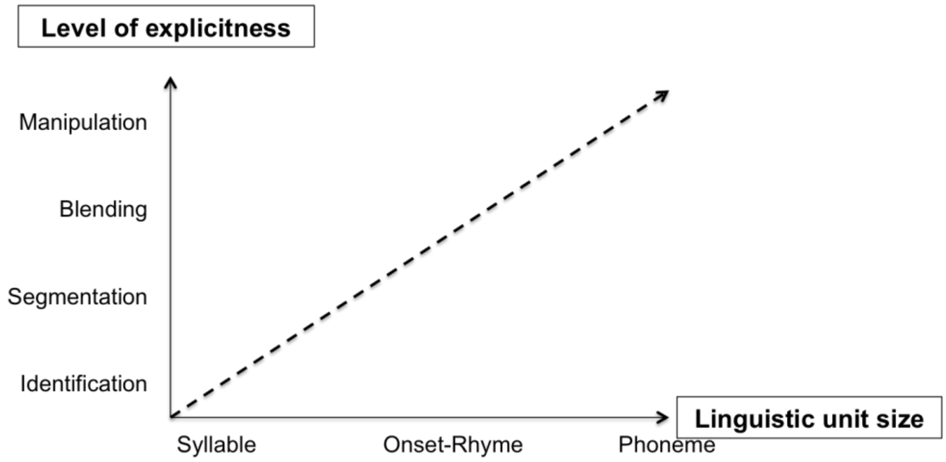
Chapter V - Literature review: Predictor study

A second aim of the thesis was to add to the understanding of L2 literacy development by examining the longitudinal cross-linguistic predictors of reading and spelling in a group of linguistically diverse children who are all learning to read in an L2. Established frameworks of early literacy development have confirmed the importance of LSK, PA, OL, rapid automatized naming (RAN) and verbal short-term memory (VSTM) in predicting learning to read and spell in monolingual and bilingual children (Lervåg & Hulme, 2009; Erdos, Genesee, Savage, & Haigh, 2011; Jongejan et al., 2007; Landerl et al., 2019; Swanson, Orosco, Lussier, Gerber, & Guzman-Orth, 2011; Lervåg & Aukrust, 2010; Melby-Lervåg, Lyster, & Hulme, 2012; Ziegler et al., 2010). In the current study, the relative importance of these five predictors in children of preschool age from Luxembourg who were learning to read in an L2 was examined. Research has shown that individual differences in literacy acquisition are already detectable before the start of formal literacy instruction and that preschool predictors could be used to identify children at-risk of reading and spelling difficulties (Landerl et al., 2013; Puolakanaho et al., 2008). This chapter reviews the literature on the relative role of LSK, PA, OL, RAN and VSTM in predicting literacy development.

5.1 Phonological awareness

PA refers to the understanding that spoken language is constituted of smaller components which can be manipulated independently from meaning (Chard & Dickson, 1999; Stahl & Murray, 1994). PA is a phonological processing skill that is primarily conceptualised along two dimension: the level of the size of the linguistic unit being processed and the level of explicitness of a phonological task (Anthony, Lonigan, Driscoll, Phillips, & Burgess, 2003; Stackhouse & Wells, 1997). The dimension of the linguistic unit being processed ranges from large units (words) to small units (phonemes). The size of the larger word-unit can be sequentially decomposed to the smallest unit of individual phonemes. For example, on a first sub-lexical level, words can be divided into syllables (e.g. bas-ket). The syllabic unit can be further broken up into the initial consonant (or consonant cluster) representing the onset (if available), the vowel (peak) and any consonants following the vowel and thus forming the rhyme unit (coda) (e.g. b-as & k-et). The onset and the rhyme can be further phonologically divided into individual speech sounds (e.g. b-a-s-k-e-t). Detecting and manipulating phonological information on the level of individual phonemes is referred to as phoneme awareness (Gillon, 2004). However, PA skills are not only distinguished by the size of the linguistic unit, but a difference is also drawn between the explicitness of the metalinguistic manipulation required to complete different PA operations. The levels of explicitness range in order of conscious access needed to perform the phonological operations, i.e. from identification (with low conscious access) through segmentation, blending, and finally to manipulation (with the highest conscious access) (Stackhouse & Wells, 1997). The two dimensions of PA also help to understand its development pattern. Larger sub-lexical units generally proceed development of awareness of phonemes and children are able to perform more implicit operations (identification tasks) before they can complete more explicit operations (blending and manipulation tasks). An overview of the developmental sequence of PA combining the level of explicitness and the linguistic unit size is presented in figure 5.1.

Figure 5.1: *Developmental sequence of PA*



Note. Adapted from Children's speech and literacy difficulties I: A psycholinguistic framework, by Stackhouse and Wells (1997). London: Whurr Publishers.

As it has been shown that preschool children often show poor phoneme awareness skills (Carroll et al., 2003) a subject of an on-going debate has been to what extent phoneme awareness emerges prior or as a consequence of early literacy instruction (Castles & Coltheart, 2004). Some authors posit that in order to learn the alphabetic principle, robust awareness of the phonemic structure of spoken language is essential to recognize that words are made up of individual sounds that are represented by letters (Burgess & Lonigan, 1998; Caravolas et al., 2012; Hulme et al., 2005; Kim et al., 2010). Conversely, other authors postulate that children's phoneme awareness remains very low until children start to learn letters. According to this view, it is the understanding that letters represent phonemes in the alphabetic writing system that ultimately catalyses sensitivity to phonemes (Bowey, 1994; Castles & Coltheart, 2004; Goswami & Bryant, 1990; Maclean, Bryant, & Bradley, 1987; Ziegler & Goswami, 2005). Experimental evidence on the controversy has shown mixed results. Some studies suggest no increase in letter knowledge after training PA (Bodé & Content, 2011; Lonigan, Purpura, Wilson, Walker, & Clancy-Menchetti, 2013; Lundberg et al., 1988), whereas, other findings suggest that pre-readers show sensitivity to phonemes, which in turn impacts their letter-knowledge acquisition (Burgess & Lonigan, 1998; Kim et al., 2010; Lerner & Lonigan, 2016; Wagner, Torgesen, & Rashotte, 1994). Similar mixed results are found for the effects of letter-knowledge training on phonemic awareness. Castles et al. (2011) report greater sensitivity to phonemes after training letter-knowledge, whereas other authors report that instruction in letter-knowledge did not boost phoneme awareness (Lonigan, Purpura, et al., 2013; Piasta, Purpura, & Wagner, 2010). All those mixed results strengthens Bowey's (2005) theory that the development of LSK and PA seem to be inextricably linked and that it would be useful to consider the two skills as co-determinants of early literacy development in alphabetic languages.

A growing body of research has investigated the development of PA in bilinguals (Cárdenas-Hagan, Carlson, & Pollard-Durodola, 2007; Chen, Xu, Nguyen, Hong, & Wang, 2010; Durgunoĝlu, Nagy, & Hancin-Bhatt, 1993; Durgunoglu & Oney, 1999; Feinauer, Hall-Kenyon, & Davison, 2013; Lindsey, Manis, & Bailey, 2003; Loizou & Stuart, 2003; López & Greenfield, 2004; Netten, Droop, & Verhoeven, 2011; Páez, Tabors, & López, 2007; Verhoeven, 2007). The question whether early speech perception and the development of PA show greater variability in children growing up bilingual have been subject of debate (Werker & Byers-Heinlein, 2008). One side of the debate argues that monolingual children ought to demonstrate more advanced PA skills due to greater exposure to a single language system as compared to bilingual children who are exposed to the phonological structure of two languages. However, the other side of the debate posits that the exposure to a second language renders structural similarities and differences between languages more salient, which allows children to attend to more abstract levels of linguistic representations resulting in more advanced PA skills (Campbell, 1995; Kuo & Anderson, 2012). However, recent studies questioned this accelerated PA of bilinguals (Bialystok, Majumder, & Martin, 2003; Chiappe, Siegel, & Wade-Woolley, 2002; Jongejan, Verhoeven, & Siegel, 2007; Lesaux & Siegel, 2003). Bialystok, Majumder and Martin (2003) compared the development of PA in English monolinguals to Spanish-English and to Chinese-English bilinguals between kindergarten and Grade 2. The authors did not find a clear and consistent bilingual advantage in the development of PA skills. In contrast, their main findings suggested that PA as part of metalinguistic awareness is rather uninfluenced by growing up exposed to one or two languages. Jongejan, Verhoeven and Siegel (2007) also suggest that the development of PA in L2 learners is not very different to PA development in L1 learners. Interestingly, the authors also concluded that PA of L2 learners does not seem to depend on OL proficiency in the L2. However, other studies seemed to show that L2 learners indeed struggle more on PA tasks in the majority language than their language majority peers (Lesaux & Siegel, 2003). Melby-Lervåg and Lervåg (2011) conducted a meta-analysis on the cross-linguistic transfer of PA in an L1 to an L2. While there were large variations between studies, data from 16 independent correlational studies (N = 1,340 children) showed that the overall correlation between L1 PA and L2 PA was strong (r = 0.57, after the removal of outliers). Overall, evidence seems to converge that PA skills in an L1 and an L2 are related, but that bilinguals do not show an advantage on PA skills through exposure to more than one language system.

5.2 Phonological awareness as a predictor of reading and spelling

Over the last decades, researchers repeatedly confirmed a unique predictive role of PA, and especially phonemic awareness, in the development of word reading, reading comprehension and spelling skills across different languages (Adams, 1990; Fricke et al., 2015; Melby-Lervåg, Lyster, & Hulme, 2012; Muter, Hulme, Snowling, & Stevenson, 2004; Rack et al., 1994; Dessemontet & De Chambrier, 2015; Stahl & Murray, 1994; Whitehurst & Lonigan, 1998). Although, it has been shown that PA is predictive of word reading in many languages with varying degrees of orthographic consistency (Caravolas et al., 2012; De Jong & van der Leij, 1999; Hulme et al., 2005; Ziegler et al., 2010), findings have shown that PA is less predictive of early word reading in more consistent writing systems (Landerl et al., 2019; Seymour et al., 2003). Especially in L1 learners in Germany, PA did not emerge as a unique predictor of word reading (Fricke et al., 2015; Landerl et al., 2019). Consistent orthographies represent the phonological structure of spoken language in a more transparent and less ambiguous way than in more inconsistent languages, which makes it phonologically less demanding for children to figure out the mappings between graphemes and phonemes (Landerl & Wimmer, 2008). This explains why PA has been found to be less predictive of later literacy skills in inconsistent orthographies.

PA has also shown to predict reading comprehension (Joshi, 2005; Muter et al., 2004; Oakhill & Cain, 2012) and early spelling (Jongejan et al., 2007). With specific reference to learning to read and write in German, Fricke et al. (2015)

showed that PA in preschool was not uniquely related to word reading and reading comprehension in Grade 1. German is considered a consistent orthography in the grapheme to phoneme direction which may explain this limited relation of PA on early reading in German. However, German is rather inconsistent in the grapheme to phoneme direction (direction of spelling), which may explain why Fricke et al. (2015) found PA to uniquely predict spelling in Grade 1.

As discussed in section 2.2, L2 children are likely to show lower vocabulary knowledge in contrast to L1 learners in the school language (Bialystok, Luk, Peets, & Yang, 2010)***.*** This lower OL skills have led to the argument that L2 learners rely more and longer on early literacy skills (e.g. PA and LSK) to read real words than L1 learners (Geva & Yaghoub Zadeh, 2006; Jongejan et al., 2007). Jongejan at al. (2007) followed L1 and L2 learners in the Netherlands and found that L1 PA was the strongest predictor of L1 word reading and L1 spelling in lower grades (grades 1-2), but L1 PA stopped being a predictor of L1 word reading and L1 spelling in higher grades (grades 3-4). In contrast, L2 PA was the strongest predictor for L2 reading and spelling in the lower grades for the L2 learners, but PA also remained the strongest predictor of L2 spelling and L2 reading in the higher grades (grades 3-4). Further interesting work on L2 learners comes from Canada. Erdos et al. (2011) showed that L1 PA in kindergarten did predict L2 pseudo-word decoding skills in Grade 1, but not L2 reading comprehension (Erdos et al., 2011) .

With reference to Luxembourg, Engel de Abreu and Gathercole (2012) cross-sectionally examined the association between L1 PA, L1 executive processes of working memory, L1 VSTM and L2 literacy measures in children at the age of eight. The results showed that L1 PA skills were specifically linked to L2 word reading and L2 spelling, but not L2 reading comprehension after controlling for Luxembourgish vocabulary knowledge.

In summary, PA skills, especially phonemic skills, have shown to be one of the best predictors of early word reading and early spelling skills in myriad languages, and this for L1 and L2 learners. Overall, PA seems to be less predictive of early decoding skills in more consistent orthographies than in more inconsistent scripts.

5.3 Letter-sound knowledge

In alphabetic scripts letters represent the phonemes of the spoken language. Letter knowledge refers to the to “children’s familiarity with letter forms, names, and corresponding sounds, as measured by recognition, production, and writing tasks” (Piasta & Wagner, 2010, p.1). There is a distinction made between letter-sound knowledge (LSK- knowledge of the letters or groups of letters which represent the individual speech sounds) and letter-name knowledge (LNK - knowledge of the names of the individual letters). In order to read, children need to be able to recognize letters (or graphemes), match the letters with their corresponding sound (phoneme) and then blend the sounds to produce a spoken response (Castles, Coltheart, Wilson, Valpied, & Wedgwood, 2009; Foulin, 2005). In this way, letters morph the basis of the alphabetic code that allows children to sound out words. It has been argued that the systematic matching of letters in print to phonemes in spoken words is the single most reliable clue for the ability to identify words at their first encounter (Phillips & Torgesen, 2006).

With regards to the development to LSK in L2 learners, Chiappe, Siegel and Wade-Woolley (2002) compared LSK in preschool between English L2 learners and English L1 learners and found no differences in LSK between the two groups. However, other studies have reported lower LSK for L2 learners than L1 Learners (Muter & Diethelm, 2001; Páez et al., 2007). For example, Verhoeven (2000) found that L2 learners (age 6-7) in the Netherlands showed lower L2 LSK in comparison to their L1 peers. Verhoeven hypothesized that L2 learners may display lower levels of auditory discrimination in Dutch compared to their L1 peers, which could impede initial mapping of sounds to letters. This theory finds support by an analysis of developmental trajectories of literacy skills of English L2 learners and English L1 learners (age 3-5) growing up in the US (Lonigan, Farver, Nakamoto, & Eppe, 2013). Lonigan and colleagues (2013) showed that almost all of the initial differences in letter knowledge between the two groups could be mediated by differences in initial English OL skills, suggesting that L2 learners are disadvantaged in learning LSK in contrast to L1 learners based on their lower L2 OL skills (Lonigan, Farver, Nakamoto, & Eppe, 2013).

Interestingly, results from Canada challenge the theory that lower levels of auditory discrimination skills of L2 learners compared to L1 learners impede the development of L2 letter knowledge. Lipka and Siegel (2007) found that L2 learners performed better on an L2 letter identification task than English-speaking L1 students in Kindergarten (age 5-6), although the difference was not significant. Yet, the authors admit that their L1 and L2 sample comprised students from different socioeconomic backgrounds, which may have confounded results. However, a few other studies reported similar results of equal performances on LSK between L1 and L2 learners (Goodrich & Lonigan, 2017; Lesaux, Rupp, & Siegel, 2007).

In all, mixed results have been reported in the development of LSK between L1 and L2 learners. However, it needs to be noted that it is difficult to generalise across studies as the educational contexts of each study are highly different. There is also evidence for L1 LSK influencing the L2 LSK, yet the extent to which LSK transfers across languages may be highly dependent on the degree of similarity of letter-sounds between languages (Erdos, Genesee, Savage, & Haigh, 2014; Goodrich & Lonigan, 2017). Thus, L2 learner that speak a language at home that is linguistically similar to the language of instruction in school may face a smaller risk of falling behind in L2 LSK than L2 learners who speak a language at home that is highly dissimilar to the school language.

5.4 Letter-sound knowledge as a predictor of reading and spelling

As letters represent the code of written language, letter knowledge lays the foundation for literacy skills as it enables children to bridge the gap between a visual cue strategy and phonetic strategy. As a consequence, failure to acquire letter-sound associations leads later literacy difficulties (Hammill, 2004; Storch & Whitehurst, 2002). LSK has been found to account for unique variance in word decoding, spelling, reading fluency, reading accuracy and comprehension in the early grades of formal schooling (Adams, 1990; Anthony et al., 2003; Bowey, 2005; Burgess & Lonigan, 1998; Hammill, 2004; Huang, Tortorelli, & Invernizzi, 2014; Lerner & Lonigan, 2016; Piasta & Wagner, 2010; Puranik, Lonigan, & Kim, 2011). Schatschneider et al. (2004) followed children from Kindergarten to Grade

2 in the US and showed that LNK tended to be initially more predictive of reading skills than LSK, yet LNK seems to ceil prior to LSK. After LNK has ceiled, LSK became a better predictor of reading skills. It needs to be mentioned that this research has been conducted in the US, and its generalizability to other countries is questionable. For example, different degrees of phonetic iconicity (i.e. letter names containing the phoneme that the letter represent) in different alphabetic systems might lead to different relationships between LNK and LSK (Erdos, Genesee, Savage, & Haigh, 2014; Goodrich & Lonigan, 2017). In addition, different cultural and educational environments may shape the prevalence of early preschool LNK. For example, research has shown that in Germany most children enter first grade with no, or hardly any LSK, whereas in the US children enter Kindergarten being already acquainted with letters (Mann & Wimmer, 2002).

The issues of universality aside, research has shown that LNK, and specifically LSK, explain substantial individual differences in later literacy acquisition in many alphabetic languages. For example, Caravolas et al. (2012) assessed LSK in L1 leaners (age 5-6) in four countries (England, Spain, Slovakia, and Czech Republic) and found LSK to be an consistent unique predictive of word reading and spelling abilities 10 months later. Fricke et al. (2015) confirmed the predictive role of LSK for German L1 learners in preschool by identifying LSK as a unique predictor of word reading, spelling and early reading comprehension in Grade 1. Leppänen et al. (2008) identified LSK at the beginning of preschool (age 5-6) as the best predictor of reading fluency and reading comprehension in Grade1 in a Finish-speaking sample.

With regards to children learning to read in an L2, Chiappe, Siegel and Wade-Woolley (2002) found English LSK to be a strong predictor of English decoding skills in first grade for both English L1 and English L2 learners. Interestingly, even more variance in decoding skills was explained by letter knowledge for the L2 learners (25%) as compared to L1 leaners (12%). Erdos et al. (2011) found that LSK in English in preschools in Canada was a unique predictor of decoding skills in French in Grade 1.

In summary, there are mixed findings on differences on performances on LSK tasks between L1 and L2 learners in the early school years. A few studies rather reveal comparable LSK skills between the two groups, whereas others have found that L2 learners underperform in comparison to their L1 peers. However, a clearer picture emerges on the predictive role of LSK in explaining individual differences in literacy skills. LSK has emerged as an important predictor of literacy abilities for L1 and L2 learners (Bellocchi, Tobia, & Bonifacci, 2017; Lonigan, Farver, et al., 2013; Melby-Lervåg & Lervåg, 2014; Muter & Diethelm, 2001; Verhoeven, 2000).

5.5 Oral language

Connected text represents the written form of spoken language. Thus, it seems self-explaining that listening comprehension is intimately associated to literacy acquisition, and particularly to the reading comprehension. OL skills, such as grammar, syntactic knowledge, narrative skills and particularly vocabulary size have all been connected to literacy skills (Duff, Reen, Plunkett, & Nation, 2015; Harlaar, Dale, & Plomin, 2007; Lervåg, Hulme, & Melby-Lervåg, 2017; Muter et al., 2004; NICHD, 2005; Scarborough, 2001; Van Viersen et al., 2018). For ease of reviewing the literature, the term OL will be employed broadly to refer to any or all of these oral language components.

Support for the importance of OL in literacy development comes with children with a family risk of dyslexia who have been shown to be more likely to develop dyslexia if they show additional OL problems (Carroll et al., 2014; Moll et al., 2016; Snowling et al., 2003). A recent meta-analysis by Snowling and Melby-Lervåg (2016) suggests that approximately 45% of the children with a family risk of dyslexia go on to become dyslexic, with estimates ranging from 29%, for a Dutch study, to 66% for an English study. However, the authors showed that children at family risk of dyslexia who have later developed dyslexia tend to consistently demonstrate lower performance on measures of language, articulation, vocabulary and grammar in early preschool years in comparison to controls without a family risk of dyslexia.

With regards to OL development of L2 learners, research suggest that simultaneous bilinguals meet the developmental milestones approximately in a similar time frame as monolinguals (Genesee, 2006). In line with monolinguals, simultaneous bilingual children also tend to produce their first words roughly at the age of 12 months. The two groups show comparable vocabulary development when total vocabulary is assessed (Hoff et al., 2012) and show similar PA development (Fabiano-Smith & Barlow, 2010). However, many of the L2 learners are not simultaneous bilinguals, but rather sequential bilinguals who are sometimes exposed to the L2 for the first time with the beginning of formal education (Hoff, 2013).

There is plenty of evidence showing that L2 learners tend to show lower levels of vocabulary knowledge in the school language (Bialystok et al., 2010; Mayo & Leseman, 2008). For example Bialystok, Luk, Peets and Yang (2010) showed that English L2 learners in Canada tend to consistently show lower vocabulary performance in the school language of the school than their L1 peers between the ages 3-10. The majority of the evidence supports that L2 learners show lower vocabulary knowledge in the language of instruction than L1 learners (Droop & Verhoeven, 2003; Melby-Lervåg & Lervåg, 2014), and there is evidence for a certain longitudinal stability of this deficit across grades (Jean & Geva, 2009). According to Melby-Lervåg and Lervåg (2014), the lower OL performances of L2 learners are unsurprising as L2 learners would need to develop their language at a much faster pace than L1 learners if they were to achieve the same level. It has been argued that when assessing total vocabulary in the two languages, L2 learners may show similar vocabulary knowledge than L1 peers (Bialystok et al., 2010). However, evidence from Luxembourg has shown that this does not seem to be true for the largest community of L2 learners in Luxembourg. Engel de Abreu, Cruz-Santos, Tourinho, Martin and Bialystok (2012) showed that children from a Portuguese-speaking home background in Luxembourg underperformed their monolingual peers from Portugal even when their conceptual knowledge was considered. Lower performances on measures of OL by L2 learners have also been reported in Luxembourg’s National Education Report (Hoffmann et al., 2018). Particularly striking is the finding that L2 learners in Luxembourg are less likely to meet the required listening comprehension standards in Grade 3 than in Grade 1, suggesting that they are falling further behind over the school years.

5.6 Oral langue as a predictor of reading and spelling

The importance of OL for reading comprehension is reflected in the framework of the SVR (Gough & Tunmer, 1986; see section 1.3), which considers reading comprehension as the product of decoding skills and listening comprehension. There is plenty of empirical evidence for the link between OL and reading comprehension (Bellocchi et al., 2017; Cain & Oakhill, 1999; Elwér, Keenan, Olson, Byrne, & Samuelsson, 2013; Lervåg & Aukrust, 2010; Oakhill & Cain, 2012; Raudszus et al., 2018). Duff et al., (2015) demonstrated that vocabulary assessed at very young age (age 1-2) explained unique variance in both later reading accuracy (11%) and reading comprehension (18%). There is evidence that OL is more predictive of reading comprehension than for early word reading. For example, even after controlling for word recognition skills, Muter et al. (2004) showed that vocabulary knowledge (age 4-5) was uniquely predictive of reading comprehension two years later (age 6-7), yet vocabulary knowledge was not uniquely predictive of word reading. The reliance on OL to read and comprehend text seems to increase as children become more proficient in their reading skills and reading comprehension becomes linguistically more demanding (Storch & Whitehurst, 2002). This has been supported by Fricke et al (2015), who found that OL skills in preschools in Germany were not uniquely predictive or reading comprehension in Grade 1, but explained unique variance in Grade 2.

However, OL has been shown to influence reading comprehension beyond a direct route via listening comprehension. OL has been linked to the development of PA and letter knowledge and may influence reading via its influence on decoding skills (Burgess & Lonigan, 1998; Carroll et al., 2003; Hipfner-Boucher et al., 2014; McDowell, Lonigan, & Goldstein, 2007; Silvén et al., 2007; Zhou et al., 2015). Kendeou, van den Broek, et al., (2009) confirmed that early OL in preschool predicted later decoding skills and that the two skills were strongly interrelated in preschool. However, in line with Storch and Whitehurst (2002), this relation became weaker with advancement from preschool to Grade 2. Storch and Whitehurst (2002) found that decoding skills were also predictive of later reading comprehension, but the predictive power of decoding skills gradually diminished over school years and OL skills accounted for more variance in reading comprehension in Grade than decoding skills 2. These findings confirm a more important predictive role of OL for higher text-level processes than for early word-level decoding processes.

It is assumed that OL skills are more important in learning to read in inconsistent orthographies than in consistent orthographies. In inconsistent orthographies, children may need to tap their lexicon to read words with inconsistent grapheme to phoneme mappings that they can only partially phonologically decode (Nation & Snowling, 2004). This view is supported by research showing a stronger link between OL and exception word reading than between OL and regular word reading. This suggests that OL knowledge contributes to early word identification beyond its mediated route through early decoding skills (Ricketts, Nation, & Bishop, 2007). However, in the consistent orthographies such as Finish (Leppänen et al., 2008), Italian (Bellocchi et al., 2017) and in German (Fricke et al., 2015;), evidence for the unique predictive role of OL on early word reading beyond PA and LSK is rather limited as phonologically decoding allows children to read most of the words.

The relationship between OL and spelling has received much less attention in the literature. Caravolas et al. (2005) found that vocabulary knowledge in English L1 learners and Czech L1 learners (age 7-11) in Grade 2 was predictive of spelling in Grade 5. However, children in Grade 5 have already advanced to more orthographic spelling and relied on their orthographic representation to read words. Early readers are more likely to apply the rule of phoneme to grapheme mapping to spell words (Ehri & Wilce, 1987), which would explain why Fricke et al. (2015) did not identify OL skills in German in preschool as unique predict L1 selling in Grade 1 and Grade 2.

With regards to L2 learners, it has been established that children who do not learn to read in their L1 mainly struggle on measures of reading comprehension due to limited L2 vocabulary knowledge (Jean & Geva, 2009; Melby-Lervåg & Lervåg, 2014; Raudszus et al., 2018). Lower L2 vocabulary knowledge is believed to impede text-level comprehension skills, which in turn may compromise the creation of new lexical entries, which again exacerbates poor reading comprehension. This negative feedback loop has been referred to as the vicious circle of reading difficulties (Verhoeven, 2011). In a study from Norway, Lervåg and Aukrust (2010) found that children with Norwegian as an L2 showed lower levels of vocabulary in Norwegian at the beginning of Grade 2, and the lower L2 vocabulary primarily explained lower reading comprehension performances in Norwegian. Many studies have reported similar results, suggesting that L2 vocabulary knowledge was lower for L2 learners than for monolingual peers and that this difference is a critical factor for poor reading comprehension abilities, even when decoding skills are controlled for (Babayiğit, 2015; Bowyer-Crane et al., 2017; Burgoyne, Kelly, Whiteley, & Spooner, 2009; Lervåg & Aukrust, 2010; Mancilla-Martinez & Lesaux, 2010; Melby-Lervåg & Lervåg, 2014; Van Viersen et al., 2018).

Less research has looked at the relationship between spelling and OL for L2 learners than L1 learners. Rolla San Francisco, Mo, Carlo, August and Snow (2006) examined the role of L1 and L2 vocabulary in spelling development in L2 learners in Grade 1 in the US and found that good L2 vocabulary is related to L2 spelling. However, Westwood (2018) argues that L2 learners, if taught correctly, can be as proficient early spellers as L1 learners across all grade levels.

In summary, OL seems critically important for reading comprehension, and deficits in OL skills in the school language have been shown to disadvantage L2 learners in comparison to their L1 peers on measures of reading comprehension. The relative predictive role of OL on early word reading skills seems to be contingent on the consistency of the orthography. OL language seems to be more predictive of early decoding and spelling skills in more inconsistent orthographies.

5.7 Rapid automatized naming

Rapid automatized naming (RAN) speed is defined as the ability to name, as quickly as possible, items on an array of highly familiar visual stimuli. Typical visual stimuli used in RAN tasks consist of digits, letters, colours, or objects (Denckla & Rudel, 1976). There are at least two types of RAN tasks: discrete RAN and serial RAN tasks. In the discrete format, the visual stimuli are presented one by one on a computer screen and the naming latency is measured for each item and averaged across items. The dependent variable is the mean variable latency over all stimuli (De Jong, 2011). In the serial format, participants are presented a row or column of stimuli and the participants must name them sequentially as quickly as possible. The dependent variable is the total time needed to name all items, or the number of items named within a fixed amount of time (e.g. in 30 seconds).

There remains a considerable lack of clarity about what RAN tasks really measure and what renders rapid automatized naming speed so essential for reading development. Lervåg and Hulme (2009) argue that RAN may tap neural circuits involved in object identification and naming, which also represent a critical component of the child's development of a visual word recognition system. To a certain extent, on a most rudimentary level, both reading aloud and RAN can be considered naming tasks (Norton & Wolf, (2011). According to Wolf and Bowers (1999), a RAN letters task, for example, taps specific cognitive processes such as:

> (a) attentional processes to the stimulus; (b) bihemispheric, visual processes that are responsible for initial feature detection, visual discrimination, and letter and letter–pattern identification; (c) integration of visual feature and pattern information with stored orthographic representations; (d) integration of visual information with stored phonological representations; (e) access and retrieval of phonological labels; (f) activation and integration of semantic and conceptual information; and (g) motoric activation leading to articulation. (p. 418).

A detailed analysis of the anatomy of RAN and whether RAN taps more into sub-lexical correspondence fluency or word-specific phonological fluency knowledge is beyond the scope of this work (for reviews see e.g. Georgiou, Parrila, &

Papadopoulos, 2016; Papadopoulos, Spanoudis, & Georgiou, 2016; Savage et al., 2018).

A few studies suggested that bilingual children tend to perform better on RAN measures than monolinguals (Geva & Farnia, 2012; Jean & Geva, 2009; Jongejan et al., 2007; Lesaux & Siegel, 2003) and there is first evidence from neuroimaging studies suggesting a possible bilingual advantage effect in RAN tasks (Arredondo, Hu, Satterfield, & Kovelman, 2017). However, other studies indicated a slight disadvantage on RAN tasks for bilingual children compared to their L1-speaking peers in preschool (Chiappe, Siegel, & Gottardo, 2002). Yet evidence seems to converge that speaking a second language does not result in any differential performances on RAN tasks between L1 and L2 learners, and that possible differences between the two groups seem to disappear within the first years of formal education (Chiappe, Siegel, & Gottardo, 2002; Chiappe, Siegel, & Wade-Woolley, 2002; Geva & Yaghoub Zadeh, 2006).

5.8 Rapid automatized naming as a predictor of reading and spelling

RAN, and especially serial alphanumeric RAN, has been shown to be one of the best predictors of early literacy in consistent languages, such as Dutch (De Jong, 2011), Finish (Torppa et al., 2013), German (Landerl & Wimmer, 2008; Moll, Fussenegger, Willburger, & Landerl, 2009), Greek (Georgiou, Parrila, & Liao, 2008)), and inconsistent languages such as English (Bowey, 2005; Jongejan et al., 2007) and French (Savage et al., 2018). This led to claims of RAN being a potentially universal predictor of reading fluency (Araújo, Reis, Petersson, & Faísca, 2015). Landerl et al. (2018) investigated the robustness of RAN as a universal predictor of reading in a large scale longitudinal study in five writing systems (English, French, German, Dutch, and Greek). Their findings showed that RAN was a more stable predictor of reading fluency across languages than for example PA. This led the authors to conclude that RAN probably taps a cognitive process underlying reading fluency across all alphabetic orthographies. However, RAN has also been shown to play a predictive factor in non-alphabetic languages, such as Chinese (Araújo et al., 2015; Georgiou et al., 2008; Tan, Spinks, Eden, Perfetti, & Siok, 2005). Fewer studies have investigated the predictive role of RAN on reading comprehension. Some studies provided support for RAN predicting performance on reading comprehension measures in English-speaking children (age 7-9) (Manis, Seidenberg, & Doi, 1999) and German-speaking children (age 5-8) (Fricke et al., 2015), whereas other studies have reported a more limited unique contribution by RAN to reading comprehension (Savage et al., 2005). With regards to spelling, the evidence is more conclusive. RAN has repeatedly be shown to be a unique contribution to early spelling in English, (Scarborough, 1998), in Spanish, Slovak, and Czech (Caravolas et al., 2012) and German (Fricke et al., 2015; Landerl & Wimmer, 2008).

A recent meta-analysis on the relationship between reading and an overall RAN score composed of different stimulus type RAN measures (i.e., letters, digits, objects and colours) presented moderate meta-correlations between RAN and word reading. RAN significantly correlated with word reading (r = .45), non-word reading (r =.40) and reading comprehension (r = .39) (Araújo et al., 2015). The meta review also indicated that alphanumeric RAN tasks often presented a stronger predictive role in predicting later literacy outcomes than non-alphanumeric RAN tasks. However, Lervåg and Hulme (2009) found that variations in non-alphanumeric RAN (objects or colours) can also be strong independent predictors of subsequent variations in text-reading fluency. The magnitude of the relationship between RAN and reading proficiency appears to depend on the age of the participants (Araújo et al., 2015; Lervåg & Hulme, 2009). It is often argued that RAN may be particularly useful as an early diagnostic measure that is predictive of later reading difficulties, as it reliably predicts the later growth of reading fluency (Lervåg & Hulme, 2009).

While the effect of speaking more than one language on RAN tasks performances remains unclear, there seems to be clearer evidence for the cross-linguistic predictive utility of RAN. Indeed, a corollary of RAN tapping underlying neural circuits necessary to read would mean that RAN shows a similar predictive role in literacy development in an L1 as in an L2 (Lervåg & Hulme, 2009; Wolf & Bowers, 1999). Jongejan et al. (2007) compared the predictive role of an object naming RAN task between English L1 and English L2 learners in Canada (age 6-10) and showed that PA and RAN were the only two significant predictors of L2 word reading ability for the L2 learners. Together the two predictors explained 40% of the variance in word reading in the group of L2 learners. Similar findings emerged from a study with English speaking children in French immersion students in Canada, who are learning to read in French (age 4-6) (Erdos et al., 2011). The authors showed that preschool RAN in English was significantly predictive of reading comprehension in French at the end of Grade 1.

In all, RAN has been found to be strong predictor of literacy skills, particularly reading fluency, in various languages and its predictive role can also be found in L2 learners.

5.9 Verbal short-term memory

Working memory has been conceptualised as “a system for the temporary holding and manipulation of information during the performance of a range of cognitive tasks such as comprehension, learning, and reasoning” (Baddeley, 1986, p. 34). One of the most influential models in the field has been Baddeley and Hitch's (1974) multi-component model, which was subsequently revised by Baddeley (2000). According to this revised model, the working memory system is composed of four subcomponents: an attentional control system (i.e. the central executive), two storage buffers responsible for the short-term storage of verbal (i.e. phonological loop) and visuo-spatial information (i.e. the visuo-spatial sketchpad), and a, more recently added, fourth component (i.e. the episodic buffer) mainly responsible for linking information across domains (Baddeley, 2000). This multicomponent model of working memory has received empirical support from experimental and cognitive in adults (Baddeley, Allen, & Hitch, 2010; Baddeley & Wilson, 2002; Smith & Jonides, 1997), as well as in children (Alloway, Gathercole, Willis, & Adams, 2004; Engel de Abreu, Conway, & Gathercole, 2010; Gathercole, Pickering, Ambridge, & Wearing, 2004; Swanson, 2008).

One of the most researched components of the model is the phonological loop (Melby-Lervåg et al., 2012), also referred to as verbal short-term memory (VSTM). It is responsible for the short-term storage of predominantly phonological information. In their seminal article, Baddeley, Gathercole and Papagno (1998) refer to the phonological loop as a “language learning device”. The authors argue that VSTM plays a crucial role in language learning by temporarily storing unfamiliar sound patterns while more permanent representations of the phonological structure of new words are constructed (Baddeley et al., 1998). VSTM has also been shown to play a central role in literacy development (Brandenburg et al., 2015; De Jong, 1998; Gathercole, Alloway, Willis, & Adams, 2006; Peng et al., 2018; Savage, Lavers, & Pillay, 2007).

VSTM is traditionally assessed by so-called simple span tasks, such as digit span, that require the maintenance of verbal information without the explicit need to manipulate information. Together with RAN, VSTM tasks are also often referred to as implicit phonological tasks because they require access to phonological codes without the need to explicitly reflect on the sound structure of words (Melby-Lervåg et al., 2012). They stand in contrast to measures of PA that require awareness of or explicit reflection on the sound structure of spoken words.

As for many other cognitive processes, research examined whether learning more than one language positively effects VSTM capacity (Bialystok, Craik, Green, & Gollan, 2009; Calvo, Ibáñez, & García, 2016; Engel de Abreu, Cruz-Santos, Tourinho, Martin, & Bialystok, 2012; Harrison et al., 2016; Jongejan, Verhoeven, & Siegel, 2007; Kieffer & Vukovic, 2013; Lesaux, Rupp, & Siegel, 2007; Lesaux & Siegel, 2003; Mancilla-Martinez & Lesaux, 2010; Paap, Johnson, & Sawi, 2015). Most studies report similar performances on measures tapping VSTM between monolingual and bilingual children. The general consensus seems to be that a bilingual advantage effect on VSTM is probably minimal, and if anything it might be task specific.

An important issue in relation to assessing VSTM capacity in multilingual contexts revolves around the reliability of the measures. Verbal short-term memory such as digit span or word span tasks contain verbal material. In children with limited proficiency in the task language it is therefore often difficult to determine whether low task performance on a verbal simple span task reflect poor VSTM capacity or may be a consequence of poor L2 oral language skills. Indeed, it is not uncommon that L2 learners present lower performance on L2-VSTM tasks than their monolingual peers (Jongejan et al., 2007; Lesaux et al., 2007). It is clearly established that it is desirable to assess VSTM in young bilingual child's L1 (Baddeley & Logie, 1999; Da Fontoura & Siegel, 1995; Yeong & Rickard Liow, 2011).

5.10 Verbal short-term memory as a predictor of reading and spelling

VSTM is thought to support learning to read by storing of the sounds of the decoded graphemes whilst decoding the subsequent graphemes to build a complete representation of the word. In terms of later text-level reading comprehension, VSTM stores previously read content and combines it with novel information to create meaning (Van den Broek, Mouw, & Kraal, 2016). Significant relationships between VSTM and word reading, reading comprehension and spelling emerged across a range of studies. For instance, the NELP (2008) reported that correlation between VSTM and decoding (r = .26), spelling skills (r =.31) and reading comprehension abilities (r =.39). A recent meta-analysis Peng et al. (2018) replicated the results of the NELP report and indicated a significant relationship between VSTM and reading comprehension (r = .31) and decoding (r = .28).

An on-going debate in the field of literacy development has been whether the contributions of VSTM to literacy are specific or shared with other measures of phonological processing. According to one view, there is a causal connection between VSTM and literacy that is separate from other aspects of phonological processing, such as PA and also of general language ability (Gathercole & Baddeley, 1993). Others have instead argued that variations in VSTM are not directly linked to literacy development once phonological processing skills are taken into account. According to this position, VSTM capacity in itself does not contribute much to literacy development (McDougall, Hulme, Ellis, & Monk, 1994). This view was supported by a meta-review by (Melby-Lervåg, Lyster and Hulme (2012), who showed that the relationships between VSTM and reading is explained in terms of shared variance with other phonological processing skills.

Yeong and Rickard Liow (2011), compared the cross-linguistic predictive role of VSTM on English spelling between children with English L1 learners and L2 learners. The authors found that VSTM tasks were only predictive of English spelling when the assessment was conducted in the L2 learners' L1. Yet, assessing VSTM in the first language of a multilingual child is often not possible for practical reasons. In this case it has been argued that verbal span tasks should be employed for which the verbal material is most likely to be equally familiar to multilinguals and monolinguals. Digit span has been shown to not place heavy linguistic demands on L2 learners and provide a more robust estimates of VSTM than sentence span tasks in young L2 learners (Chiappe, Siegel, & Wade-Woolley, 2002)***.***

There is evidence that preschool VSTM is linked to L2 word reading and reading comprehension development in later grades (Gholamain & Geva, 1999; Swanson, Orosco, & Lussier, 2015; Lesaux et al., 2007)***.*** Swanson and colleagues have conducted several large scaled studies exploring the role of VSTM in L2 literacy acquisition. Repeatedly the authors identified VSTM as an independent contributor to individual differences in literacy skills in bilinguals beyond PA (Swanson, Orosco, & Lussier, 2015; Swanson, Orosco, Lussier, Gerber, & Guzman-Orth, 2011; Swanson, Sáez, Gerber, & Leafstedt, 2004). Other work points, however towards a more indirect relative role of VSTM to L2 reading measures and L2 spelling (Harrison et al., 2016).

With specific reference to Luxembourg, Engel de Abreu and Gathercole (2012) analysed specific links among executive processes of working memory, VSTM and early L2 literacy measures in Luxembourg in Children learning to read in an L2. The results showed that VSTM in Luxembourgish made indirect contributions to word decoding and reading comprehension, but was directly linked to spelling.

5.11 Summary on predictors

Although a few studies have reported similar performances between L1 and L2 learners on early literacy (i.e. LSK and PA) skills (Goodrich & Lonigan, 2017; Lesaux, Rupp, & Siegel, 2007), the majority of the literature seems to support that L2 learners face a greater risk of underperforming on early literacy skills in comparison to L1 learners (Muter & Diethelm, 2001; Páez et al., 2007). This is unsurprising as L2 learners have to follow school instruction in a language they have to yet to fully acquire and might miss out on some content during regular class teaching (August, Carlo, Dressler, & Snow, 2005). There seems to be conclusive evidence that L2 learners perform lower on L2 vocabulary knowledge than their L1 peers in early school years (Melby-Lervåg & Lervåg, 2014), and that this deficit is stable over the grades (Bialystok et al., 2010). However, according to Melby-Lervåg and Lervåg (2014), this deficit is unsurprising as L2 learners would need to develop their language at a much faster pace than L1 learners if they were to achieve the same level. VSTM and RAN seem to be rather language independent skills and no clear differences in performances differences between L1 and L2 learners could be identified (Calvo et al., 2016; Farnia & Geva, 2013). However, an important issue in research on L2 populations seems to revolve around the reliability of the measures. Many tests tapping cognitive and linguistic skills contain verbal material. In children with limited proficiency in the task language, it is therefore often difficult to determine whether low task performance reflects less proficient capacity in the assessed skills, or may be the consequence of poor L2 oral language skills (Chiappe, Siegel, & Wade-Woolley, 2002).

With regards to the predicting literacy, a pattern of similar importance of the predictors for children acquiring literacy in an L1 or in an L2 has emerged. There is some evidence that L2 learners rely longer on early literacy skills (Jongejan et al., 2007), however, limited work on L2 learners in multilingual settings has suggested that individual differences L1 and L2 literacy skills are explained by the same predictor variables.

5.12 The current study

A wealth of research has established that PA, RAN, LSK, vocabulary knowledge and VSTM contribute to literacy development. However, most of the available studies have been conducted with monolingual children learning to read in English. Less is known about literacy development in children who are learning to read in an L2. The current study addresses this gap by longitudinally following linguistically diverse children from preschool to Grade 1 in Luxembourg. All the children in the sample learn to read and spell in an L2. The results contribute to the theoretical cross-linguistic understanding of L2 literacy acquisition and how linguistic and literacy skills at early stages of development are connected. The results have important practical implications in the context of Luxembourg. The exploration of reliable predictors of individual differences in reading and spelling acquisition would allow for the early identification of children at-risk of experiencing later reading difficulties, which in turn would enable the provision of early targeted interventions (Scanlon, Gelzheiser, Vellutino, Schatschneider, & Sweeney, 2008).

The two specific research questions of the current study were as follows:

1. What are the cross-sectional relationships between PA, LSK, vocabulary knowledge, RAN and VSTM in Luxembourgish in preschool, and between word reading, reading comprehension and spelling in German in Grade 1?

2. What is the relative importance of PA, LSK, vocabulary knowledge, RAN and VSTM in learning to read and write in a group of multilingual children who were all learning to read in the L2 German?

Chapter VI – Methods: Predictor study

6.1 Participants

Participants were a subsample of the control group (*N* = 100) from the early literacy intervention study (Study 1). Children came from four small to medium sized public preschools (14 classrooms) from rural areas (i.e. village schools) in the Central and the North regions of Luxembourg, with an average of 25 children per school (range 12-36 children per school). Children were followed longitudinally from Year 2 of preschool to Grade 1 of primary school. Notably, two children who spoke German (the language of literacy instruction in Luxembourg) as a first language at home were excluded. Nine children could not be followed-up after the transition from preschool to Grade 1 because they were either retained in preschool (*n* = 2) or they had moved to a school district for which no study consent had been obtained (*n* = 7). All children in the sample acquired German as a second language. In total, full data for 89 children were available and analysed .

Out of the 89 children, 50 were male (56%) and age ranged between 5;9 – 7;4 (mean 6;3) in preschool, and 6;4 – 7;11 (mean 6;10) in Grade 1. Background information were gathered via caregiver questionnaires (see section 3.1.2). Participants came from a broad range of socio-economic backgrounds. The average SES (ISEI-08) score was 55.89 (*SD* = 21.98) and ranged from 14 (e.g. agricultural farmers, cleaners) to 89 (e.g. medical doctors, managers). Children were classified as Luxembourgish language learners if they predominantly spoke another language than Luxembourgish at home. Thirty-one (35%) of the 89 children in the sample were Luxembourgish language learners. An overview of the main home languages spoken by the sample is presented in table 6.1.

Table 6.1: *Overview of the main home languages spoken*

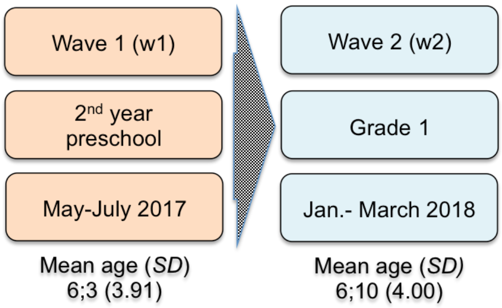
N = 89	
Language	n (%)
1. Luxembourgish	58 (65%)
2. Portuguese	15 (17%)
3. French	7 (8%)
4. Polish	2 (2%)
5. Dutch	1 (1%)
6. Bosnian	1 (1%)
7. Russian	1 (1%)
8. Serbo-Croatian	1 (1%)
9. Slovakian	1 (1%)
10. Chinese	1 (1%)
11. Créole	1 (1%)

Luxembourgish was the language spoken most frequently at home by the children (55%), followed by Portuguese (17%) and French (8%). In total, 11 different languages were spoken in the home context of the participants. According to national school data, this is broadly representative of the wider pupil population in Luxembourg. Luxembourgish-speaking pupils represent 56%, Portuguese-speaking pupils represent 19.9% and the French-speaking pupils represent 6.9% of the pupil population in Luxembourg preschool (MENFP, 2015).

6.2 Study design

Children were followed longitudinally and assessed at the end of preschool (*w*1) and in Grade 1 (*w*2), with nine months between the two waves. An overview of the study design is presented in figure 6.1.

Figure 6.1*: Timeline of the study design showing the two test points, grades, mean ages and SDs (in months)*



6.3 Measures

Children completed an exhaustive battery of cognitive, linguistic and literacy measures. An overview of the measures with cross-references to the specific task descriptions in Chapter 2 is presented in table 6.2. Two additional tasks – RAN and digit span – were administered for the purpose of the present analyses and are described in details below.

Table 6.2: *Overview of measures with cross-references to the task description*

	Preschool (*w*1)	Grade 1 (*w*2)
PRESCHOOL MEASURES LUXEMBOURGISH		
Large-unit PA measures		
Rhyme detection	see section 3.5.1	
Onset–rhyme blending	see section 3.5.1	
Onset identification	see section 3.5.1	
Small-unit PA measures		
Phoneme blending	see section 3.5.1	
Onset manipulation	see section 3.5.1	
Phoneme segmenting	see section 3.5.1	
LSK	see section 3.5.1	
Vocabulary (PPVT)	see section 3.5.1	
RAN (objects)	described below	
VSTM (digit recall)	described below	
GRADE 1 MEASURES GERMAN		
Word reading SLRT		see section 3.6.2
Non–word reading SLRT		see section 3.6.2
Reading comprehension ELFE		see section 3.6.2
Spelling HSP		see section 3.6.2

Note. (),PPVT, Peabody Picture Vocabulary Test; SLRT II, Salzburger Lese- und Rechtschreibtests [Salzburg reading and Orthography Test]; ELFE 1– 6, Ein Leseverständnistest für Erst-bis Sechstklässler [ELFE 1– 6: A reading comprehension test for students in Grades 1 through 6]; HSP, Hamburger Schreib-Probe [Hamburg spelling-task].

6.3.1 Rapid automatized naming

RAN objects (w1). An adapted Luxembourgish version of the RAN object task developed by Fricke et al., (2015) for German was used. Pictures of five objects were presented in a pseudo-random sequence over six rows of nine objects each. Each object appeared 10 to 11 times. Children were asked to sequentially name as many objects as they could in 30 seconds. The five objects represented high frequency words of the same total vowel length as in the original task (Fricke et al.

(2015). The RAN objects were: <Auto> /ɑʊtoː/ (car), <Kou> /kəʊ/ (cow), <Bam> /baːm/ (tree), <Schéier> /ˈʃɜɪɐ / (scissors), <Haus> /hɑʊs/ (house). The pictures of the objects were easy differentiable by colour. The test was only administered if a child could correctly name all of the objects on the practice sheet. Correct answers received a score of one, and the possible maximum score on the test was 54. Auto-corrections were scored as correct. Cronbach's alpha on the study data indicated an excellent internal consistency (α = .94).

6.3.2 Verbal short-term memory

Digit recall (w1). Children completed a computerized Luxembourgish version of the digit recall subtest of the Automated Working Memory Assessment (AWMA, Alloway, 2007; Engel de Abreu, Conway, & Gathercole, 2010). Children were asked to immediate recall sequences of spoken digits, in the same order as they were presented. The test contained nine blocks of six items each. The span of digits increased progressively in each block. In the first block, one digit is presented, leading up to nine digits in block nine. If four items in one block are recalled correctly the examiner should automatically move to the next block of test items. The test was discontinued after three non-consecutive mistakes within one block. An answer is considered correct if the all digits are recalled in the correct order. Correct responses received a score of one and incorrect answers a score of zero, with a possible maximum score of 54. Cronbach's alpha on the study data indicated a good internal consistency (α = .89).

6.4 Testing procedure

The first wave of the data was gathered at the end of Year 2 of preschool. The second assessment wave was administered approximately nine months later in Grade 1, after about five months of instruction in reading and spelling in German. Children were tested individually. The exact testing procedures are described under section 3.2.

Chapter VII - Results and discussion: Predictor study

The aims of Study 2 were twofold. Firstly, to cross-sectionally explore the relationships between PA, LSK, vocabulary knowledge, RAN and VSTM in Luxembourgish in preschool, as well as between word reading, reading comprehension and spelling in German in Grade 1 in a population of linguistic diverse children in Luxembourg. The second aim was to examine the longitudinal predictors of reading and spelling in this group of multilingual children who were all learning to read in the L2 German. To answer these two research questions, data from two study waves (one at the end of preschool and one in middle Grade 1, approximately nine months apart) were analysed cross-sectionally and longitudinally. Children in Luxembourg are formally introduced to literacy at the beginning of Grade 1 in German. By the time of the assessment in Grade 1, children had been following literacy instruction for five months.

The chapter is split into four sections. Firstly, the descriptive statistics of the measures administered in preschool (i.e. potential predictors) and the Grade 1 reading and spelling measures are presented. In the second section, the cross-sectional analyses of relationships between the potential predictor tasks at preschool as well as the literacy measures in Grade 1 are considered. The third section presents the longitudinal analyses on the relative importance of the potential predictors in learning to read and write. The chapter concludes with a discussion of the major findings in relation to the research aims and the literature.

7.1 Performance on measures in preschool and Grade 1

Mean raw scores, standard deviations and range for measures administered in preschool and the Grade 1 are presented in table 7.1.

Table 7.1*: Mean raw scores (SD) and range of potential predictors in preschool and literacy measures in Grade 1*

	Preschool		Grade1	
	Mean (*SD*)	Range	Mean (*SD*)	Range
PRESCHOOL MEASURES LUXEMBOURGISH				
Large-unit PA measures				
Rhyme detection (12)	9.29 (3.17)	0 - 12		
Onset–rhyme blending (12)	6.86 (3.94)	0 - 12		
Onset identification (12)	6.17 (3.38)	0 - 12		
Small-unit PA measures				
Phoneme blending (12)	2.93 (3.38)	0 - 12		
Onset manipulation (12)	1.46_F (2.98)	0 - 12		
Phoneme segmenting (12)	1.71_F (2.75)	0 - 11		
LSK (20)	10.01 (6.58)	0 - 20		
Vocabulary (PPVT) (40)	31.77 (5.04)	17 - 39		
RAN (objects) (54)	23.84 (6.74)	9 - 42		
VSTM (digit recall) (54)	21.16 (4.28)	1 - 34		
GRADE 1 MEASURES GERMAN				
Word reading SLRT (72)			9.28 (9.69)	0 - 69
Non–word reading SLRT (72)			14.92 (8.04)	0 - 52
Reading comprehension ELFE (72)			11.74 (5.61)	4 - 40
Spelling HSP (40)			32.26 (7.54)	0 - 38

Note. (), maximum raw scores; PPVT, Peabody Picture Vocabulary Test; SLRT II, Salzburger Lese- und Rechtschreibtests [Salzburg reading and Orthography Test]; ELFE 1– 6, Ein Leseverständnistest für Erst-bis Sechstklässler [ELFE 1– 6: A reading comprehension test for students in Grades 1 through 6]; HSP, Hamburger Schreib-Probe [Hamburg spelling-task].

F = floor effect.

Mean raw scores on the large-unit PA variables showed that children scored higher on rhyme detection than on onset-rhyme blending and onset-identification. An ANOVA showed that the differences between the large-unit PA measures were statistically significant, $F(2, 264) = 18.61$, $p > 001$, $\eta^2 = .12$. Post-hoc comparisons with Bonferroni corrections showed that the mean for rhyme detection ($M = 9.29$, $SD = 3.17$) was significantly higher than the mean for onset-rhyme blending ($M = 6.86$, $SD = 3.94$) and for onset identification ($M = 6.17$, $SD = 3.38$). However, onset rhyme did not significantly differ from onset identification. Concerning the small-unit PA measures, the onset manipulation and phoneme segmenting presented floor effects. The mean raw score on the phoneme blending task was slighter higher than for onset manipulation and phoneme segmenting, but still 27 out of 89 (30%) children scored zero and 43 out of 89 (48%) children scored the two lowest scores. As performances on the small-unit PA measures showed a floor effect (onset manipulation and phoneme segmenting), or a performances approaching a floor effect (48% of the sample scored 0 or 1 on the phoneme blending task), the decision was taken to not include the small-unit PA measures in the further statistical analyses. No meaningful conclusions could be drawn from the analyses. Children knew 10 letters on average at the end of preschool even without formal explicit introduction to literacy. However, the standard deviation and the range for the LSK task were large, suggesting great differences in LSK between the children. While no scores of zero were observed for vocabulary knowledge, RAN and VSTM, a few children scored zero on the code-related early literacy skills in preschool, i.e. PA and LSK.

With regards to the literacy measures, children scored significantly lower on the word reading tasks than on the non-word reading, $t(88) = 10.97$, $p > .000$. In addition, the word reading task showed a standard deviation that was higher than the mean and also a large range of scores, suggesting substantial variability in performance. In relation to their means, the reading comprehension and spelling tasks showed smaller standard deviations than the word and non-word reading tasks. Frequency tables indicated that four children scored zero on the word reading SLRT task, three children scored zero on the non-word SLRT task and two children scored zero on the spelling task. No child scored zero on the reading comprehension task, which may be explained by task structure because it allowed for guessing[7], which was not possible for the other literacy tasks.

[7] In the reading comprehension task, children had to select and underline the printed word (out of four options) that matched the adjacent picture.

7.2 Concurrent correlations between potential preschool predictors

To examine the relationships between the preschool measures, concurrent zero-order correlations (Pearson's *r*) were computed and are presented in table 7.2

Table 7.2: *Pearson's correlations between potential preschool predictor measures*

	2	3	4	5	6	7
1 Rhyme identification	.46**	.53**	.43**	.70**	.38**	.50**
2 Onset-rhyme blending		.55**	.42**	.54**	.43**	.42**
3 Onset identification			.53**	.61**	.52**	.50**
4 LSK				.63**	.51**	.50**
5 Vocabulary Luxembourgish					.47**	.52**
6 RAN						.57**
7 VSTM						

Note. LSK, Letter-sound knowledge; RAN, Rapid automatized naming; VSTM, verbal short tem memory.
Strength of correlations: very strong (.80-1), strong (.50-.79), moderate (.30-.49), weak (.10-.29), negligible (<.09).
**p* = .05; ***p* < .01.

Rhyme identification presented a moderate correlation with onset-rhyme blending (*r* = .46), and a strong correlation with onset identification (*r* = .53). The strongest correlation between the PA variables emerged between onset-rhyme blending and onset identification (*r* = .55). Out of the three PA tasks, onset identification showed the strongest correlation with LSK (*r* = .53). LSK was moderately correlated to rhyme identification (*r* = .43) and onset-rhyme blending (*r* = .42). With the exception for RAN (*r* = .47), vocabulary knowledge in Luxembourgish strongly correlated with all other measures with *r*s ranging from .50 for VSTM to .70 for rhyme identification. RAN moderately correlated with rhyme identification, onset-rhyme blending and vocabulary knowledge (*r*s ranging from .38 - .47), and correlated strongly with onset identification (*r* = .52), LSK (*r* = .51) and VSTM (*r* = .57). VSTM correlated moderately with onset-rhyme blending (*r* = .42), and strongly with all other variables (*r*s ranging from .50 - .57).

7.3 Concurrent correlations between literacy measures in Grade 1

An overview of the zero-order correlations (Pearson's *r*) between the literacy measures in Grade 1 (after five months of literacy instruction) is presented in table 7.3.

Table 7.3: *Pearson's correlations between literacy measures in Grade 1*

	2	3	4
1 Word reading SLRT	.81**	.67**	.74**
2 Non word reading SLR		.55**	.74**
3 Reading comprehension ELFE			.49**
4 Spelling HSP			

Note. SLRT II, Salzburger Lese- und Rechtschreibtests [Salzburg reading and Orthography Test]; ELFE 1– 6, Ein Leseverständnistest für Erst-bis Sechstklässler [ELFE 1– 6: A reading comprehension test for students in Grades 1 through 6]; HSP, Hamburger Schreib-Probe [Hamburg spelling-task].

Strength of correlations: very strong (.80-1), strong (.50-.79), moderate (.30-.49), weak (.10-.29), negligible (<.09).

*p = .05; **p < .01.

The SLRT single word reading and non-word reading, correlated very strongly with each other (*r* = .81). Both tasks also strongly correlated with reading comprehension (*r*s ranging from .55 to .67) and with spelling (*r*s of .74). Reading comprehension presented a moderate correlation with spelling (*r* = .49).

7.4 Longitudinal correlations between potential preschool predictors and literacy measures in Grade 1

Table 7.4 presents zero-order correlations (Pearson's *r*) between the preschool measures in Luxembourgish and the German literacy measures in Grade 1.

Table 7.4: *Pearson's correlations between potential predictors in Luxembourgish and Grade 1 literacy measures in German*

	Literacy measures in Grade 1			
Preschool predictors	Word reading	Non-word reading	Reading comprehension	Spelling
Rhyme identification	.57**	.44**	.35**	.57**
Onset-rhyme blending	.49**	.40**	.34**	.51**
Onset identification	.55**	.57**	.48**	.60**
LSK	.59**	.45**	.47**	.52**
Vocabulary	.50**	.46**	.40**	.58**
RAN	.48**	.43**	.45**	.40**
VSTM	.44**	.40**	.34**	.42**

Note. Strength of correlations: very strong (.80-1), strong (.50-.79), moderate (.30-.49), weak (.10-.29), negligible (<.09).
*p = .05; **p < .01.

All correlation coefficients were significant. PA variables correlated moderately to strongly with all four literacy measures (*r*s ranging from .35 - .60). The overall pattern of relationship indicates that the PA variables presented stronger relationships with the spelling than with the reading measures. Onset identification presents the strongest relation with the four literacy measures out of the three individual PA measures (*r*s ranging from .48 - .60). The rhyme identification task showed a stronger correlation with word reading than onset identification.
LSK presented strong correlations with word reading and spelling (*r*s of .59 and 52 respectively) and was moderately correlated with non-word reading and reading comprehension (*r*s of .45 and .47 respectively). Vocabulary knowledge showed a strong correlation with word reading and spelling (*r*s of .50 and .58 respectively) and moderate correlations with non-word reading and reading comprehension (*r*s of .46 and .40 respectively). RAN presented a consistent pattern of moderate correlations with all four literacy measures (*r*s ranging from .40 to .48). Overall the weakest correlations emerged between VSTM and the literacy measures: correlations ranged from .34 for reading comprehension to .44 for word reading. Overall, table 7.4 showed that, except for RAN and VSTM, all the predictors correlated strongly with spelling. In contrast, all predictors presented a moderate relationship with reading comprehension.

7.5 Predicting Grade 1 reading and spelling from preschool predictors

To explore the relative importance of the different predictors in explaining individual differences in reading and spelling in Grade 1, path analyses were performed. Due to the relatively small sample size and the lack of multiple indicators for most measured domains, path analyses with observed variables and not structural equation modelling (SEM) with latent variables were used. SEM is generally considered a large sample technique (Kline, 2015), but using only observed variables substantially reduces the complexity of the models and improves model fit indices for smaller samples. To further reduce model complexity, separate path analyses were run for the different literacy outcome measures. The same five preschool predictor variables were included in each model.

7.5.1 Information on data preparation

Multiple tasks tapped into PA skills in Luxembourgish. However, as mentioned above, it was not appropriate to use latent constructs in the longitudinal path models due to the low ratio of sample size to free parameters to estimate (Kline, 2015). Thus, rhyme identification, onset-rhyme blending and onset identification were transformed into a single linear component using principal component analysis (PCA) prior to the path analyses. PCA was used because the primary purpose was to reduce the data to a single composite, while retraining as much of the original information as possible (Field, 2013). The three individual PA variables in preschool correlated significantly to each other (rs > .46, ps < .001; see table 7.2) and were combined into a PA component prior to inclusion in the path analyses using PCA. A single component with an eigenvalue above 1.00 emerged which accounted for 68% of the variance and showed satisfactory indices (i.e. high component loadings for the three PA measures (> .80), KMO = 0.69, Bartlett's test of sphericity, $p < .001$). Out of the four literacy tasks, the word reading and the non-word reading tasks presented a very strong correlation ($r = .81$) and both measures assessed word reading. Thus, the two measures were combined into a word reading component prior to inclusion in the path analyses using PCA. A single component with eigenvalue above 1.00 emerged that accounted for 90% of the variance and showed satisfactory indices (i.e. component loadings of .950, KMO = 0.50, Bartlett's test of sphericity, $p < .001$). The reading comprehension and spelling measures were kept as single indicators.

Path analyses can be impacted by sample size, outliers, multivariate normality and missing data; and good data characteristics are particularly important for path analyses with small sample sizes (Bentler & Chou, 1987; Tabachnick & Fidell, 2012; Ullman, 2006). Inspection of normality and data transformation procedures are described under section 4.3. All analyses were run on the square root transformed reading measures and the log transformed spelling measure. Multivariate outliers in the data were examined using Mahalanobis distances. Two cases[8] exceeded the critical Mahalanobis distance of 11.07 for *p*-value of .050 and five degrees of freedom. Cases were removed listwise and the final data set consisted of 87 children. The final dataset included normally distributed variables, no missing data, no univariate and multivariate outliers and only observed indicators, or component scores of combined observed indicators (Bentler & Chou, 1987; Ullman, 2006).

The following sections present the path diagrams exploring the predictive value of PA, LSK, vocabulary, RAN and VSTM in Luxembourgish in preschool for explaining variance in word reading, reading comprehension and spelling in German in Grade 1. Path models allow the examination of the unique contribution of each predictor to the outcomes, after controlling for shared variance with the other predictors. Double headed arrows represent correlations between the predictor variables and single-headed arrows represent path coefficients (standardized regression weights) from the predictor variables to the endogenous (outcome) variable. The proportion of variance in the literacy outcome variable accounted for in the model (R^2) is presented above the endogenous variable. Goodness of fit of models was assessed by using the following four fit indices (Byrne, 2010):

[8] Mahalanobis distance of the two cases: 21.28, 16.01*.*

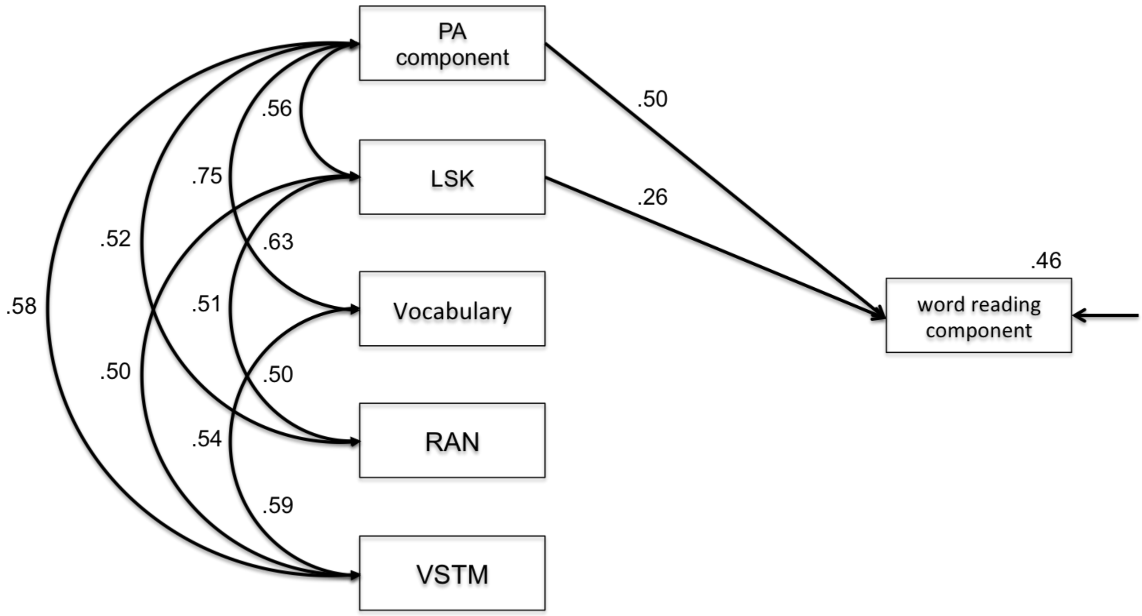
- Chi-Square statistic (χ^2): good fit = value close the degrees of freedom and nonsignificant
- CMIN/DF (χ^2 / df): < 2 fair fit and near 1 = good fit
- Comparative Fit Index (CFI): < .90 good fit and < .95 superior fit
- Root Mean Square Error of Approximation (RMSEA): < .08 acceptable fit and <.05 good fit

The following section presents the path diagrams modelling the relationships between preschool predictors and Grade 1 reading and spelling skills. In the initial model, all predictor measures were allowed to covary and all paths from the predictors to word reading were estimated. To improve model fit, nonsignificant paths were deleted successively to obtain simplified models in which all remaining paths were statistically significant (as done by Caravolas, Volín, and Hulme, 2005; Fricke, Szczerbinski, Fox-Boyer and Stackhouse, 2015; Muter, Hulme, Snowling, & Stevenson, 2004).

7.5.2 Predicting Grade 1 word reading from preschool predictors

The deletion of nonsignificant paths resulted in a reduced model for word reading with excellent fit to the data, $\chi^2(3, N = 87) = 1.72$, $p = .633$, CMIN/DF = .572, CFI = .1000, RMSEA = .000 (CI_{90} = .000 - .147) and is shown in figure 7.1.

Figure 7.1: *Path analyses model predicting word reading in Grade 1 from preschool predictor variables*



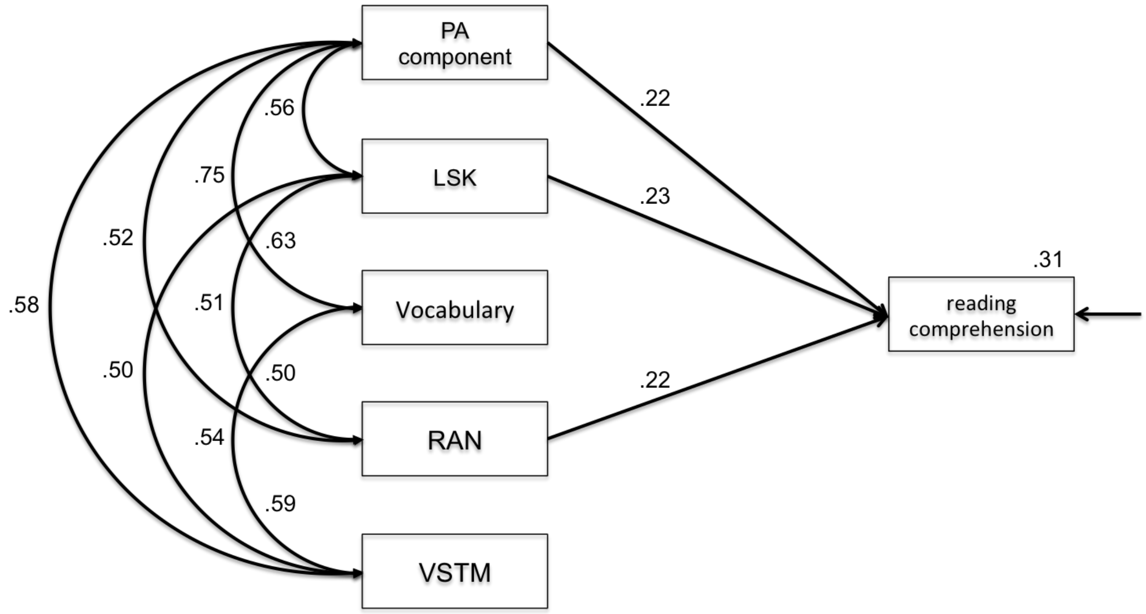
Note: $\chi^2(3, N = 87) = 1.72$, $p = .633$, CMIN/DF = .572, CFI = .1000, RMSEA = .000 (CI_{90} = .000 - .147)

In the final model, only PA ($\beta = .50$, $p < .001$) and LSK ($\beta = .26$, $p = .006$) emerged as unique predictors of word reading and explained 46% of the variance in word reading.

7.5.3 Predicting Grade 1 reading comprehension from preschool predictors

Nonsignificant paths were deleted for reading comprehension. However, PA was approaching significance (β = .22, p = .060) and was kept in the model. The final reduced model showed excellent fit to the data, $\chi^2(2, N = 87) = 0.27$, p = .874, CMIN/DF = .135, CFI = .1000, RMSEA = .000 (CI_{90} = .000 - .107), and is shown in figure 7.2.

Figure 7.2: *Path analysis model predicting reading comprehension in Grade 1 from preschool predictor variables*



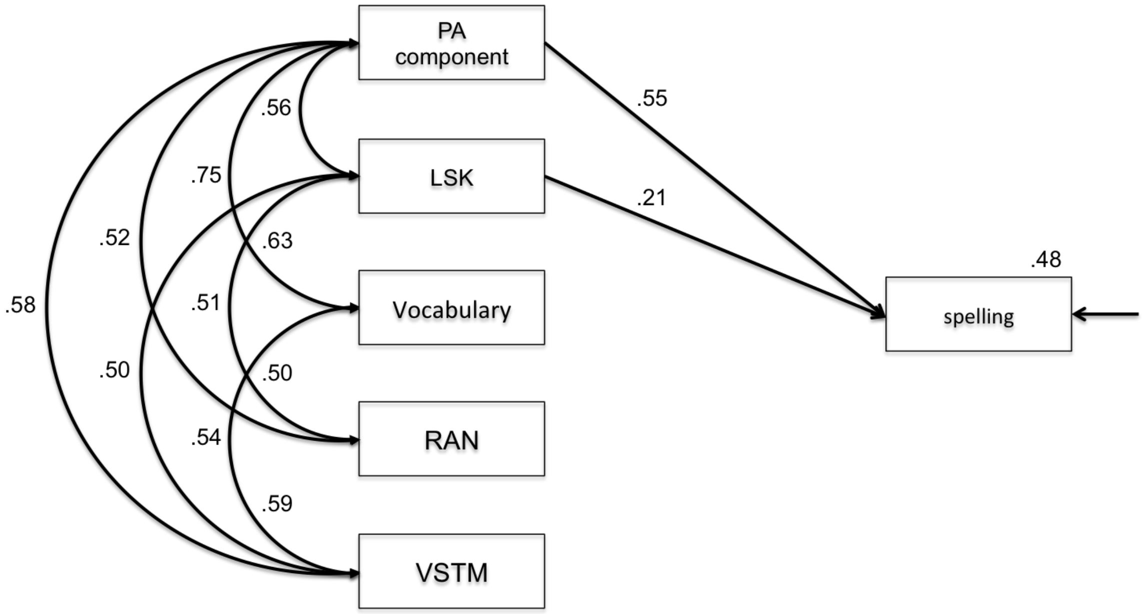
Note. $\chi^2(2, N = 87) = 0.27$, p = .874, CMIN/DF = .135, CFI = .1000, RMSEA = .000 (CI_{90} = .000 - .107)

In the final model, LSK (β = .23, p = .042), RAN (β = .22, p = .042) and PA (β = .22, p = .060) were identified as unique predictors of reading comprehension. Together the three predictors explained 31% of the variance in reading comprehension.

7.5.4 Predicting Grade 1 spelling from preschool predictors

The deletion of nonsignificant paths resulted in a reduced model for spelling with excellent fit to the data, $\chi^2(3, N = 87) = 1.26$, $p = .739$, CMIN/DF = .420, CFI = .1000, RMSEA = .000 (CI_{90} = .000 - .128), and is shown in figure 7.3.

***Figure 7.3**: Path analysis model predicting spelling in Grade 1 from preschool predictor variables*



Note. $\chi^2(3, N = 87) = 1.26$, $p = .739$, CMIN/DF = .420, CFI = .1000, RMSEA = .000 (CI_{90} = .000 - .128).

In the final model, only PA (β = .55, $p < .001$) and LSK (β = .21, $p = .026$) emerged as unique predictors and explained 48% of the variance in spelling.

7.6 Discussion

Study 2 had two aims: (1) to cross-sectionally explored the relationships between PA, LSK, vocabulary knowledge, RAN and VSTM in Luxembourgish in preschool, as well as between word reading, reading comprehension and spelling in German in Grade 1, and (2) to examine the longitudinal predictors of reading and spelling in multilingual children who were all learning to read in the L2 German.

7.7 Cross-sectional relationships between preschool measures

The first section discusses results with reference to the research question on the cross-sectional relationships between PA, LSK, vocabulary, RAN and VSTM in Luxembourgish in preschool. As all the predictor measures correlated significantly to each other only the size of the correlations is reported in the discussion.

7.7.1 Cross-sectional relationships for phonological awareness at preschool

Rhyme identification presented a moderate correlation with onset-rhyme blending and a strong correlation with onset identification. Onset identification correlated strongly with LSK, whereas rhyme identification and onset-rhyme blending correlated moderately with LSK. All three PA tasks showed a strong correlation to vocabulary knowledge. Onset identification correlated strongly with RAN, and rhyme identification and onset-rhyme blending correlated moderately with RAN. The rhyme identification and the onset identification task correlated strongly with VSTM, and onset-rhyme blending presented a moderate correlation with VSTM.

The individual PA tasks presented moderate to strong correlations to each other, which is consistent with previous studies showing strong correlations between individual PA measures (Anthony & Francis, 2005; Anthony & Lonigan, 2004; Muter, Hulme, Snowling, & Stevenson, 2004). Differences in the strength of the correlation may be explained by either the explicitness of the PA operation, the linguistic unit of the PA task, or the format of task. Rhyme identification and onset identification showed a strong correlation and the two tasks differed in the linguistic unit size, but shared the same explicitness of operation (i.e. identification) and task format (i.e. input with picture prompts). Onset identification and onset-rhyme blending required awareness of the onset, but differed on the explicitness and task format (i.e. input vs. output). Rhyme identification and onset-rhyme blending task identification presented the weakest correlation within the PA tasks, but the two tasks did not share the same level of explicitness, the linguistic unit, nor the task format.

All the PA measures significantly correlated with LSK, confirming a close link between PA and LSK (Bowey, 2005; Hulme, Snowling, Caravolas, & Carroll, 2005). Out of the three PA measures included in the current analysis, the onset identification task tapped phonological processing on the smallest linguistic unit, and also presented the strongest correlation with LSK. Small-unit PA is thought to provide the necessary insight into the phonological structure of words to understand that words are made up of individual sounds, which are represented by letters (Burgess & Lonigan, 1998; Caravolas et al., 2012; Hulme, Snowling, et al., 2005; Kim et al., 2010).This view is in line with the current findings which suggested more pronounced relationships between LSK and PA measures tapping smaller linguistic units (Lerner & Lonigan, 2016; Melby-Lervåg, Lyster, & Hulme, 2012; Muter et al., 2004; Ziegler & Goswami, 2005).

Strong correlations between the three PA measures and vocabulary emerged. Thus, previously reported intimate links between PA and vocabulary could be replicated in the current study (Dickinson, McCabe, Anastasopoulos, Peisner-Feinberg, & Poe, 2003; Silvén, Poskiparta, Niemi, & Voeten, 2007). This could be interpreted as support for Metsala and Walley's (1998) Lexical Restructuring Model. According to this model, growing early vocabulary knowledge raises the demands on PA to accurately recognize and articulate words, which in turn fine-tunes PA skills (Walley, Metsala, Victoria, & Garlock, 2003).

The data further showed that PA and RAN tasks correlated moderately to strongly. This finding was expected as PA and RAN are often thought of as subcomponents of phonological processing skills (De Jong, 2011; Norton & Wolf, 2011). Both PA and RAN measure phonological processing. RAN measures implicit phonological processing, i.e. tasks where phonological processing is automatically engaged without any awareness of or explicit reflection on the sub-lexical sound structure of spoken words. Conversely, PA is conceived as an explicit phonological processing component, which is activated in tasks that require the explicit reflection and manipulation of the sound structure of spoken words (Melby-Lervåg et al., 2012). Moderate correlations between PA and RAN have been observed in previous studies (e.g. Caravolas et al., 2012; Landerl et al., 2019).

PA measures also correlated moderately to strongly with the VSTM task. Links between PA and VSTM have been established in precious research (Melby-Lervåg et al., 2012). PA and VSTM are thought to be related because of their shared reliance on holding verbal information in VSTM. Phonological operations such as phonological blending or segmenting require the short-term storing of phonological information (Gathercole & Baddeley, 1993; Van den Broek, Mouw, & Kraal, 2016). However, two out of the three PA tasks used in the current study were input tasks (i.e. rhyme identification and onset identification) that included pictures to reduce the memory load (Schaefer et al., 2009). Only the onset-rhyme blending (output) task was not supported by pictures. Yet interestingly, the two tasks that were supported by pictures (i.e. the two input tasks: rhyme identification and onset identification) showed a stronger correlation with VSTM than the onset-rhyme blending tasks (output) that was not supported by visual stimuli. This finding, however, may be explained by the administration procedure of the two input tasks. As explained in section 3.4, the pilot study led to changes in the administration procedures of the PA input asks. In the original German TPB (Fricke & Schaefer, 2008), the test items of the input tasks are not named by the experimenter and children had to complete the tasks solely on the basis of the visual stimuli. However, this administration procedure was too difficult for the linguistically divers sample of the current study. Thus, the decision was taken to name all the pictures of the test items during the administration of the two input tasks (see section 3.4 for details on why this decision was taken). This, verbal input by the experimenter may have led to a greater reliance of VSTM to complete the input tasks and may explain the observed strong correlations between the two PA tasks and VSTM.

7.7.2 Cross-sectional relationships for letter-sound knowledge at preschool

In addition to the relationship between LSK and PA discussed above, the data also revealed a strong correlation between LSK and vocabulary knowledge in Luxembourgish, RAN and VSTM.

Previous work has suggested that low vocabulary knowledge results in lower levels of auditory discrimination in the language of letter knowledge instruction, which could impede initial mapping of sounds to letters. This would also explain the correlation between vocabulary and LSK (Lonigan, Farver, Nakamoto, & Eppe, 2013; Verhoeven, 2000). However, future research would need to examine whether knowing more letter-sounds is the direct consequence of superior vocabulary skills in Luxembourgish, or whether a higher proficiency in the instruction language allows L1 children to access the implicit holistic teaching of LSK in preschools in Luxembourg more easily.

The data revealed a strong correlation between LSK and RAN. Although, assessing LSK does not involve a time component and therefore a rapid access and retrieval of phonological labels does not seem to be required in letter learning, RAN is sometimes conceived as taxing parts of the pathway that underlie the ability to quickly learn arbitrary relations (Manis, Seidenberg, & Doi, 1999). Manis, Seidenberg and Doi (1999) argue that the speed aspect of RAN may tap a crucial component to the learning of the arbitrary relations between letter-sounds and their graphical representations. The current findings would further provide support for this view and are in line with the observed link between LSK and RAN in previous work done with monolingual and bilingual children (Lervåg, Bråten, & Hulme, 2009; Lindsey, Manis, & Bailey, 2003; Schatschneider, Fletcher, Francis, Carlson, & Foorman, 2004).

A strong relationship has been observed between LSK and VSTM in the current data, which is in line with previous research findings (De Jong & Olson, 2004). Working memory skills are thought to be a major contributor to individual differences in acquiring new knowledge and may play a key-role in storing and linking phonological with orthographic representations and to secure those links in long-term memory (Gathercole & Alloway, 2004).

7.7.3 Cross-sectional relationships for vocabulary at preschool

Beyond the relationships already discussed in the previous sections, vocabulary knowledge showed a moderate relationship with RAN and a strong correlation with VSTM.

The moderate correlation between vocabulary and RAN is interesting. Previous work has shown that RAN can usually be assessed independently of vocabulary (Geva & Yaghoub Zadeh, 2006; Harrison et al., 2016). However, the studies only used alphanumeric RAN tasks. The moderate correlation between vocabulary and RAN could be explained by the use of a RAN object task in the current study. There has been debate around the use of RAN object measures to examine rapid naming in L2 learners. Discussions revolve around the extent to which young L2 learners show appropriate lexical access and retrieval speed due to their weaker connections between phonological and semantic information than in L1 learners (Brysbaert & Duyck, 2010; Luo, Luk, & Bialystok, 2010; Van Hell & Tanner, 2012). Erdos, Genesee, Savage, and Haigh (2011) have shown that a RAN objects task, in contrast to alphanumeric RAN tasks, indeed also taps into a semantic component. Although, it was ensured that all linguistically diverse children were familiar with the RAN stimuli in the current study, retrieval of RAN objects could potentially be less automatized in L2 learners than in L1 learners (Erdos et al., 2011).

Vocabulary knowledge strongly correlated with the VSTM task. This is unsurprising as a seminal work by (Baddeley, Gathercole, & Papagno, 1998) has shown how VSTM is related to language learning. VSTM plays an important role in the long-term learning of previously unfamiliar words by allowing short-term retention before long-term representations are formed (cf. Baddeley, Gathercole, & Papagno, 1998; Masoura & Gathercole, 1999). The current study could be interpreted as support for that view. The relationship between vocabulary knowledge and VSTM in the findings is slightly stronger than observed previous studies with monolingual or pure bilingual populations (Baddeley & Logie, 1999; Da Fontoura & Siegel, 1995; Yeong & Rickard Liow, 2011).

7.7.4 Cross-sectional relationships for rapid automatized naming at preschool

Beyond the relationships between RAN in the other preschool predictors already discussed in the previous sections, RAN presented a strong correlation with VSTM.

The finding that RAN and VSTM presented a strong relationship is surprising, as RAN is thought to rather assess retrieval speed of phonological representations from long-term memory, which would not require much involvement from VSTM (Georgiou, Parrila, & Papadopoulos, 2016). While some research has suggested a limited relationship amongst VSTM and RAN (Babayiğit & Stainthorp, 2010), other studies have shown moderate relations between span and short-term memory - and RAN tasks (Erdos, Genesee, Savage, & Haigh, 2011; Georgiou, Das, & Hayward, 2008). However, to my knowledge no study has revealed a strong relationship between RAN and VSTM to date. Lesaux, Lipka and Siegel (2006) examined working memory and RAN measure in children from linguistically diverse backgrounds and only found weak relationships between VSTM and RAN measures. However, the children in Lesaux at al.'s (2006) study were in Grade 4 and there is evidence that early differences on RAN tasks wash out over time once L2 learners become more proficient in the L2 (Chiappe, Siegel, & Gottardo, 2002; Chiappe, Siegel, & Wade-Woolley, 2002; Geva & Yaghoub Zadeh, 2006). More research is needed to replicate and further explore the relationship between VSTM and RAN in children growing up in Luxembourg.

7.7.5 Cross-sectional relationships for verbal short-term memory at preschool

The relationships between VSTM and the other preschool predictors have been discussed in the previous sections (see sections 7.7.1 to 7.7.4).

7.8 Cross-sectional relationships between Grade 1 literacy measures

The relationship between spelling and reading comprehension presented only a moderate correlation. All other relationships between individual literacy measures were strong to very strong.

The very strong correlation between SLRT word reading and SLRT non-word reading is unsurprising as the two tasks come from same SLRT test and are structured and administered in the same way. Furthermore, at such an early stage in reading development (i.e. five months after starting school), the two tasks tap into the same phonological recoding skills. Research in consistent orthographies has revealed that, in the early stages of reading development, children have not yet progressed to more advanced sight word reading of real words (Ehri, 2005), but rely on the recoding and blending of all the constituent graphemes to read real words and non-words alike. There is work showing that the reliance on lower-level skills to read words may be even greater for L2 than for L1 learners (Jongejan et al., 2007). Learning to read in an L2 may slow down development of word specific representations in memory necessary for fast and efficient word recognition (Jongejan et al., 2007). Both the early word and non-word reading measures were strongly related (r = .74) to spelling, confirming that reading and spelling are highly interrelated skills (Georgiou et al., 2019; Moll et al., 2014). Reading comprehension showed the weakest correlation to all other literacy measures, which is unsurprising as reading for comprehension, in addition to lower level decoding skills, places greater demands on children’s broader OL skills than word reading or spelling tasks (Melby-Lervåg & Lervåg, 2014).

7.9 Longitudinal relationships between predictor and literacy variables

The following section discusses the results from the longitudinal analysis of the preschool predictors of Grade 1 word reading, reading comprehension and spelling abilities among children learning to read in an L2.

7.9.1 Preschool phonological awareness as a predictor of Grade 1 literacy

The longitudinal simple correlations revealed moderate to strong significant relationships between the large-unit PA component and all the literacy outcomes. In the path models, after controlling for LSK, vocabulary, RAN and VSTM, PA accounted for unique variance in word reading and spelling. The predictive role in explaining unique variances in reading comprehension was approaching significance and kept in the model as a unique predictor ($\beta = .22$, $p = .060$).

The finding that PA accounted for unique variance in word reading is line with previous work showing that PA is a reliable predictor of word reading skills in other consistent European orthographies (Caravolas et al., 2012; De Jong & van der Leij, 1999; Hulme et al., 2005; Ziegler et al., 2010). However, this finding does not concur with work from L1 learners in Germany showing that PA did not emerge as a unique predictor of word reading (Fricke et al., 2015; Landerl et al., 2019). Several factors could explain this. It has been shown that the timeframe might influence the predictive importance of PA. There is evidence that PA is more predictive of literacy measures if PA and word reading and spellings skills are both measured within a shorter period of time, e.g. less than 10 months at the very early stages of reading development (Caravolas et al., 2012). In the current study, PA was assessed at the end of preschool and literacy skills were assessed in the second term of Grade 1, approximately nine months after the assessment in preschool and roughly after five months of literacy instruction. As children have just recently been introduced to formal teaching of the alphabetic principles, it might be that children in the current study still heavily rely on initial code-related skills to decode words. It would be interesting to further follow-up the children of the current study into later Grades and continue to examine the importance of PA. Based on the literature on this topic, one would expect the role of PA to be less predictive of more advanced reading skills (Storch & Whitehurst, 2002; Vellutino, Tunmer, Jaccard, & Chen, 2007). An alternative reason for the importance of PA in learning to read in German in the context of Luxembourg is that L2 learners are likely to show lower vocabulary knowledge in contrast to L1 learners in the school language (Bialystok, Luk, Peets, & Yang, 2010). Therefore, L2 learners may rely more on lower level decoding skills (e.g. PA and LSK) to read words than L1 learners in monolingual settings (Geva & Yaghoub Zadeh, 2006; Jongejan et al., 2007).

Out of the three separate models for word reading, reading comprehension and spelling, PA explained the most unique variance in spelling. A finding that aligns with previous work in German studies showing that PA is more predictive of spelling skills than of reading abilities (Fricke et al., 2015; Landerl & Wimmer, 2008). The finding is also in line with previous work from Luxembourg showing that PA assessed in Luxembourgish was specifically linked to spelling in German (Engel de Abreu & Gathercole, 2012). In German, the encoding from phoneme to grapheme involved in spelling is more inconsistent than the rather consistent decoding from grapheme to phoneme involved in reading (Landerl & Thaler, 2006), which makes PA a better predictor of spelling than of reading.

Concerning reading comprehension, PA approached statistical significance (p = .060) as a unique predictor. Previous work has shown rather limited importance of PA on measures of reading comprehension (Fricke et al., 2015; Melby-Lervåg & Lervåg, 2014; Muter et al., 2004; Oakhill & Cain, 2012), and there is evidence for a more limited role of PA in reading comprehension compared to word reading in the current study. However, the time frame of the study may help to understand the importance of PA in reading comprehension. It has been argued that reading comprehension skills in the early stages of literacy development are still mediated by decoding skills (Cutting & Scarborough, 2006; Perfetti, Landi, & Oakhill, 2005).

Overall, in the current study, preschool PA was found to be more important for predicting individual differences in literacy measures in German in Grade 1 than in previous work (Fricke et al., 2015; Landerl et al., 2019). The strong observed predictive values of PA in the current setting are particularly surprising as PA was assessed in Luxembourgish and reading and spelling in German. This could be interpreted as providing support for the cross-linguistic nature of PA. It needs to be noted that Luxembourgish and German are linguistically similar languages, which may have facilitated any potential cross-linguistic effects of PA. However, conclusive conclusions on this matter cannot be drawn based on the design of this study. However, the current work supports the finding that preschool PA assessed in one language can be predictive of literacy acquisition in a second language in Grade 1 (Erdos et al., 2011).

7.9.2 Preschool letter-sound knowledge as a predictor of Grade 1 literacy

The longitudinal correlations confirmed that preschool LSK strongly and significantly correlated with all literacy skills. Controlling for PA, VSTM, vocabulary knowledge and RAN in preschool, LSK emerged as a unique contributor to individual differences in all three literacy measures, i.e. word reading, reading comprehension and spelling.

As letters represent the code of written language, it was unsurprising that LSK in preschool showed predictive value for all three literacy measures in Grade 1. Letter knowledge enables children to bridge the gap between a visual cue strategy and a phonological strategy (Hammill, 2004; Storch & Whitehurst, 2002), and thus, lays the foundation for reading and spelling. The findings of the current study align with a body of research showing that LSK consistently accounts for unique variance in word reading, reading comprehension and spelling in children learning to read in an L1 (Adams, 1990; Anthony, Lonigan, Driscoll, Phillips, & Burgess, 2003; Chiappe, Siegel, & Gottardo, 2002; Hammill, 2004; Huang et al., 2014; Lerner & Lonigan, 2016; Lesaux, Geva, Koda, Siegel, & Timothy, 2008; Piasta & Wagner, 2010; Puranik, Lonigan, & Kim, 2011), as well as in children learning to read in an L2 (Chiappe, Siegel, & Wade-Woolley, 2002; Erdos et al., 2011).
The 20 letter-sounds assessed in preschool in the current study are identical in Luxembourgish and in German. Hence, no conclusions about the potential cross-linguistic nature of LSK can be drawn.

7.9.3 Preschool vocabulary as a predictor of Grade 1 literacy

Vocabulary knowledge in Luxembourgish in preschool showed significant moderate to strong simple correlations to all L2 literacy measures in German in Grade 1. Yet, the path models indicated, that after controlling for PA, LSK, RAN and VSTM, vocabulary did not uniquely predict either of the three literacy skills in Grade 1.

Vocabulary did not account for unique variance in early word reading, reading comprehension, nor spelling, which is in line with a limited direct contribution of OL skills on early literacy skills in monolingual children reported in previous work (Fricke et al., 2015; Muter et al., 2004; Nation & Snowling, 2004). In addition, vocabulary knowledge is considered a language specific skill, as the labels are given to certain concepts are language specific (Goodrich & Lonigan, 2017). Hence, it would have been very surprising if vocabulary knowledge in Luxembourgish directly contributed to L2 literacy skills in German, if its direct contribution is already weak in L1 learners (Fricke et al., 2015; Lervåg, Hulme, & Melby-Lervåg, 2017; Nation & Snowling, 2004; Van Viersen et al., 2018). The significant simple correlations between vocabulary knowledge in Luxembourgish and the literacy measures in German observed in the current study rather provide support of an indirect contribution of vocabulary through PA in Luxembourgish and LSK (Burgess & Lonigan, 1998; Carroll et al., 2003; Hipfner-Boucher et al., 2014; McDowell, Lonigan, & Goldstein, 2007; Silvén et al., 2007; Zhou et al., 2015).
As OL was only assessed using a language-specific receptive vocabulary task in Luxembourgish and not in German (Goodrich & Lonigan, 2017), the current research design does not allow to draw conclusions on whether the direct contribution of OL in early stages of reading or spelling in German may be limited, or whether OL is more important for reading comprehension than for word reading or spelling (Muter et al., 2004). Children had not been introduced to German in preschool in Luxembourg and vocabulary knowledge in German was not assessed in preschool in the current study.

However, future work could look at whether less language-specific OL skills than vocabulary knowledge in preschool in Luxembourgish would contribute to unique variance in later reading comprehension measures in German. For example, inferencing and narrative skills have been found to show cross-linguistic associations (Rodina, 2017). It seems paramount that future studies include a more comprehensive assessment of linguistic competency in linguistically diverse children to conclusively replicate the weak direct relationship between OL skills in Luxembourgish and L2 literacy development in the current study.

7.9.4 Preschool rapid automatized naming as a predictor of Grade 1 literacy

The simple longitudinal correlations revealed significant moderate relationships between RAN and all the literacy outcome measures. Controlling for PA, VSTM, vocabulary knowledge and LSK in preschool, the path analyses revealed that RAN did not emerge as a unique contributor for word reading and spelling. However, RAN emerged as a significant unique predictor of reading comprehension.

RAN objects did not account for unique variance in word reading, which is not in line with previous work. RAN has been repeatedly identified as an important independent contributor of early reading in consistent orthographies (De Jong, 2011; Georgiou, Parrila, & Liao, 2008; Landerl et al., 2019; Moll, Fussenegger, Willburger, & Landerl, 2009; Torppa et al., 2013). Particularly for L1 learners acquiring literacy in German, RAN has shown to predict more unique variance in word reading than PA (Fricke et al., 2015; Landerl et al., 2019). In addition, RAN was a particular important predictor of reading measures showing a speed component, for example word and text reading fluency tasks (Moll et al., 2009; Savage et al., 2005).

There are a few possible explanations why RAN did not emerge as a predictor of word reading in the current study. Firstly, alphanumeric RAN tasks have shown to be more predictive of later literacy skills, and the use of a RAN object task in the current study may attenuate a direct contribution from RAN to word reading (Araújo et al., 2015; Lervåg & Hulme, 2009). Due to the lack of formal LSK and numerical instruction in preschools in Luxembourg and the resulting lack of automaticity of that knowledge in preschool, alphanumeric RAN scores could not be used as a meaningfully measure of access speed. Thus, no alphanumeric RAN task was administered in this study, which may have decreased the predictive power of RAN, but also reduced potential measurement error. A further explanation for the lower predictive value of RAN observed in the current study than in the other studies in German literacy contexts was the limited reliability of a single RAN measurement. Landerl et al. (2019) and Fricke et al (2015) administered at least two different RAN tasks, and created an average score (Landerl at al., 2019) or a component score (Fricke et al., 2015) of the multiple tasks. This represents a more reliable estimate of RAN than the single measure used in the current study and may explain why RAN has been shown to be an independent predictor of reading skills in German.

A third reason why RAN did not emerged as a predictor of word reading might have to do with the fact that the children in this sample were in the very early stages of learning to read and use a decoding strategy to read words (Moll et al., 2009). This is supported by the results showing the high predictive value of PA and LSK in word reading. RAN and reading are associated because they both tap into the speed with which phonological representations can be retrieved from long term memory (Wimmer, Mayringer, & Landerl, 2000) and the literature shows that RAN is a better predictor of reading fluency than of reading accuracy (e.g. Arne Lervåg & Charles Hulme, 2009; Young & Bowers, 1995). Although the current measure is considered a reading fluency measure, the children may still laboriously retrieve the links between graphemes and phonemes, which may be a modulating factor on the reading fluency score in the current sample. Additionally, it has been shown that L2 learners may rely more on a decoding strategy due to lower broader OL skills (Jongejan et al., 2007), which may again support the lack of preditive value of RAN on word reading in the current sample. It would be interesting to further follow-on the current sample into Grade 2 to explore if RAN becomes predictive once children have transitioned from an alphabetic to a faster and more efficient orthographic reading strategy (Melby-Lervåg et al., 2012).

Concerning the predictive path between RAN and reading comprehension, the results show that RAN accounted for unique variance in L2 reading comprehension. While some work has found limited evidence of RAN predicting reading comprehension (Savage et al., 2005), others found a direct link in L1 learners (Fricke et al., 2015; Manis et al., 1999), as well as L2 learners (Erdos et al., 2011). As mentioned above, RAN objects could tapping into a semantic component which would explain why RAN was uniquely predictive of reading comprehension skills but not of word reading skills. In addition, the unique contribution of RAN to reading comprehension may be explained by the structure of the reading comprehension task, because the reading comprehension tasks included a speed component (max. number of words read in three minutes). Further research, including the use of alphanumeric RAN tasks to assess this skill, would be necessary to examine the semantic component of RAN objects on L2 reading comprehension in linguistically diverse children in Luxembourg.

When it came to predicting spelling, RAN was not found to uniquely predict L2 spelling skills, which is not in line with previous work indicating a close link between RAN and early spelling in consistent orthographies (Caravolas et al., 2012; Fricke et al., 2015; Georgiou et al., 2012; Landerl et al., 2018). In contrast to the reading tasks, the spelling measure did not include a time component, so processing speed may not directly have influenced performance on the spelling task. However, similarly to word reading, the indirect contribution of RAN on early spelling after controlling for PA, LSK, VSTM and vocabulary, could reflect the fact that children in the current sample have not yet transitioned from an alphabetic spelling to an orthographic spelling. Again, it would be interesting to further follow the children into Grade 2 to explore whether preschool RAN becomes predictive of more advanced spelling.

In summary, it seems that in stage of reading development children still rely mainly on decoding strategies and the relationship between RAN and the literacy skills might be mediated by the laboriously decoding of words.

7.9.5 Preschool verbal short-term memory as a predictor of Grade 1 literacy

VSTM in preschool showed significant moderate longitudinal correlations with all Grade 1 literacy measures. However, controlling for PA, LSK, vocabulary and RAN, VSTM did not emerge as a unique contributor for word reading, reading comprehension and spelling.

While limited evidence for a direct contribution of VSTM to early reading and spelling has been reported in the literature (Farnia & Geva, 2013; Swanson, Orosco, & Lussier, 2015), most work has not associated VSTM to differences in literacy development after controlling for phonological processing (Harrison et al., 2016; McDougall, Hulme, Ellis, & Monk, 1994; Melby-Lervåg et al., 2012). There has been considerable debate whether VSTM plays a unique predictive role in individual differences in literacy development, or whether they are related to literacy acquisition via an indirect sublexical route through PA and RAN (Savage, Lavers, & Pillay, 2007; Tunmer & Hoover, 1992). The current findings support the view that VSTM does not account for additional unique variance in reading skills over and above its shared variance with PA in the case of word reading and spelling, and over and above its shared variance with PA and RAN for reading comprehension.

As mentioned above, pictures were used to reduce the memory load of the input PA tasks (i.e. rhyme identification and onset identification). However, as previously discussed, a modification to the administration procedures of the input tasks may have led to a greater reliance on VSTM to complete the PA tasks. Thus, the PA and VSTM may present greater shared variances in the current data than if PA had been measured more independently from VSTM. In addition, only a span memory task (i.e. digit recall) was used to assess VSTM abilities. While the use of digit naming in the assessment of VSTM in L2 learners has been recommended (Chiappe, Siegel, & Gottardo, 2002), adding multiple indicators tapping into the skill of actively processing information could have resulted in a more reliable VSTM measures and possibly revealed a more unique predictive role of WM in L2 literacy acquisition.

7.10 Summary of findings

The findings of Study 2 revealed that the cross-sectional simple correlations between the preschool measures were in line with what the literature suggested. The simple correlation between the input PA tasks and VSTM observed in the current data seems higher than observed in previous work, but this was probably related to the format of the input PA tasks used. The longitudinal simple correlations between the preschool predictors and the literacy measures in Grade 1 were all significant and moderate to strong in magnitude. This suggests that all the preschool predictors in Luxembourgish in the current study play a foundational role in L2 literacy development in German.

The path models confirmed that two of the most reliable predictors (i.e. PA and LSK) of learning to read and write in many European orthographies also uniquely contribute to L2 literacy acquisition in Luxembourg, after controlling for vocabulary, RAN and VSTM. The most surprising finding was that, contrary to previous findings in learning to read and write in German, RAN has not been found to account for unique variance in L2 word reading and L2 spelling in German in the linguistically diverse sample of this study (Fricke et al., 2015; Georgiou et al., 2019). However, this could be related to the fact that in the current study only a single RAN object task was used and that children were still at very early stages of reading development.

The predictors in the current study accounted for a larger percentage of total variance in the literacy skills than reported in previous work in German L1 learners (Fricke et al., 2015), which suggests that children growing up in Luxembourg heavily rely on PA in Luxembourgish and LSK to learn how to read and spell in German. Yet still, roughly half of the variance for word reading and spelling, and two thirds of the variance for reading comprehensions in Grade 1 were unaccounted for by the predictor measures in preschool. It could be that either other cognitive, early literacy skills and oral language skills in Luxembourgish or in German, or environmental and family factors would account for a large proportion of the unexplained variance.

7.10.1 Strengths, weaknesses and future directions

A major strength of Study 2 was its longitudinal research design and that data were not only collected cross-sectionally. Concerning causality, a longitudinal design is superior to a cross-sectional design as the development of literacy skills in Grade 1 could not have influenced the development of early predictors in preschool. A possible confounding factor that cannot be excluded in cross-sectional research (Tabachnick & Fidell, 2012). Another asset of the current work is that children completed an extensive PA battery taping into various linguistic units and degrees of explicitness. In some previous studies other authors have only used a single measure of PA (cf. Erdos et al., 2011; Harrison et al., 2016). In addition, the assessment of literacy skills in Grade 1 only involved standardised German measures, which increased the reliability of the measurements (Coolican, 2009). However, the normative data from the native German-speaking normed sample could not be applied to the German L2 learners in Luxembourg.

With regards to the predictor measures in preschool, no standardized tests existed in Luxembourgish. Thus, only researcher-developed assessments could be administered in preschool. However, these measures showed good psychometric properties. Although a large panoply of predictor domains were measured, with the exception of PA, all predictors were only represented by single indicators due to time constraints. For example, vocabulary knowledge, was only assessed by one receptive task and OL may become more predictive of reading comprehension if various aspects of OL had been assessed, e.g. expressive vocabulary task, narrative skills, morphological knowledge (Ouellette, 2006). Including multiple indicators per domain of interest would have also allowed for more advanced statistical analyses (e.g. latent SEM models), which would have further strengthened the reliability of the findings (Byrne, 2010; Kline, 2015). In addition to the examination of the direct contribution of the preschool predictions to literacy skills, it would have been interesting to perform a mediation analysis including PA and OL in German in Grade 1, when predicting literacy outcomes in German from predictors in Luxembourgish. However, although the presented path models included only single indicators and the models are most parsimonious for the data collected, the sample size provides limited statistical power for more complex mediation analysis. A replication study including a larger sample size would be desirable. In addition, future studies may want to also investigate other environmental and family factors influencing literacy development (for review see e.g. Duncan & Seymour, 2003; Heath et al., 2014). The study of these factors might be particular important in the context of Luxembourg due to its high cultural diversity (OECD, 2010).

Another limitation of this study was the RAN object measure used. RAN objects tasks have been shown to be less predictive of later literacy skills. An alphanumeric RAN task would have increased the predictive importance of RAN (Araújo et al., 2015). However, an alphanumeric RAN could not have been administered due a lack of robust familiarity of letters and digit names to meaningfully measure the automatic retrieval of letter and digits. Yet a potentially a RAN colours could have been added to the battery to create a more reliable measurement of RAN (cf. Fricke et al. 2015).

While the longitudinal nature of the study is an asset, an even longer follow up of the sample could have strengthened the findings further. However, due to constrictions in the time frame available for the study it was not possible to include data points beyond Grade 1 in the current work. It would be valuable to follow-up children further into Grade 2 and assess if the prediction patterns would hold once children become more proficient in reading and their strategies change from decoding strategies to a more global reading based on orthographic patterns (Ehri, 2005). A project in this regard is currently underway.

Chapter VIII: General discussion

The rapid increase in linguistic and cultural diversity in schools worldwide (Grosjean, 2010; OECD, 2018) has prompted a bulk of research on children acquiring literacy in multilingual educational settings and children learning to read and write in an L2 (Murphy & Unthiah, 2015; Oxley & de Cat, 2019). So far, the field of early literacy research had predominantly been focused on monolingual children learning to read in English (Share, 2008). The necessity to extend the scope of existing research on learning to read and write to L2 learners has been repeatedly acknowledged (August & Shanahan, 2010; Murphy & Unthiah, 2015). This thesis aimed to contribute to the understanding of the literacy development of linguistically diverse children learning to read and write in German.

The thesis was split into two studies. The first study explored the efficacy of an early literacy programme that was developed for preschool-aged linguistically diverse learners from Luxembourg. The programme was evaluated in a quasi-experimental longitudinal study in mainstream preschool settings in Luxembourg. The second part of the thesis consisted of a correlational examination of cross-sectional and longitudinal early predictors of reading and spelling in children learning to read in an L2 in Luxembourg. A key aspect of the two studies was that almost all the children followed literacy instruction in an L2. To the best of my knowledge, this was the first study to evaluate the efficacy of systematic support of PA combined with LSK in children growing up in Luxembourg, and to longitudinally explore the importance of PA, LSK, RAN, VSTM and vocabulary knowledge in Luxembourgish on L2 literacy skills in German.

The following section summarizes and discusses main findings, highlights practical implications and presents a final conclusion of the work.

8.1 Supporting early literacy skills in linguistically diverse children

Luxembourg is a culturally and linguistically heterogeneous country in Central Europe, where approximately 59% of the total pupil population comes from a non-Luxembourgish speaking background (Lenz & Heinz, 2018). National standardized student assessments in Luxembourg show that over 40% of Luxembourg's nine-year-olds do not meet the required reading standards, and this figure increases to 71% if only the Portuguese-speaking pupil population is considered (Hoffmann et al., 2018).

The importance of high-quality early education can hardly be overstated. At-risk children who have been provided targeted early support are less likely to repeat school grades, and are also less likely to require additional support in future years (Knudsen et al., 2006). Over the last decades a large body of international research has investigated how to optimally prepare children for the task of learning to read and write. The conclusion has been drawn that early support in foundational literacy skills is important and should combine explicit and systematic teaching in PA and LSK in a language-rich environment (Bus & van IJzendoorn, 1999; Ehri et al., 2001; Rose, 2006). The preschool curriculum in Luxembourg is traditionally predicated on play-based learning in a holistic approach and only includes very limited and unstructured teaching of LSK and PA (MENFP, 2011). Thus, Luxembourg may be missing an important opportunity to provide children with a solid foundation for literacy development. The intervention study of the thesis examined the efficacy of an early literacy programme in a real world context. Regular preschool teachers delivered a newly developed early literacy programme over 12 weeks to their Year 2 preschool classes (age 5-6). In addition to the overall efficacy of the intervention in supporting early literacy skills, the multilingual context of the study provided the opportunity to explore cross-linguistic facilitative effects of early literacy training in one language and literacy development in another language.

The following three major findings emerged from Study 1. Firstly, the intervention successfully supported the development of the targeted early literacy skills (i.e. PA and LSK) in Luxembourgish and the effect sizes immediately post-intervention ranged from large to huge for measures of PA and LSK. A second particularly promising finding was that children with LOL, 79% of whom were L2 learners, also benefited from the intervention. This result is practically relevant as those children often face the greatest risk of literacy difficulties and general educational underachievement (August & Shanahan, 2005; Hoffmann et al., 2018; Lesaux, Crosson, Kieffer, & Pierce, 2011). A third major finding was that the intervention effects on early literacy skills in Luxembourgish seem to have generalised to PA in German and to the development of learning to read and spell in German in Grade 1, nine months after the end of the intervention.

From a theoretical perspective, the results are in line with previous findings, showing smaller transfer effects from early literacy support to early word reading skills than to reading comprehension or spelling (Fischer & Pfost, 2015). As observed in other consistent orthographies (Caravolas et al., 2013), the current work provides support that linguistically diverse children in Luxembourg also show a steep increase in early decoding abilities within the few months of formal literacy instruction in German. The current work seems to extend the view that reliable mappings of graphemes onto phonemes in German allow children to quickly learn grapheme to phoneme correspondences, which in turn leads to more proficient decoding and encoding skills at early stages of reading development (Aro & Wimmer, 2003). It has been argued that stronger foundational literacy skills would free up cognitive resources that could then be used for higher level reading comprehension, which would explain the stronger effects on reading comprehension for the intervention group in the current study (Oakhill et al., 2015; Perfetti, 2007). In addition, it could be presumed that the Grade 1 teachers of the children of the intervention group focused more on the teaching of comprehension aspects in early months of formal literacy instruction because the children showed a solid basis in decoding skills. In a meeting with the Grade 1 teachers of the children from the intervention group explained the research team (CW & PW) that the strong foundation of early literacy skills of the children allowed the teachers to focus less on targeted teaching of decoding skills and to include broader oral language activities in their teaching. This may also have contributed to the better performance of the intervention group on the reading comprehension task.

The results from the current study conducted in Luxembourg are in line with conclusions drawn from studies in other languages, orthographies and countries (Bowyer-Crane et al., 2008; Ehri, Nunes, Stahl, et al., 2001; Fischer & Pfost, 2015; Lundberg, Frost, & Petersen, 1988; Schneider, Roth, & Ennemoser, 2000; Stuart, 1999), showing that early interventions focusing on foundational literacy skills can successfully be implemented and lead to educational improvements (Network Promising Practices, 2007; What Works Clearing House, 2007). This study further showed that it is possible for regular teachers to foster literacy development using a structured and explicit teaching approach in a play-based and holistic learning context in Luxembourg (MENFP, 2011).

The study also adds to a growing body of research examining the language specific or language independent nature of PA (Goodrich & Lonigan, 2017; Melby-Lervåg & Lervåg, 2011; Wawire & Kim, 2018). The current study is the first study to employ a quasi-experimental design showing that training PA in Luxembourgish may generalise to improvements in PA skills in German. This could be interpreted as support of the view on PA as an underlying cross-linguistic ability (Branum-Martin & Garnaat, 2015; Branum-Martin, Tao, Garnaat, Bunta, & Francis, 2012; Melby-Lervåg & Lervåg, 2011). The results of the current study clearly demonstrated that effective support of early literacy skills in Luxembourgish in preschool generalizes to measures of PA, reading comprehension and spelling in German in Grade 1.

Previous research has shown that underneath the overall efficacy of the intervention lie wide variances in response and non-responders (Duff & Clarke, 2011; Fallis, 2013). It is well-established that a small percentage of at-risk children tend to not respond to classroom-based early literacy training and that those children require targeted support in smaller groups or in individual sessions (Duff et al., 2008; McMaster, Fuchs, Fuchs, & Compton, 2005). Estimates of non-responders to phonological based early literacy intervention range from 10 to 46% (Torgesen, 2000). It is well understood that it is not a single short-term intervention that represents the magic solution for all literacy difficulties, but that on-going support is needed to remediate long-standing reading problems and long-term negative educational consequences (Duff et al., 2008). It is, however, clear that for children with a severe risk of failing to read, whole-class support alone is not intense and targeted enough (e.g. Hatcher, Hulme, & Snowling, 2004). Future research in this area could focus more specifically on those children and establish the extent to which their reading difficulties might be remediated if the whole-class support is complemented by small group teaching or individual tutoring.

As the study by Bodé and Content (2011) and the current study are the only two experimental studies that explored the efficacy of early support in foundational literacy skills, no previous work in Luxembourg has examined the effect of the children that do not respond to those phonological based interventions. High priority should be given to future work examining individual differences in children's responsiveness to early literacy interventions in Luxembourg and what other support non-responders in Luxembourg may require.

8.2 Predictors of reading and spelling in second language learners

In order to gain a better understanding of how children learn to read and write in the multilingual education context in Luxembourg, Study 2 considered the role of phonological skills (PA, RAN, VSTM), LSK and vocabulary in predicting individual differences in L2 literacy in Grade1 in the context of Luxembourg. Eighty-nine preschool children, all of whom were acquiring German as an L2, were longitudinally followed and tested in two waves: at the end of preschool and in the middle of Grade 1.

The following three major findings emerged. Firstly, in line with findings from L1 and L2 learners, the study replicated previous work showing that individual differences in literacy development could be predicted from early predictors before the formal start of literacy instruction (Duff et al., 2015; Fricke et al., 2015). This suggests that it would be possible to identify at-risk children before they begin literacy instruction and already provide support in preschool. A second major finding was that PA, RAN, VSTM, LSK and vocabulary all presented significant correlations with word reading, reading comprehension and spelling skills. These findings are in line with previous work showing the importance of these linguistic and cognitive skills in literacy acquisition (Engel de Abreu & Gathercole, 2012;

Swanson et al., 2015; Melby-Lervåg & Lervåg, 2014; Melby-Lervåg et al., 2012; Muter et al., 2004). A third major finding revealed that, after controlling for the effects of vocabulary, VSTM and RAN, LSK and PA in preschool emerged as unique contributors to individual differences in word reading, reading comprehension and spelling. RAN only emerged as unique contributor to reading comprehension. This pattern of findings seems to support the view that different phonological processing measures (PA, RAN, VSTM) tap into a common underlying ability for word reading and spelling (Melby-Lervåg et al., 2012). VSTM and RAN do not seem to account for unique variance in learning to read and spell above and beyond their shared variances with PA. RAN only contributed directly to reading comprehension, which may be partly related to the speed component of the task and the RAN object task used in the current study (Lenhard & Schneider, 2006). Future work would need to further explore the importance of RAN in literacy development in children growing up in Luxembourg using different RAN measures and different reading comprehension tasks before any reliable conclusions can be drawn.

Children in the current study acquired literacy in an L2 and most of the previous work has looked at literacy development in children learning to read and write in their L1 (e.g. Fricke et al., 2015; Leppänen, Aunola, Niemi, & Nurmi, 2008; Muter et al., 2004). However, the findings on L2 learners of the current study align with previous work on L2 literacy acquisition showing that L2 literacy development seems to rely on the same foundational skills than in an L1 learners (Erdos et al., 2011).

Some studies conducted in consistent orthographies have suggested that PA plays a universal role in explaining individual differences in word reading skills across more and less consistent alphabetic orthographies (Caravolas et al., 2005; Ziegler et al., 2010). However, other studies on consistent orthographies (Aarnoutse et al., 2005; Georgiou et al., 2012; Silvén et al., 2007) and in German particularly (Fricke et al., 2015) questioned the predictive role of PA in the development of learning to read. The current results support the view that PA is a unique predictor in learning to read in linguistically diverse children learning to read in German as an L2.

Two main reasons could explain the predictive role of PA in Luxembourgish in learning to read and spell in German. Firstly, Luxembourgish and German share a very high degree of structural and linguistic commonalties which render causal cross-linguistic facilitative effects in early literacy skills between the two languages more likely (Melby-Lervåg & Lervåg, 2011). Secondly, 98% of the pupil population follow reading instruction in a second language in Grade 1 (MENJE, 2017). It stands to reason that the double task of learning to read while simultaneously acquiring the language of reading instruction poses distinct challenges. Yet children all over the world learn to read in an L2, for example Spanish children in the US (Manis, Lindsey, & Bailey, 2004; Páez, Tabors, & López, 2007), or Turkish children in the Netherlands (Van Tuijl et al., 2001; Verhoeven, 1987) and in Germany (Klein, Biedinger, & Becker, 2014). However, in those countries the language of literacy instruction is also the majority language of the country. This contrasts sharply with Luxembourg, where children learn to read and write in German, but the main language used in Luxembourg is Luxembourgish. German is only spoken as an L1 by 2% of the student population (MENJE, 2017). Although German is present in the national media in Luxembourg (e.g. TV or newspapers, Gilles, 2014), pupils in Luxembourg have a lot less exposure to the language of literacy instruction outside the school context than L2 learners from the US, the Netherlands or Germany for example. This is highly problematic as OL skills and literacy development are intrinsically linked (Lervåg et al., 2017; Storch & Whitehurst, 2002). The lack of exposure to German OL outside the school context may partly explain why so many French-speaking and Portuguese-speaking struggle to meet the national reading standards in Luxembourg (Hoffmann et al., 2018).

The finding that PA and LSK in Luxembourgish emerged as unique predictors of literacy development in linguistically diverse children is particularly encouraging from an educational point of view, as it is well-established that PA and LSK skills can be successfully promoted (Bowyer-Crane et al., 2008; NRP, 2000; Wawire & Kim, 2018). From a clinical perspective, RAN, VSTM and vocabulary could also be reliably assessed, but current findings seem to support that these variables only explain individual differences in literacy development via shared variance with LSK and PA.

It was unexpected that the finding of RAN as a unique predictor of word reading could not be replicated in the current work (Fricke et al., 2015; Landerl et al., 2019). However, this may be related to the nature of the tasks and the reading strategy employed by the sample. Only a single RAN object task was administered in the current study, whereas Fricke et al. (2015) used a composite score of RAN object and RAN colour, and Landerl et al (2019) used a composite score of RAN colours and RAN digits. Previous work has suggested that the role of RAN as a predictor of reading skills increases with age and reading experience (Scarborough, 1998). Older children have automatized letter- and digit knowledge enough to administer RAN letters or RAN digits task, which have shown to be more predictive of reading (Araújo et al., 2015). In addition, the children in the current sample were in the very early stages of learning to read and RAN may become more predictive of later advanced orthographic reading (Lervåg & Charles Hulme, 2009; Young & Bowers, 1995).

Vocabulary knowledge in Luxembourgish in preschool did not emerge as a unique predictor of learning to read or spell in German in Grade 1. This could be tentatively interpreted as Luxembourgish and German developing orthogonally and that Luxembourgish vocabulary plays no unique role in the development of German literacy skills. Yet this finding could also be interpreted as showing that vocabulary knowledge may only play a limited role in earlier stages of literacy development and may become more important for more advanced stages of literacy skills (Caravolas et al., 2012; Fricke et al., 2015; Tilstra, McMaster, van Den Broek, Kendeou, & Rapp, 2009). However, vocabulary knowledge is considered a language specific skill (Goodrich & Lonigan, 2017) and it would had been very surprising if vocabulary knowledge in Luxembourgish directly contributed to L2 word level literacy skills in German.

In summary, Study 2 clearly provides support for the unique contribution of PA in Luxembourgish and LSK for L2 literacy development in Grade 1. These findings extend previous work to a population of linguistically diverse learners growing up in Luxembourg and learning to read and spell in German. More longitudinal data is needed to further explore the role of the predictors in Luxemburgish on literacy skills at later stages. In this regard, a follow-up assessment of the current sample in Grade 2 is underway.

8.3 Implications

There is a pressing need for early and accurate identification of young children at-risk of developing literacy difficulties to provide timely and appropriate interventions (Beeghly, 2006; Hulme & Snowling, 2011). There are several practical implications that can be drawn from the current thesis to help address this need. The predictor study showed that the foundational skills could be reliably assessed in the second year in preschool in Luxembourgish and that they are predictive of individual differences in literacy in German. This suggests that it would be possible to screen and identify children at risk of literacy difficulties before children begin formal literacy instruction in German. This in turn, would allow practitioners and teachers in multilingual context to customize individual interventions or educational strategies as early as preschool (Snowling & Hulme, 2012). So far, the importance of support early literacy skills in preschool in Luxembourg has not been recognized and incorporated into the preschool curriculum. Preschool education in Luxembourg remains predicated on play-based learning in a holistic approach (MENFP, 2011). In this regard, the preschool curriculum contrasts sharply with the highly structured teaching style in Grade 1, where children are confronted with the double task of learning to read and write in a language that they have yet to fully acquire. The high numbers of children failing to meet the national reading standards in Grade 3 (age 9) in Luxembourg are alarming and clearly indicative of an underlying shortcoming in the way that children are taught to read and write in Luxembourg (Hoffmann et al., 2018). With so many children failing within the first years of formal schooling, it seems evident that providing children with a solid basis in the foundational skills of literacy should be a priority in preschool education.

In the intervention study of the thesis, a new theoretical motivated early literacy preschool programme targeting foundational literacy skills was developed and evaluated. The intervention study therefore offers an important suggestion on how literacy difficulties could be tackled in Luxembourg. From an educational viewpoint, the results are very encouraging as children who struggled with the language instruction of the intervention seemed to have benefitted even more from the intervention. From a theoretical point of view, the study showed that training foundational skills in one language can have a positive impact on the development of early literacy and literacy skills in another language. This finding provides further support for the theory of cross-language transfer of early literacy skills (Goodrich & Lonigan, 2017; Goodrich, Lonigan, & Farver, 2013). It has been suggested that PA and LSK should be combined with OL training for children at-risk of reading failure (Bowyer-Crane et al., 2008). Although combining the provision of early literacy and oral language support in preschools in Luxembourg would be helpful, as many non-Luxembourgish and non-German speaking children struggle with oral comprehension in Luxembourg (Hoffmann et al., 2018), it raises educational as well political problems for preschool education in Luxembourg. As preschooler's OL skills are considered a rather language specific skill (Goodrich & Lonigan, 2017), the long terms benefits of additional OL support in Luxembourgish in preschool with regards to learning to read and write in German in Grade 1 remain questionable. To the best of my knowledge, no rigorously conducted experiment has looked at the effectiveness of OL training in Luxembourgish in preschool on literacy development in German. Supporting oral German in preschool would be more likely to show positive effects on literacy development as more proficient OL in German may reduce the cognitive load in following literacy instruction in German in Grade 1. However, introducing German in preschools in Luxembourg is politically controversial as the two years in preschool are the only two years in the educational system, where the main language of instruction is Luxembourgish. The Luxembourgish language in preschool is seen as the "langue d'intégration" (integration language) for all children from a non-Luxembourgish speaking background (Kirsch, 2018). In addition, instead of orally introducing the language of literacy instruction in preschool, the ministry of Education decided on introducing oral French in preschools in 2018. Before 2018, oral French only used to be introduced in Grade 2 in Luxembourg, and written French in Grade 3. The reason for introducing French in preschools was to facilitate the introduction of oral French in Grade 2. As French has been recently introduced in preschool, it is very unlikely that oral German would be added to the preschool programme in the near future.

Arguably, there is a meaningful alternative to the training of OL skills in German in preschool: to ensure that children in preschool in Luxembourg receive a solid basis in the more language independent code-related skills important for literacy development (i.e. PA and LSK). Providing children with a strong basis in foundation literacy skills in preschool would allow teachers in Grade 1 to shift valuable teaching time from teaching the alphabetic principle to the teaching of oral German. The newly developed intervention programme would offer an opportunity to follow this solution by providing children with a solid basis in PA and LSK in preschool. As such an intervention could be delivered by regular teachers, it could be relatively easily integrated into the existing preschool context. It would not require additional human resources and an implementation at scale would be feasible.

Ahead of scalable implementations, further studies would need to determine the quality assurance mechanisms necessary to ensure that the intervention benefits remain replicable. It is well-known that intervention effects have stronger impacts when the research team both develops and leads the evaluation of the intervention programme (Asmussen et al., 2019). Thus, steps could be undertaken to assess the effectiveness of the intervention independent from the research team to ensure that they remain effective when offered at scale.

8.4 Conclusion

The linguistically diverse pupil population within the multilingual early educational context in Luxembourg raises distinct challenges to all stakeholders, i.e. caregivers, speech and language therapists, teachers and policy makers. The current thesis sheds some light upon how the pathway to literacy of children growing up in Luxembourg can be facilitated.

The thesis showed that it is possible to reliably identify preschool predictors in Luxembourgish of literacy attainment in German in Grade 1. This provides empirical support for teachers and speech and language therapists to identify at-risk children in preschool based on their PA and LSK skills in Luxembourgish. The thesis also demonstrated that it was possible to provide preventive support in preschool targeting PA and LSK skills, and that this could be a promising way to support children literacy development in Grade 1. Particularly encouraging was the result that children with low oral language skills in the language of instruction seem to benefit long-term from such additional early support.

In the introduction of this thesis, the research-practice gap and its negative consequences for public health were highlighted. Moreover, in the last decade, repeated calls have been made for more research on the literacy development of L2 learners and on supportive strategies for linguistically diverse children. The current thesis represents an important step into addressing both of these challenges. Taken together, the intervention and the predictor study provide fruitful first steps in the direction of strengthening the evidence-base for the identification strategies and prevention initiatives of literacy difficulties in Luxembourg.

On average, it has been estimated that it takes seventeen years for any research results to flourish into practical application (Morris et al., 2011).

Let us hope Luxembourg assigns their at-risk pupils a higher priority!

Glossary

ALPHABETIC PRINCIPLE

The understanding that sounds in spoken words can be represented consistently by specific written letters.

ALPHABETIC READING

Application of the alphabetic principle (connecting letters with their sounds) to read words.

BASIC RESEARCH

Basic research, also called fundamental research, is meant to increase the scientific knowledge base with the intent of increasing our understanding of certain phenomena or behaviour.

CODE EMPHASIS TEACHING

Teaching reading by focusing on teaching the specific “code” (i.e. the alphabet in alphabetic languages) or set of letters and words and how to decode the code to read words.

DECODING

Act of converting write language into speech.

ENCODING

Act of converting speech into written language.

GRAPHEME

Graphic representation of a phoneme in a particular language.

INPUT TASK

A verbal response is not required for the input tasks, participants could complete the task by pointing.

LETTER-NAME KNOWLEDGE (LNK)

Knowledge of the names, which represent the individual speech sounds in a language.

LETTER-SOUND KNOWLEDGE (LSK)

Knowledge of the letters, or groups of letters, which represent the individual speech sounds in a language.

LEXICON

Repertoire of words of an individual person.

ORTHOGRAPHIC CONSISTENCY:

The degree of mapping consistency from phonology to orthography.

ORTHOGRAPHIC READING

Orthographic reading-, or orthographic processing skills, refer to the ability to identify spelling patterns of specific letters as words, eventually leading to word recognition. The spelling, pronunciation, and meaning of a word are unified and the information is accessed simultaneously upon visual presentation of an individual word.

OUTPUT TASK

Output tasks required a spoken response from the participants.

PHONEME

The smallest unit of sounds in spoken language distinguishing one word from another.

PHONICS

Teaching approach based on the teaching of this systematic relationship between letters and their corresponding sounds, and how to decode words by blending the individual sounds together; or how to spell by segmenting words into individual sounds.

PHONOLOGICAL AWARENESS

The ability to recognize and manipulate the sounds of spoken words.

TRANSLATION RESEARCH

Strand of research that deals with the application of new knowledge generated by advances in basic science research into new approaches to prevention, diagnosis, and treatment of diseases, or cognitive and learning difficulties.

Appendices

Appendix A

Caregiver background questionnaire

This questionnaire should be completed by the child's parent(s) or legal guardian(s).
This questionnaire is voluntary. If there are questions that you do not want or cannot answer, please just leave them blank.
The questionnaire is anonymous. The name of your child will be removed and replaced with an anonymous code, and will not be identifiable to others outside the project.
We would be extremely grateful if you could complete this questionnaire because the information will help us to better understand what is important for children's learning.

INFORMATION ON THE CHILD

1. Nationality of your child:
- ☐$_1$ Luxembourgish
- ☐$_2$ German
- ☐$_3$ French
- ☐$_4$ Portuguese
- ☐$_5$ Other (which one?) ___________________________

2. Date of birth of your child:
______/______/__________
day month year

3. How long has your child lived in Luxembourg? ☐$_1$ all his life ☐$_2$ since s/he was _____ years old

4. Did your child attend *Précoce* in Luxembourg? ☐$_1$ yes ☐$_2$ no

5. Did your child attend the first year of Kindergarten *(cycle 1.1)* **in Luxembourg?** ☐$_1$ yes ☐$_2$ no

6. How many languages does your child hear at home?

✎ **Only one**
That is? (please only tick <u>one</u> box)
- ☐$_1$ Luxembourgish
- ☐$_2$ German
- ☐$_3$ French
- ☐$_4$ Portuguese
- ☐$_5$ Other (which one?) ___________________________

✎ **More than one**
Those are? (please tick **<u>all</u>** the boxes that apply)
- ☐$_1$ Luxembourgish
- ☐$_2$ German
- ☐$_3$ French
- ☐$_4$ Portuguese
- ☐$_5$ Other (Which one?) ___________________________

7. Which of the following languages is spoken <u>most</u> by your child <u>at home</u>? (please only tick <u>one</u> box)
- ☐$_1$ Luxembourgish
- ☐$_2$ German
- ☐$_3$ French
- ☐$_4$ Portuguese
- ☐$_5$ Other (which one?) ___________________________

8. At what age did your child start to talk? (approximately) __________ years

9. Does your child have a significant health problem? ☐$_2$ No ☐$_1$ Yes
(Which one?) ____

10 . Is your child seeing a professional with regards to his language and/ or speech development? (e.g. speech language therapist, psychologist)
☐$_1$ No ☐$_2$ Yes (why)____

11. How many books do you have at home? (approximately)
(Normally 40 books fit on a shelf of a length of 1 meter: please do **not** include newspapers and school books)

☐$_1$ 0 - 10 Books	☐$_4$ 101 - 200 Books
☐$_2$ 11 - 25 Books	☐$_5$ 201 - 500 Books
☐$_3$ 26 - 100 Books	☐$_6$ more than 500 Books

INFORMATION ON THE PARENTS (or the persons/caregivers that take up this role)

If you do not know the answer to one of the questions, please just leave that question blank.

MOTHER	FATHER
12. Language that the mother speaks most with the child: (please only tick **one** box) ☐1 Luxembourgish ☐2 German ☐3 French ☐4 Portuguese ☐5 Other (Which one?) ____________________________	**13. Language that the father speaks most with the child:** (please only tick **one** box) ☐1 Luxembourgish ☐2 German ☐3 French ☐4 Portuguese ☐5 Other (Which one?) ____________________________
14. Does the mother also speak one or more other languages with the child? ☐2 No ☐1 Yes (Which one/s)______________________	**15. Does the father also speak one or more other languages with the child?** ☐2 No ☐1 Yes (Which one/s)______________________
16. How often does the mother speak this other language(s) with the child (approximately)**?** ☐1 50% ☐2 30-50% ☐3 10-30% ☐4 < 10%	**17. How often does the father speak this other language(s) with the child** (approximately)**?** ☐1 50% ☐2 30-50% ☐3 10-30% ☐4 < 10%
18. Highest level of education completed (e.g. 13ième, Bachelor in Educational Sciences, …) ____	**19. Highest level of education completed** (e.g. 13ième, Bachelor in Educational Sciences, …) ____
20. How many years did it take the mother to complete this education? (please do not include Kindergarten and do not count repeated school years, but do include postsecondary / university education) e.g. 13ième in Luxembourg = 13 years). _______ years	**21. How many years did it take the father to complete this education?** (please do not include Kindergarten and do not count repeated school years, but do include postsecondary / university education) e.g. 13ième in Luxembourg = 13 years). _______ years
22. What is the mother's job title? (e.g. butcher, housewife, educator, vendor in a clothes shop). If the mother is currently not in work, please indicate the most recent work: ____	**23. What is the father's job title?** (e.g. butcher, househusband, educator, vendor in a clothes shop). If the father is currently not in work, please indicate the most recent work: ____
24. Please briefly describe the activities in this job (e.g. sell meat, take care of children, manage a hotel): ____ ____	**24. Please briefly describe the activities in this job** (e.g. sell meat, take care of children, manage a hotel): ____ ____

26. Do you have any other comments that may be important with regards to the development of your child?

____

____

If you have further **questions**, please do not hesitate to get in touch with us. You can find **further information** about our research on the website of our research group **Language and Cognitive Development:** https://langcog.uni.lu/

Thank you very much for your support.

Cyril WEALER
Doctoral Candidate in Psychology
Tel: 46 6644 9387
Email: cyril.wealer@uni.lu
Belval, Université du Luxembourg

Prof. Dr. Pascale ENGEL de ABREU
PhD in Developmental Psychology
Tel: 46 6644 9779
Email: pascale.engel@uni.lu
Belval, Université du Luxembourg

Appendix B

Overview of the intervention structure: 48 intervention sessions administered over 12 weeks with four sessions per week

Week	1	2	3	4	5	6	7	8	9	10	11	12
Introduction of new content / consolidation	introduction of new content					consolidation	introduction of new content					consolidation
LSK	no letter-sounds		introduction and consolidation of letter-sounds									
Targeted PA skills: linguistic unit and level of explicitness	large-unit PA: syllables, rhyme & onset-rhyme; identification, blending, segmentation		small-unit PA: phoneme identification									
				phoneme blending								
									phoneme segmentation			
1st session per week	1	5	9	13	17	21	25	29	33	37	41	45
Letter introduced				M	S	/	R	F	N	W	EI	/
Letters consolidated				L-A	L-A-M-I-T	L-A-M-I	M-I-S-O-B	S-O-R-E-G	R-E-F-U-H	F-U-N-K-D	N-K-W-Z-P	23 sounds
2nd session per week	2	6	10	14	18	22	26	30	34	38	42	46
Letter introduced			L	I	O	/	E	U	K	Z	SCH	/
Letters consolidated				L-A-M	A-M-I-T-S	T-S-O-B	T-S-O-B-R	B-R-E-G-F	G-F-U-H-N	H-N-K-D-W	D-W-Z-P-EI	23 sounds
3rd session per week	3	7	11	15	19	23	27	31	35	39	43	47
Letter introduced			A	T	B	/	G	H	D	P	AU	/
Letters consolidated				L-A-M-I	M-I-T-S-O	L-A-M-I-T-S-O-B	S-O-B-R-E	R-E-G-F-U	F-U-H-N-K	N-K-D-W-Z	W-Z-P-EI-SCH	23 sounds
4th session per week	4	8	12	16	20	24	28	32	36	40	44	48
Letters consolidated			L-A	L-A-M-I-T	M-I-T-S-O-B	L-A-M-I-T-S-O-B	S-O-B-R-E-G	R-E-G-F-U-H	F-U-H-N-K-D	N-K-D-W-Z-P	W-Z-P-EI-SCH-AU	23 sounds

Appendix C

Post-intervention caregiver questionnaire on the intervention material

Parent questionnaire for the "LALA - Lauter lëschteg Lauter" project

During the last 2 trimesters of the "Spillschoul" last year, your child participated in the "LALA - Lauter lëschteg Lauter" Project. We would appreciate if you could take the time to complete the following survey.

1. **How often did you child talk to you about the LALA programme or the LALA parrot?**
   □ 1 often □ 2 sometimes □ 3 never
2. **Do you feel your child enjoyed the LALA programme?** □ 1 yes □ 2 no □ 3 I don't know
3. **Do you feel your child learned something in the LALA programme?** (e.g., new sounds, is more attentive to letters…) □ 1 yes □ 2 no □ 3 I don't know
   - **If yes, what have you noticed:** ______________________________

____

____

____

Last year, you received a "LALA bag" with a storybook, CD, songbook, flash cards and a guide for parents.

4. **Does your child like those materials?** □ 1 very much □ 2 neutral □ 3 not at all □ 4 I don't know
5. **Which materials does your child like in particular?** (please tick all the boxes that apply)
   □ 1 everything
   □ 2 storybook
   □ 3 CD
   □ 4 songbook
   □ 5 flash cards
   □ 6 none
   □ 7 I don't know
6. **How often does your child use the material?** □ 1 often □ 2 sometimes □ 3 never □ 4 I don't know
7. **With whom does your child generally use the material?**
   □ 1 alone □ 2 with his siblings / friends □ 3 with an adult (e.g. Mummy, Daddy, Grandma…)
8. **Have you read the parent guide?** □ 1 yes □ 2 no
9. **If you have NOT read the guide, why not?**
   □ 1 I did not have time
   □ 2 the guide was too long
   □ 3 the guide was too complicated
   □ 4 the guide was written in a language that I did not understand well
   □ 5 other reason ______________________________
10. **Would you have preferred to read it in another language?** □ 1 yes (which?)__________ □ 2 no
11. **Was it easy to understand?** □ 1 yes (it was written clearly) □ 2 no (it was very difficult to understand)
12. **Any additional comments about the LALA Program or the materials:**

1

Appendix D

Overview of data transformations for normality.

Individual variables with a standardized skew above 3.29 were transformed. If a component score showed a standardized skew larger than 3.29, all individual measures included in the component were transformed using the same transformation prior to the PCA.

	Time 2					Time 3				
	Interv. *n*=85	Control *n*=100		Interv. *n*=85	Control *n*=100	Interv. *n*=81	Control *n*=91		Interv. *n*=81	Control *n*=91
	Before transformation			After transformation		Before transformation			After transformation	
	Stand. Skew	Stand. Skew	Type of transformation	Stand. Skew	Stand. Skew	Stand. Skew	Stand. Skew	Type of transformation	Stand. Skew	Stand. Skew
Letter-sound knowledge	**-4,782**	**0.647**	**SQRT(K-X)**	**-2,962**	**-2,353**					
Phoneme blending Ger.[1]						-7.640	-5.877	SQRT(K-X)	3.491	3.372
Onset manipulation Ger.[1]						-0.843	-0.945	SQRT(K-X)	-1.064	-0.802
Phoneme segmentation Ger.[1]						-6.139	-4.028	SQRT(K-X)	3.187	1.913
Small-unit PA component Ger.						**-5.715**	**-3.901**		**2.753**	**2.083**
Word reading SLRT[2]						8.401	13.257	SQRT(X)	4.783	3.601
Non-word reading SLRT[2]						1.926	4.209	SQRT(X)	-1.799	3.597
SLRT reading component						**5.228**	**9.372**		**1.755**	**0.632**
Reading comprehension ELFE						**1.348**	**6.466**	**SQRT(X)**	**-1.888**	**2.577**
Spelling HSP						**-4.295**	**-12.055**	**LOG10(K-X)**	**0.531**	**2.925**
Number naming	**6.115**	**3.822**	**LOG10(X)**	**2.360**	**-0.851**					

Note. [1]No inferential statistics were computed on the individual PA measures in German. The inferential statistics were only performed on the small-unit PA component. However, the individual PA measures in German were transformed prior to the PCA, as the PCA on the untransformed measures resulted in a component with a standardized skew larger than the 3.29 threshold.

[2]No inferential statistics were computed on the individual reading measures of the SLRT. The inferential statistics were only performed on the SLRT reading component. However, the individual SLRT reading measures were transformed prior to the PCA, as the PCA on the untransformed SLRT reading measures resulted in a component with a standardized skew larger than the 3.29 threshold.

Appendix E

Overview of the PA and literacy components with factorability statistics at pre-test(*t*1), post-test(*t*2) and delayed follow-up(*t*3)

Components	component variables	Factor loadings	factorability
Pre-test (*t*1)			
PA large-unit (Lux.)	rhyme Identification	.66	KMO= .67
	onset-rhyme blending	.60	Bartlett's test, $p< .001$
	onset Identification	.71	explained var.= 66%
PA small-unit (Lux)	phoneme blending	.88	KMO= .50
	phoneme manipulation.	.88	Bartlett's test, $p< .001$
			explained var.= 78%
Basic reading	basic word reading	.99	KMO= .50
	basic non-word reading	.99	Bartlett's test, $p< .001$
			explained var.= 97%
Post-test (*t*2)			
PA large-unit (Lux.)	rhyme Identification	.73	KMO= .65
	onset-rhyme blending	.83	Bartlett's test, $p< .001$
	onset Identification	.86	explained var.= 65%
PA small-unit (Lux.)	phoneme blending	.90	KMO= .68
	phoneme manipulation	.81	Bartlett's test, $p< .001$
	phoneme segmenting	.84	explained var.= 73%
Basic reading	basic word reading	.97	KMO= .50
	basic non-word reading	.97	Bartlett's test, $p< .001$
			explained var.= 95%
Delayed follow-up (*t*3)			
PA small-unit (Lux.)	phoneme blending	.84	KMO= .65
	phoneme manipulation	.77	Bartlett's test, $p< .001$
	phoneme segmenting	.76	explained var.= 63%
PA small-unit (Ger.)	phoneme blending	.88	KMO= .65
	phoneme manipulation	.80	Bartlett's test, $p< .001$
	phoneme segmenting	.79	explained var.= 68%
Basic reading	basic word reading	.93	KMO= .50
	basic non-word reading	.93	Bartlett's test, $p< .001$
			explained var.= 86%
SLRT reading	word reading SLRT	.95	KMO= .50
	non-word reading SLRT	.95	Bartlett's test, $p< .001$
			explained var.= 90%

Note. Lux., Luxembourgish; Ger., German; SLRT, Salzburger Lese- und Rechtschreibtests [Salzburg reading and orthography Test].

References

Aarnoutse, C., & van Leeuwe, J. (2000). Development of poor and better readers during the elementary school. *Educational Research and Evaluation*, *6*(3), 251–278. https://doi.org/10.1076/1380-3611

Aarnoutse, C., Van Leeuwe, J., & Verhoeven, L. (2005). Early literacy from a longitudinal perspective. *Educational Research and Evaluation*, *11*(3), 253–275. https://doi.org/10.1080/08993400500101054

Abbott, R. D., Berninger, V. W., & Fayol, M. (2010). Longitudinal Relationships of Levels of Language in Writing and Between Writing and Reading in Grades 1 to 7. *Journal of Educational Psychology*, *102*(2), 281–298. https://doi.org/10.1037/a0019318

Adams, M. J. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.

Aikens, N. L., & Barbarin, O. (2008). Socioeconomic Differences in Reading Trajectories: The Contribution of Family, Neighborhood, and School Contexts. *Journal of Educational Psychology*, *100*(2), 235–251. https://doi.org/10.1037/0022-0663.100.2.235

Alloway, T. P. (2007). *Automated working memory assessment*. London: Harcourt Assessment.

Alloway, T. P., Gathercole, S. E., Willis, C., & Adams, A. M. (2004). A structural analysis of working memory and related cognitive skills in young children. *Journal of Experimental Child Psychology*, *87*(2), 85–106. https://doi.org/10.1016/j.jecp.2003.10.002

Anthony, J. L., Lonigan, C. J., Driscoll, K., Phillips, B. M., & Burgess, S. R. (2003). Phonological sensitivity: A quasi-parallel progression of word structure units and cognitive operations. *Reading Research Quarterly*, *38*(4), 470–487. https://doi.org/10.1598/RRQ.38.4.3

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

Anthony, J. L., & Lonigan, C. J. (2004). The Nature of Phonological Awareness: Converging Evidence From Four Studies of Preschool and Early Grade School Children. *Journal of Educational Psychology*, *96*(1), 43–55. https://doi.org/10.1037/0022-0663.96.1.43

Araújo, S., Reis, A., Petersson, K. M., & Faísca, L. (2015). The relationship between rapid automatized naming and reading performance: a meta-analysis. *Journal of Educational Psychology*, *107*(3), 868–883. Retrieved from http://sapientia.ualg.pt/handle/10400.1/5934

Arbuckle, J. L. (2017). *Amos (Version 25.0) [Computer Program]*. Chicago: IBM SPSS.

Aro, M., & Wimmer, H. (2003). Learning to read: English in comparison to six more regular orthographies. *Applied Psycholinguistics*, *24*(4), 621–635. https://doi.org/10.1017/S0142716403000316

Arredondo, M. M., Hu, X. S., Satterfield, T., & Kovelman, I. (2017). Bilingualism alters children's frontal lobe functioning for attentional control. *Developmental Science*, *20*(3), 1–25. https://doi.org/10.1111/desc.12377.Bilingualism

Asmussen, K., Brims, L., & Mcbride, T. (2019). *AR 10 steps for evaluation success*. London.

August, D., Carlo, M., Dressler, C., & Snow, C. (2005). The Critical Role of Vocabulary Development for English Language Learners. *Learning Disabilities Research and Practice*, *20*(1), 50–57. https://doi.org/10.1111/j.1540-5826.2005.00120.x

August, D., & Shanahan, T. (2005). *Developing Reading and Writing in Second-Language Learners*. New York: Routledge.

August, D., & Shanahan, T. (2010). Response to a review and update on developing literacy in second-language learners: Report of the national literacy panel on language minority children and youth. *Journal of Literacy Research*, *42*(3), 341–348. https://doi.org/10.1080/1086296X.2010.503745

Babayiğit, S. (2015). The relations between word reading, oral language, and reading comprehension in children who speak English as a first (L1) and second language (L2): a multigroup structural analysis. *Reading and Writing*, *28*(4), 527–544. https://doi.org/10.1007/s11145-014-9536-x

Baddeley, A. D. (1986). *Working memory*. Oxford: Oxford University Press.

Baddeley, A. D. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417–423. https://doi.org/10.1016/j.apm.2016.02.027

Baddeley, A. D, Allen, R. J., & Hitch, G. (2010). Investigating the episodic buffer. *Psychologica Belgica*, *50*(3&4), 223–243. https://doi.org/10.5334/pb-50-3-4-223

Baddeley, A. D, & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (pp. 47–89). San Diego: Academic Press.

Baddeley, A. D, & Logie, R. H. (1999). Working memory: The multicomponent model. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control*. New York: Cambridge University Press.

Baddeley, A. D., & Wilson, B. A. (2002). Prose recall and amnesia: Implications for the structure of working memory. *Neuropsychologia*, *40*(10), 1737–1743. https://doi.org/10.1016/S0028-3932(01)00146-4

Baddeley, A. D, Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, *105*(1), 158–173. https://doi.org/10.1016/j.cortex.2018.05.015

Baroody, A. E., & Diamond, K. E. (2012). Links Among Home Literacy Environment, Literacy Interest, and Emergent Literacy Skills in Preschoolers At Risk for Reading Difficulties. *Topics in Early Childhood Special Education*, *32*(2), 78–87. https://doi.org/10.1177/0271121410392803

Beck, I. L., & Beck, M. (2013). *Making Sense of Phonics*. New York: Guilford Press.

Beck, I. L., McKeown, M. G., & Kucan, L. (2013). *Bringing words to life: Robust vocabulary instruction* (2nd ed.). New York: Guilford Press.

Beech, J. R. (2005). Ehri's model of phases of learning to read: A brief critique. *Journal of Research in Reading*, *28*(1), 50–58. https://doi.org/10.1111/j.1467-9817.2005.00252.x

Beeghly, M. (2006). Translational research on early language development: Current challenges and future directions. *Development and Psychopathology*, *18*(03), 737–757. https://doi.org/10.1017/S0954579406060366

Bellocchi, S., Tobia, V., & Bonifacci, P. (2017). Predictors of reading and comprehension abilities in bilingual and monolingual children: a longitudinal study on a transparent language. *Reading and Writing*, *30*(6), 1311–1334. https://doi.org/10.1007/s11145-017-9725-5

Bentler, P. M., & Chou, C. P. (1987). Practical Issues in Structural Modeling. *Sociological Methods & Research*, *16*(1), 78–117. https://doi.org/10.1177/0049124187016001004

Berg, K., Eichmeyer, A., Kunze, H., Mager, E., Stiebritz, C., & Von Werder, K. (2009). *Karibu Fibel*. Braunschweig: Westermann.

Bergmann, J., & Wimmer, H. (2008). A dual-route perspective on poor reading in a regular orthography : Evidence from phonological and orthographic lexical decisions A dual-route perspective on poor reading in a regular orthography : Evidence from phonological and orthographic lexical decis. *Cognitive Neuropsychology*, *25*(5), 653–676. https://doi.org/10.1080/02643290802221404

Bialystok, E., Luk, G., Peets, K. F., & Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition*, *13*(4), 525–531. https://doi.org/10.1097/CCM.0b013e31823da96d.Hydrogen

Bialystok, E., Craik, F. I. M., Green, D. W., & Gollan, T. H. (2009). Bilingual minds. *Psychological Science in the Public Interest, Supplement*, *10*(3), 89–129. https://doi.org/10.1177/1529100610387084

Bialystok, E., Majumder, S., & Martin, M. M. (2003). Developing phonological awareness: Is there a bilingual advantage? *Applied Psycholinguistics*, *24*(1), 27–44. https://doi.org/10.1017/S014271640300002X

Biesta, G. (2007). Why "what works" won't work: Evidence-based practice and the democratic deficit in educational research. *Educational Theory*, *57*(1), 1–22. https://doi.org/10.1111/j.1741-5446.2006.00241.x

Biltgen, C., Brosius, E., Hilger, M., Letsch, J., Neuens, O., Reding, P., … Weibes, F. (2013). *Mila. Mein erstes Lesebuch*. Luxembourg: Ministère de l'Education nationale et de la Formation professionnelle.

Bishop, D. V. M., & Snowling, M. J. (2004). Developmental dyslexia and specific language impairment: Same or different? *Psychological Bulletin*, *130*(6), 858–886. https://doi.org/10.1037/0033-2909.130.6.858

Blanca, M., Alarcón, R., Arnau, J., Bono, R., & Bendayan, R. (2017). Non-normal data: Is ANOVA still a valid option? *Psicothema*, *29*(4), 552–557.

https://doi.org/10.7334/psicothema2016.383

Bodé, S. (2004). *Phonologesch Bewosstheet am Spillschoulsalter*. Luxembourg: Saint-Paul.

Bodé, S., & Content, A. (2011). Phonological awareness in kindergarten: A field study in Luxembourgish schools. *European Journal of Psychology of Education*, *26*(1), 109–128. https://doi.org/10.1007/s10212-010-0039-0

Boets, B., De Smedt, B., Cleuren, L., Vandewalle, E., Wouters, J., & Ghesqui??re, P. (2010). Towards a further characterization of phonological and literacy problems in Dutch-speaking children with dyslexia. *British Journal of Developmental Psychology*, *28*(1), 5–31. https://doi.org/10.1348/026151010X485223

Bosman, A. M. T., & Van Orden, G. C. (1997). Why Spelling is More Difficult than Reading. In CA Perfetti, L. Rieben, & M. Fayol (Eds.), *The psycholinguistics of spelling and reading*. Mahwah, NJ: Lawrence Erlbaum.

Bowey, J. A. (1994). Phonological sensitivity in novice readers and nonreaders. *Journal of Experimental Child Psychology*, *58*(1), 134–159. https://doi.org/10.1006/jecp.1994.1029

Bowey, J. A. (1995). Socioeconomic status differences in preschool phonological sensitivity and first-grade reading achievement. *Journal of Educational Psychology*, *87*(3), 476–487. https://doi.org/10.1037/0022-0663.87.3.476

Bowey, J. A. (2005a). Predicting individual differences in learning to read. *The Science of Reading: A Handbook.* Retrieved from https://espace.library.uq.edu.au/view/UQ:72024/UQ_AV_72024.pdf

Bowey, J. A. (2005b). Predicting individual differences in learning to read. In M. J. Snowling & C. Hulme (Eds.), *The Science of Reading: A Handbook* (pp. 155–172). Oxford: Blackwell Publishing.

Bowman, M., & Treiman, R. (2004). Stepping stones to reading. *Theory into Practice*, *43*(4), 295–303. https://doi.org/10.1207/s15430421tip4304_8

Bowyer-Crane, C., Fricke, S., Schaefer, B., Lervåg, A., & Hulme, C. (2017). Early literacy and comprehension skills in children learning English as an additional language and monolingual children with language weaknesses. *Reading and Writing*, *30*(4), 771–790. https://doi.org/10.1007/s11145-016-9699-8

Bowyer-Crane, C., Snowling, M. J., Duff, F. J., Fieldsend, E., Carroll, J. M., Miles, J., … Hulme, C. (2008). Improving early language and literacy skills: Differential effects of an oral language versus a phonology with reading intervention. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *49*(4), 422–432. https://doi.org/10.1111/j.1469-7610.2007.01849.x

Bradley, L., & Bryant, P. E. (1983). Categorizing sounds and learning to read - A causal connection. *Nature*, *301*(5899), 419–421. https://doi.org/10.1038/301419a0

Bradley, R. H., Corwyn, R. F., Burchinal, M., McAdoo, H. P., & Garcia Coll, C. (2001). The Home Environments of Children in the United States Part II: Relations with Behavioral Development through Age Thirteen. *Child Development*, *72*(6), 1868–1886. https://doi.org/10.1111/1467-8624.t01-1-00383

Brandenburg, J., Klesczewski, J., Fischbach, A., Schuchardt, K., Büttner, G., & Hasselhorn, M. (2015). Working Memory in Children With Learning Disabilities in Reading Versus Spelling: Searching for Overlapping and Specific Cognitive Factors. *Journal of Learning Disabilities*, *48*(6), 622–634. https://doi.org/10.1177/0022219414521665

Branum-Martin, L., & Garnaat, S. (2015). Bilingual Phonological Awareness : Reexamining the Evidence for Relations Within and Across Languages. *Journal of Educational Psychology*, *107*(1), 111–125. https://doi.org/10.1037/a0037149

Branum-Martin, L., Tao, S., Garnaat, S., Bunta, F., & Francis, D. J. (2012). Meta-analysis of bilingual phonological awareness: Language, age, and psycholinguistic grain size. *Journal of Educational Psychology*, *104*(4), 932–944. https://doi.org/10.1037/a0027755

Brown, C. S. (2014). Language and Literacy Development in the Early Years: Foundational Skills that Support Emergent Readers. *Language and Literacy Development in the Early Years*, *24*, 35–48.

Burgess, S. R., Hecht, S. A., & Lonigan, C. J. (2002). Relations of the home literacy environment (HLE) to the development of reading-related abilities: A one-year longitudinal study. *Reading Research Quarterly*, *37*(4), 408–426. https://doi.org/10.1598/RRQ.37.4.4

Burgess, S. R., & Lonigan, C. J. (1998). Bidirectional Relations of Phonological Sensitivity and Prereading Abilities: Evidence from a Preschool Sample. *Journal of Experimental Child Psychology*, *70*(2), 117–141. https://doi.org/10.1006/jecp.1998.2450

Burgoyne, K., Kelly, J. M., Whiteley, H. E., & Spooner, A. (2009). The comprehension skills of children learning English as an additional language. *British Journal of Educational Psychology*, *79*(4), 735–747. https://doi.org/10.1348/000709909X422530

Bus, A. G., & van IJzendoorn, M. H. (1999). Phonological awareness and early reading: A meta-analysis of experimental training studies. *Journal of Educational Psychology*, *91*(3), 403–414. https://doi.org/10.1037//0022-0663.91.3.403

Butler, D. (2008). Translational research: crossing the valley of death. *Nature*, *453*(7197), 840–842.

Byrne, B. (2010). *Structural Equation Modeling with AMOS*. New York: Routledge.

Cain, K. E., & Oakhill, J. V. (1999). Inference making ability and its relation to comprehension failure in young children. *Reading and Writing*, *11*(5/6), 489–503. https://doi.org/10.1023/A:1008084120205

Calvo, N., Ibáñez, A., & García, A. M. (2016). The Impact of Bilingualism on Working Memory: A Null Effect on the Whole May Not Be So on the Parts. *Frontiers in Psychology*, *7*, 1–4. https://doi.org/10.3389/fpsyg.2016.00265

Campbell, R. (1995). Accelerated Metalinguistic (Phonological) Awareness in Bilingual Children. *British Journal of Developmental Psychology*, *13*(1), 61–68. https://doi.org/10.1111/j.2044-835X.1995.tb00664.x

Caravolas, M. (2004). Spelling development in alphabetic writing systems: A cross-linguistic perspective. *European Psychologist*, *9*(1), 3–14. https://doi.org/10.1027/1016-9040.9.1.3

Caravolas, M., Hulme, C., & Snowling, M. J. (2001). The foundations of spelling ability: Evidence from a 3-year longitudinal study. *Journal of Memory and Language*, *45*(4), 751–774. https://doi.org/10.1006/jmla.2000.2785

Caravolas, M., Lervåg, A., Defior, S., Seidlová Málková, G., & Hulme, C. (2013). Different Patterns, but Equivalent Predictors, of Growth in Reading in Consistent and Inconsistent Orthographies. *Psychological Science*, *24*(8), 1398–1407. https://doi.org/10.1177/0956797612473122

Caravolas, M., Lervåg, A., Mousikou, P., Efrim, C., Litavský, M., Onochie-Quintanilla, E., … Hulme, C. (2012). Common Patterns of Prediction of Literacy Development in Different Alphabetic Orthographies. *Psychological Science*, *23*(6), 678–686. https://doi.org/10.1177/0956797611434536

Caravolas, M., Volín, J., & Hulme, C. (2005). Phoneme awareness is a key component of alphabetic literacy skills in consistent and inconsistent orthographies: Evidence from Czech and English children. *Journal of Experimental Child Psychology*, *92*(2), 107–139. https://doi.org/10.1016/j.jecp.2005.04.003

Cárdenas-Hagan, E., Carlson, C. D., & Pollard-Durodola, S. D. (2007). The Cross-Linguistic Transfer of Early Literacy Skills: The Role of Initial L1 and L2 Skills and Language of Instruction. *Language Speech and Hearing Services in Schools*, *38*(3), 249–259. https://doi.org/10.1044/0161-1461(2007/026)

Carnine, D. W., Silbert, J., Kame'enui, E. J., & Tarver, S. G. (2004). *Direct Reading Instruction*. Merrill Prentice Hall, New Jersey: Pearson.

Carroll, J. M., Bowyer-Crane, C., Duff, F. J., Hulme, C., & Snowling, M. J. (2011). *Developing language and literacy: Effective intervention in the early years*. Oxford: Wiley Blackwell.

Carroll, J. M., Mundy, I. R., & Cunningham, A. J. (2014). The roles of family history of dyslexia, language, speech production and phonological processing in predicting literacy progress. *Developmental Science*, *17*(5), 727–742. https://doi.org/10.1111/desc.12153

Carroll, J. M., Snowling, M. J., Stevenson, J., & Hulme, C. (2003). The development of phonological awareness in preschool children. *Developmental Psychology*, *39*(5), 913–923. https://doi.org/10.1037/0012-1649.39.5.913

Carroll, J. M, & Snowling, M. J. (2004). Language and phonological skills in children at high risk of reading difficulties. *Jounral of Child Psychology and Psychiatry*, *45*(3), 631–640.

Castles, A., & Coltheart, M. (2004). Is there a causal link from phonological awareness to success in learning to read? *Cognition*, *91*(1), 77–111. https://doi.org/10.1016/S0010-0277(03)00164-1

Castles, A., Coltheart, M., Wilson, K., Valpied, J., & Wedgwood, J. (2009). The genesis of reading ability: What helps children learn letter-sound correspondences? *Journal of Experimental Child Psychology*, *104*(1), 68–88. https://doi.org/10.1016/j.jecp.2008.12.003

Castles, A., Wilson, K., & Coltheart, M. (2011). Early orthographic influences on phonemic awareness tasks: Evidence from a preschool training study. *Journal of Experimental Child Psychology*, *108*(1), 203–210. https://doi.org/10.1016/j.jecp.2010.07.006

Chard, D., & Dickson, S. (1999). Phonological Awareness: Instructional and Assessment Guidelines I LD Topics I LD OnLine. *LD Online*, (May). Retrieved from http://www.ldonline.org/article/6254/

Charmaz, K. (2006). *Constructing Grounded Theory. A practical guide through qualitative analysis*. London: SAGE.

Chen, X., Xu, F., Nguyen, T. K., Hong, G., & Wang, Y. (2010). Effects of Cross-Language Transfer on First-Language Phonological Awareness and Literacy Skills in Chinese Children Receiving English Instruction. *Journal of Educational Psychology*, *102*(3), 712–728. https://doi.org/10.1037/a0018802

Chiappe, P., Siegel, L. S., & Gottardo, A. (2002). Reading-related skills of kindergartners from diverse linguistic backgrounds. *Applied Psycholinguistics*, *23*(1), 95–116. https://doi.org/10.1017/S014271640200005X

Chiappe, P., Siegel, L. S., & Wade-Woolley, L. (2002). Linguistic Diversity and the Development of Reading Skills: A Longitudinal Study. *Scientific Studies of Reading*, *6*(4), 369–400. https://doi.org/10.1207/S1532799XSSR0604_04

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Coltheart, M, Rastle, K., Perry, C., Langdon, R., & Ziegler, J. C. (2001). A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*.

Coltheart, M. (2005). Modeling Reading: The Dual-Route Approach. In Margaret J Snowling & C. Hulme (Eds.), *The science of reading: A handbook*. Oxford: Blackwell Publishing.

Coolican, H. (2009). *Research Methods and Statistics in Psychology*. London: Hodder Education.

Corp, I. (2016). *IBM SPSS Statistics for Windows, Version 24.0*. Armonk, NY: IBM Corp.

Cossu, G., Gugliotta, M., & Marshall, J. C. (1995). Acquisition of reading and written spelling in a transparent orthography: Two non parallel processes? *Reading and Writing*, *7*(1), 9–22. https://doi.org/10.1007/BF01026945

Cutting, L. E., & Scarborough, H. S. (2006). Prediction of reading comprehension: Relative contributions of word recognition, language proficiency, and other cognitive skills can depend on how comprehension is measured. *Scientific Studies of Reading*, *10*(December 2015), 257–275. https://doi.org/10.1207/s1532799xssr1003

Da Fontoura, H. A., & Siegel, L. S. (1995). Reading, syntactic, and working memory skills of bilingual Portuguese-English Canadian children. *Reading and Writing*, *7*(1), 139–153. https://doi.org/10.1007/BF01026951

De Jong, P. F. (2011). What discrete and serial rapid automatized naming can reveal about reading. *Scientific Studies of Reading*, *15*(4), 314–337. https://doi.org/10.1080/10888438.2010.485624

De Jong, P. F. (1998). Working Memory Deficits of Reading Disabled Children. *Journal of Experimental Child Psychology*, *70*(2), 75–96. https://doi.org/10.1006/jecp.1998.2451

De Jong, P. F., & van der Leij, A. (1999). Specific contributions of phonological abilities to early reading acquisition: Results from a Dutch latent variable longitudinal study. *Journal of Educational Psychology*. https://doi.org/10.1037/0022-0663.91.3.450

Denckla, M. B., & Rudel, R. G. (1976). Rapid “automatized” naming (R.A.N): dyslexia differentiated from other learning disabilities. *Neuropsychologia*, *14*(4), 471–479. https://doi.org/10.1016/0028-3932(76)90075-0

Derrick A. Bennett. (2001). How can I deal with missing data in my study? *Australian and New Zealand Journal of Public Health*, *25*(5). Retrieved from https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-842X.2001.tb00294.x

Dessemontet, R., & de Chambrier, A. F. (2015). The role of phonological awareness and letter-sound knowledge in the reading development of children with intellectual disabilities. *Research in Developmental Disabilities*, *41–42*, 1–12. https://doi.org/10.1016/j.ridd.2015.04.001

Dimitrov, D. M., & Rumrill, P. D. (2003). Pretest-posttest designs and measurement of change. *Work*, *20*(2), 159–165. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12671209

Dombey, H. (2006). Phonics and English orthography. In M. Lewis & S. Ellis (Eds.), *Phonics: Practice, research and policy* (pp. 95–104). London, England: Paul Chapman Publishing.

Drolet, B. C., & Lorenzi, N. M. (2011). Translational research: Understanding the continuum from bench to bedside. *Translational Research*, *157*(1), 1–5. https://doi.org/10.1016/j.trsl.2010.10.002

Droop, M., & Verhoeven, L. (2003). Language proficiency and reading ability in first- and second-language learners. *Reading Research Quarterly*, *38*(1), 78–103. https://doi.org/10.1598/rrq.38.1.4

Duff, F. J., & Clarke, P. J. (2011). Practitioner Review: Reading disorders: What are the effective interventions and how should they be implemented and evaluated? *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *52*(1), 3–12. https://doi.org/10.1111/j.1469-7610.2010.02310.x

Duff, F. J., Fieldsend, E., Bowyer-Crane, C., Hulme, C., Smith, G., Gibbs, S., & Snowling, M. J. (2008). Reading with vocabulary intervention: Evaluation of an instruction for children with poor response to reading intervention. *Journal of Research in Reading*, *31*(3), 319–336. https://doi.org/10.1111/j.1467-9817.2008.00376.x

Duff, F. J., Reen, G., Plunkett, K., & Nation, K. (2015). Do infant vocabulary skills predict school-age language and literacy outcomes? *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *56*(8), 848–856. https://doi.org/10.1111/jcpp.12378

Duncan, L. G., & Seymour, P. H. K. (2000). Socio-economic differences in foundation-level literacy. *British Journal of Psychology*, *91*(2), 145. https://doi.org/10.1348/000712600161736

Duncan, L. G., Colé, P., Seymour, P. H. K., & Magnan, A. (2006). Differing sequences of metaphonological development in French and English. *Journal of Child Language*, *33*(2), 369–399. https://doi.org/10.1017/S030500090600732X

Duncan, L. G., & Seymour, P. H. K. (2003). Socio-economic differences in foundation-level literacy. *British Journal of Psychology*, *91*(2), 145–166. https://doi.org/10.1348/000712600161736

Dunn, D. M., & Dunn, L. M. (2007). *Peabody picture vocabulary test: Manual*. Pearson.

Durgunoĝlu, A. Y., Nagy, W. E., & Hancin-Bhatt, B. J. (1993). Cross-Language Transfer of Phonological Awareness. *Journal of Educational Psychology*, *85*(3), 453–465. https://doi.org/10.1037/0022-0663.85.3.453

Durgunoglu, A. Y., & Oney, B. (1999). A cross-linguistic comparison of phonological awareness and word recognition. *Reading and Writing*, *11*(4), 281–299. https://doi.org/Doi 10.1023/A:1008093232622

Durham, R. E., Farkas, G., Hammer, C. S., Bruce Tomblin, J., & Catts, H. W. (2007). Kindergarten oral language skill: A key variable in the intergenerational transmission of socioeconomic status. *Research in Social Stratification and Mobility*, *25*(4), 294–305. https://doi.org/10.1016/j.rssm.2007.03.001

Dusenbury, L., Brannigan, R., Falco, M., & Hansen, W. B. (2003). A review of research on fidelity of implementation: Implications for drug abuse prevention in school settings. *Health Education Research*, *18*(2), 237–256. https://doi.org/10.1093/her/18.2.237

Edele, A., & Stanat, P. (2016). The role of first-language listening comprehension in second-language reading comprehension. *Journal of Educational Psychology*, *108*(2), 163–180. https://doi.org/10.1037/edu0000060

Ehri, L. C. (1997). Learning to Read and Learning to Spell are One and the Same, Almost. In CA Perfetti, L. Rieben, & M. Fayol (Eds.), *Learning to spell: Research, theory, and practice across languages* (pp. 237–269). Mahwah, NJ: Lawrence Erlbaum.

Ehri, L. C., & Wilce, L. S. (1987). Does Learning to Spell Help Beginners Learn to Read Words ? *Reading Reasearch Quarterly*, *22*(1), 47–65.

Ehri, L.C. (2005). Learning to Read Words: Theory, Findings, and Issues. *Scientific Studies of Reading*, *9*(2), 167–188. https://doi.org/10.1207/s1532799xssr0902_4

Ehri, L.C., & Wilce, L. S. (1985). Movement into Reading : Is the First Stage of Printed Word Learning Visual or Phonetic ? *Research Quarterly*, *20*(2), 163–179.

Ehri, L. C. (1991). Development of the ability to read words. In R. Barr, M. L. Kamil, P. Mosenthal, & P. D. Pearson (Eds.), *Handbook of reading research* (2nd ed., pp. 383–417). New York: Longman.

Ehri, L. C. (2014). Orthographic Mapping in the Acquisition of Sight Word Reading, Spelling Memory, and Vocabulary Learning. *Scientific Studies of Reading*, *18*(1), 5–21. https://doi.org/10.1080/10888438.2013.819356

Ehri, L. C., Nunes, S. R., Willows, D. M., Schuster, B. V., Yaghoub-Zadeh, Z., & Shanahan, T. (2001). Phonemic Awareness Instruction Helps Children Learn to Read: Evidence From the National Reading Panel's Meta-Analysis. *Reading Research Quarterly*, *36*(3), 250–287. https://doi.org/10.1598/RRQ.36.3.2

Ehri, L. C, Nunes, S. R., Stahl, S. A., Willows, D. M., Ehri, L. C., Nunes, S. R., & Stahl, S. A. (2001). Systematic Phonics Instruction Helps Students Learn to Read : Evidence from the National Reading Panel ' s Meta-Analysis Published by : American Educational Research Association Stable URL : http://www.jstor.org/stable/3516004 Your use of the JSTOR archiv. *Review of Educational Research*, *71*(3), 393–447. https://doi.org/10.3102/00346543071003393

Elwér, Å., Keenan, J. M., Olson, R. K., Byrne, B., & Samuelsson, S. (2013). Longitudinal stability and predictors of poor oral comprehenders and poor decoders. *Journal of Experimental Child Psychology*, *115*(3), 497–516. https://doi.org/10.1016/j.jecp.2012.12.001

Engel de Abreu, P. M. J., Hornung, C., & Martin, R. (2015). Wie lernen Kinder Sprache(n)? Überlegungen zum Spracherwerb und Alphabetisierung in Luxemburg aus Sicht der Kognitionswissenschaften. In MENJE: Ministère de l'Éducation nationale de l'Enfance et de la Jeunesse (Ed.), *Bildungsbericht Luxemburg 2014* (pp. 15–21). https://doi.org/10.14800/ics.95

Engel de Abreu, P. M. J., Conway, A. R. A., & Gathercole, S. E. (2010). Working memory and fluid intelligence in young children. *Intelligence*, *38*(6), 552–561.

Engel de Abreu, P. M. J., Cruz-Santos, A., Tourinho, C. J., Martin, R., & Bialystok, E. (2012). Bilingualism Enriches the Poor. *Psychological Science*, *23*(11), 1364–1371. https://doi.org/10.1177/0956797612443836

Engel de Abreu, P. M.J., & Gathercole, S. E. (2012). Executive and phonological processes in second-language acquisition. *Journal of Educational Psychology*, *104*(4), 974–986. https://doi.org/10.1037/a0028390

Erdos, C., Genesee, F., Savage, R., & Haigh, C. (2014). Predicting Risk for Oral and Written Language Learning Difficulties in Students Educated in a Second Language 1. *Applied Psycholinguistics, 35*(2), 371–398. https://doi.org/10.1017/S0142716412000422

Erdos, C., Genesee, F., Savage, R., & Haigh, C. A. (2011). Individual differences in second language reading outcomes. *International Journal of Bilingualism*, *15*(1), 3–25. https://doi.org/10.1177/1367006910371022

Fabiano-Smith, L., & Barlow, J. A. (2010). Interaction in bilingual phonological acquisition: Evidence from phonetic inventories. *International Journal of Bilingual Education and Bilingualism*, *13*(1), 81–97. https://doi.org/10.1080/13670050902783528

Fallis, A. (2013). Non – Responders: Current Findings in Reading Instruction as Related to Students. *Journal of Chemical Information and Modeling*, *53*(9), 1689–1699. https://doi.org/10.1017/CBO9781107415324.004

Farnia, F., & Geva, E. (2013). Growth and predictors of change in English language learners' reading comprehension. *Journal of Research in Reading*, *36*(4), 389–421. https://doi.org/10.1111/jrir.12003

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191. https://doi.org/10.1088/1755-1315/148/1/012022

Feinauer, E., Hall-Kenyon, K. M., & Davison, K. C. (2013). Cross-Language Transfer of Early Literacy Skills: An Examination of Young Learners in a Two-Way Bilingual Immersion Elementary School. *Reading Psychology*, *34*(5), 436–460. https://doi.org/10.1080/02702711.2012.658142

Ferguson, C. J. (2009). An Effect Size Primer: A Guide for Clinicians and Researchers. *Professional Psychology: Research and Practice*, *40*(5), 532–538. https://doi.org/10.1037/a0015808

Field, A. (2013). *Discovering statistics using SPSS*. London: Sage.

Fischer, M. Y., & Pfost, M. (2015). Wie effektiv sind Maßnahmen zur Förderung der phonologischen Bewusstheit?: Eine meta-analytische Untersuchung der Auswirkungen deutschsprachiger Trainingsprogramme auf den Schriftspracherwerb. *Zeitschrift Fur Entwicklungspsychologie Und Padagogische Psychologie*, *47*(1), 35–51. https://doi.org/10.1026/0049-8637/a000121

Florit, E., & Cain, K. (2011). The Simple View of Reading: Is It Valid for Different Types of Alphabetic Orthographies? *Educational Psychology Review*, *23*(4), 553–576. https://doi.org/10.1007/s10648-011-9175-6

Forget-Dubois, N., & Dionne, G. (2009). Early child language mediates the relationship between home environment and school readiness. *Child Development*, *80*(3), 736–749.

Foulin, J. N. (2005). Why is letter-name knowledge such a good predictor of learning to read? *Reading and Writing*, *18*(2), 129–155. https://doi.org/10.1007/s11145-004-5892-2

Fricke, S., & Schaefer, B. (2008). *Phonologische Bewusstheitsfähigkeiten (TPB) [Test for phonological awareness skills]*. Idstein: Schulz-Kirchner Verlag.

Fricke, S, Bowyer-Crane, C., Haley, A. J., Hulme, C., & Snowling, M. J. (2013). Efficacy of language intervention in the early years. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *54*(3), 280–290. https://doi.org/10.1111/jcpp.12010

Fricke, S, Burgoyne, K., Bowyer-Crane, C., Kyriacou, M., Zosimidou, A., Maxwell, L., … Hulme, C. (2017). The efficacy of early language intervention in mainstream school settings: a randomized controlled trial. *Journal of Child Psychology and Psychiatry*, *58*(10), 1141–1151. https://doi.org/10.1111/jcpp.12737

Fricke, S, Szczerbinski, M., Fox-Boyer, A., & Stackhouse, J. (2015). Preschool Predictors of Early Literacy Acquisition in German-Speaking Children. *Reading Research Quarterly*. https://doi.org/10.1002/rrq.116

Fricke, S, Szczerbinski, M., Stackhouse, J., & Fox-Boyer, A. V. (2008). Predicting individual differences in early literacy acquisition in German: The role of speech and language processing skills and letter knowledge. *Written*

Language & Literacy, *11*(2), 103–146. https://doi.org/10.1075/wll.11.2.02fri

Frith, U. (1985). Beneath the surface of developmental dyslexia. In K. Patterson, J. Marshall, & M. Coltheart (Eds.), *Surface dyslexia, neuropsychological and cognitive studies of phonological reading* (pp. 301–330). London: Erlbaum.

Frost, R. (2005). Orthographic systems and skilled word recognition processes in reading. In Margaret J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 272–295). Oxford: Blackwell Publishing.

Furtak, E. M., Ruiz-Primo, M. A., Shemwell, J. T., Ayala, C. C., Brandon, P. R., Shavelson, R. J., & Yin, Y. (2008). On the fidelity of implementing embedded formative assessments and its relation to student learning. *Applied Measurement in Education*, *21*(4), 360–389. https://doi.org/10.1080/08957340802347852

Ganzeboom, H. B., De Graaf, P. M., & Treiman, D. J. (1992). A standard international socio-economic index of occupational status. *Social Science Research*, *21*(1), 1–56. https://doi.org/10.1016/0049-089X(92)90017-B

Gathercole, S. E, Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The Structure of Working Memory from 4 to 15 Years of Age. *Developmental Psychology*, *40*(2), 177–190. https://doi.org/10.1037/0012-1649.40.2.177

Gathercole, S. E., Alloway, T. P., Willis, C., & Adams, A. M. (2006). Working memory in children with reading disabilities. *Journal of Experimental Child Psychology*, *93*(3), 265–281. https://doi.org/10.1016/j.jecp.2005.08.003

Genesee, F. (2006). Bilingual first language acquisition in perspective. In P. McCardle & E. Hoff (Eds.), *Childhood bilingualism: Research on infancy through school age* (pp. 45–67). Clevedon: Multilingual Matters.

Georgiou, G. K., Parrila, R., & Liao, C. H. (2008). Rapid naming speed and reading across languages that vary in orthographic consistency. *Reading and Writing*, *21*(9), 885–903. https://doi.org/10.1007/s11145-007-9096-4

Georgiou, G. K., Parrila, R., & Papadopoulos, T. C. (2016). The anatomy of the RAN-reading relationship. *Reading and Writing*, *29*(9), 1793–1815. https://doi.org/10.1007/s11145-016-9653-9

Georgiou, G. K., Torppa, M., Landerl, K., Desrochers, A., Manolitsis, G., de Jong, P. F., & Parrila, R. (2019). Reading and Spelling Development Across Languages Varying in Orthographic Consistency: Do Their Paths Cross? *Child Development*, *00*(0), 1–14. https://doi.org/10.1111/cdev.13218

Georgiou, G. K., Torppa, M., Manolitsis, G., Lyytinen, H., & Parrila, R. (2012). Longitudinal predictors of reading and spelling across languages varying in orthographic consistency. *Reading and Writing*, *25*(2), 321–346. https://doi.org/10.1007/s11145-010-9271-x

Geva, E., & Farnia, F. (2012). Developmental changes in the nature of language proficiency and reading fluency paint a more complex view of reading comprehension in ELL and EL1. *Reading and Writing*, *25*(8), 1819–1845. https://doi.org/10.1007/s11145-011-9333-8

Geva, E., & Yaghoub Zadeh, Z. (2006). Reading efficiency in native English-speaking and English-as-a-second- language children: The role of oral proficiency and underlying cognitive-linguistic processes. *Scientific Studies of*

Reading, *10*(1), 31–57. https://doi.org/10.1207/s1532799xssr1001_3

Gholamain, M., & Geva, E. (1999). Orthographic and cognitive factors in the concurrent development of basic reading skills in English and Persian. *Langauge Learning*, *49*(2), 183–217.

Gilles, P. (2014). Phonological domains in Luxembourgish and their relevance for the phonological system. In J. Caro Reina & R. Szczepaniak (Eds.), *Syllable and Word Languages* (pp. 1–22). Berlin: de Gruyter. https://doi.org/10.1515/9783110346992.279

Gilles, P., & Trouvain, J. (2013). Luxembourgish. *Journal of the International Phonetic Association*, *43*(1), 67–74. https://doi.org/10.1017/S0025100312000278

Gillon, G. T. (2004). *Phonological awareness: From research to practice*. New York: Guilford Press.

Goodrich, J. M., & Lonigan, C. J. (2017). Language-independent and language-specific aspects of early literacy: An evaluation of the common underlying proficiency model. *Journal of Educational Psychology*, *109*(6), 782. https://doi.org/10.1037/edu0000179

Goodrich, J. M., Lonigan, C. J., & Farver, J. M. (2013). Do Early Literacy Skills in Children's First Language Promote Development of Skills in Their Second Language? An Experimental Evaluation of Transfer. *Journal of Educational Psychology*, *105*(2), 414–426. https://doi.org/10.1021/nl061786n.Core-Shell

Goswami, U., & Bryant, P. (1990). *Phonological skills and learning to read*. London: Lawrence Erlbaum.

Goswami, U. (1986). Children's use of analogy in learning to read: A developmental study. *Journal of Experimental Child Psychology*, *42*(1), 73–83. https://doi.org/10.1016/0022-0965(86)90016-0

Goswami, U. (2002). Phonology, reading development, and dyslexia: A cross-linguistic perspective. *Annals of Dyslexia*, *52*(1), 139–163. https://doi.org/10.1007/s11881-002-0010-0

Gough, P. B., & Tunmer, W. (1986). Decoding, reading, and reading disability. *Remedial and Special Education*, *7*(1), 6–10. https://doi.org/10.1177/074193258600700104

Graham, S. (2000). Should the natural learning approach replace spelling instruction? *Journal of Educational Psychology*, *92*(2), 235–247. https://doi.org/10.1037/0022-0663.92.2.235

Graham, S., & Hebert, M. (2011). *Writing to Read: A Meta-Analysis of the Impact of Writing and Writing Instruction on Reading*. *Harvard Educational Review* (Vol. 81, pp. 710–744). https://doi.org/10.17763/haer.81.4.t2k0m13756113566

Graham, S., & Santangelo, T. (2014). Does spelling instruction make students better spellers, readers, and writers? A meta-analytic review. *Reading and Writing*. https://doi.org/10.1007/s11145-014-9517-0

Green, L. W., & Glasgow, R. E. (2006). Evaluating the Relevance, Generalization, and Applicability of Research. *Evaluation & the Health Professions*, *29*(1), 126–153. https://doi.org/10.1177/0163278705284445

Grosjean, F. (2010). *Bilingual: Life and reality*. Cambridge, MA: Harvard University.

Gupta, S. K. (2011). Intention-to-treat concept: a review. Perspectives in clinical research. *Perspectives in Clinical Reseach*, *2*(3), 109–112.

Hammer, C. S., Jia, G., & Uchikoshi, Y. (2011). Language and Literacy Development of Dual Language Learners Growing Up in the United States: A Call for Research. *Child Development Perspective*, *5*(1), 4–9. https://doi.org/10.1016/j.surg.2006.10.010.Use

Hammill, D. (2004). What We Know About Correlates of Reading. *Exceptional Children*, *70*(4), 453–468. https://doi.org/10.1177/001440290407000405

Harlaar, N., Dale, P. S., & Plomin, R. (2007). Reading exposure: A (largely) environmental risk factor with environmentally-mediated effects on reading performance in the primary school years. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *48*(12), 1192–1199. https://doi.org/10.1111/j.1469-7610.2007.01798.x

Harm, M. W., & Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia: Insights from connectionist models. *Psychological Review*, *106*(3), 491–528. https://doi.org/10.1037/0033-295X.106.3.491

Harrison, G. L., Goegan, L. D., Jalbert, R., McManus, K., Sinclair, K., & Spurling, J. (2016). Predictors of spelling and writing skills in first- and second-language learners. *Reading and Writing*, *29*(1), 69–89. https://doi.org/10.1007/s11145-015-9580-1

Hatcher, P. J., Duff, F. J., & Hulme, C. (2000). *Sound linkage: An integrated programme for overcoming reading difficulties* (2nd ed.). London: Whurr.

Hatcher, P. J, Hulme, C., & Ellis, A. W. (1994). Ameliorating Early Reading Failure by Integrating the Teaching of Reading and Phonological Skills : The Phonological Linkage Hypothesis. *Child Development*, *65*(1), 41–57.

Hatcher, P. J, Hulme, C., & Ellis, A. W. (2009). Ameliorating Early Reading Failure by Integrating the Teaching of Reading and Phonological Skills : The Phonological Linkage Hypothesis. *Child Development*, *65*(1), 41–57.

Hatcher, P. J, Hulme, C., & Snowling, M. J. (2004). Explicit phoneme training combined with phonic reading instruction helps young children at risk of reading failure. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *45*(2), 338–358.

Hipfner-Boucher, K., Milburn, T., Weitzman, E., Greenberg, J., Pelletier, J., & Girolametto, L. (2014). Relationships between preschoolers' oral language and phonological awareness. *First Language*, *34*(2), 178–197. https://doi.org/10.1177/0142723714525945

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), 1–27. https://doi.org/10.1017/S0305000910000759.Dual

Hoff, E. (2006). How social contexts support and shape language development. *Developmental Review*, *26*(1), 55–88. https://doi.org/10.1016/j.dr.2005.11.002

Hoff, E. (2013). Interpreting the early language trajectories of children from low-SES and language minority homes: implications for closing achievement gaps. *Developmental Psychology*, *49*(1), 4–14. https://doi.org/10.1037/a0027238.Interpreting

Hoff, E, & Tian, C. (2005). Socioeconomic status and cultural influences on language. *Journal of Communication Disorders*, *38*(4 SPEC. ISS.), 271–278. https://doi.org/10.1016/j.jcomdis.2005.02.003

Hoffmann, C., Hornung, C., Gamo, S., Esch, P., Keller, U., & Fischbach, A. (2018). *National Educational Report 2018: Schulische Kompetenzen von Erstklässlern und ihre Entwicklung nach zwei Jahren*. *Nationaler Bildungsbericht Luxemburg 2018 [National Education Report].* Luxembourg : University of Luxembourg & Service de Coordination de la Recherche et de l´Innovation pédagogiques et technologiques.

Hofslundsengen, H., Gustafsson, J., & Hagtvet, B. E. (2018). Contributions of the Home Literacy Environment and Underlying Language Skills to Preschool Invented Writing. *Scandinavian Journal of Educational Research*, *00*(0), 1–17. https://doi.org/10.1080/00313831.2017.1420686

Hoover, W., & Gough, B. (1990). The Simple View of Reading. *Reading and Writing*, *2*(2), 127–160.

Huang, F. L., Tortorelli, L. S., & Invernizzi, M. A. (2014). An investigation of factors associated with letter-sound knowledge at kindergarten entry. *Early Childhood Research Quarterly*, *29*(2), 182–192.

Huguenin, C., & Dubois, O. (2013). *Die Alphas: Mein Alphabet lebt. Mit allen Sinnen lesen lernen für alle Kinder von 4 - 7 Jahren!* Braunschweig: Georg Westermann Verlag.

Hulme, C., Bowyer-Crane, C., Carroll, J. M., Duff, F. J., & Snowling, M. J. (2012). The Causal Role of Phoneme Awareness and Letter-Sound Knowledge in Learning to Read: Combining Intervention Studies With Mediation Analyses. *Psychological Science*, *23*(6), 572–577. https://doi.org/10.1177/0956797611435921

Hulme, C., Nash, H. M., Gooch, D., Lervåg, A., & Snowling, M. J. (2015). The Foundations of Literacy Development in Children at Familial Risk of Dyslexia. *Psychological Science*, *26*(12), 1877–1886. https://doi.org/10.1177/0956797615603702

Hulme, C., Snowling, M., Caravolas, M., & Carroll, J. M. (2005). Phonological skills are (probably) one cause of success in learning to read: A comment on Castles and Coltheart. *Scientific Studies of ReadingStudies of Reading*, *9*(4), 351–365.

Hulme, C., & Snowling, M. J. (2009). *Developmental Disorders of Language Learning and Cognition*. Oxford, England: Wiley Blackwell.

Hulme, C., & Snowling, M. J. (2011). Children's reading comprehension difficulties: Nature, causes, and treatments. *Current Directions in Psychological Science*, *20*(3), 139–142. https://doi.org/10.1177/0963721411408673

Huo, S., & Wang, S. (2017). The Effectiveness of Phonological-Based Instruction in English As a Foreign Language Students at Primary School Level: A Research Synthesis. *Frontiers in Education*, *2*, 1–13. https://doi.org/10.3389/feduc.2017.00015

Innocenti, M. S., Coyne, M., Gersten, R., Greenwood, C., Compton, D., & Fuchs, L. S. (2014). Quality Indicators for Group Experimental and Quasi-Experimental Research in Special Education. *Exceptional Children*, *71*(2), 149–164. https://doi.org/10.1177/001440290507100202

Jean, M., & Geva, E. (2009). The development of vocabulary in English as a second language children and its role in predicting word recognition ability. *Applied Psycholinguistics*, *30*(01), 153–185. https://doi.org/doi:10.1017/S0142716408090073

Jongejan, W., Verhoeven, L., & Siegel, L. S. (2007). Predictors of Reading and Spelling Abilities in First- and Second-Language Learners. *Journal of Educational Psychology*, *99*(4), 835–851. https://doi.org/10.1037/0022-0663.99.4.835

Joshi, M. R. (2005). Vocabulary: A Critical Component of Comprehension. *Reading & Writing Quarterly*, *21*(3), 209–219. https://doi.org/10.1080/10573560590949278

Joshi, R. M., & Aaron, P. G. (2000). The component model of reading: Simple view of reading made a little more complex. *Reading Psychology*, *21*(2), 85–97. https://doi.org/10.1080/02702710050084428

Joshi, R. M, Tao, S., Aaron, P. G., & Quiroz, B. (2012). Cognitive Component of Componential Model of Reading Applied to Different Orthographies. *Journal of Learning Disabilities*, *45*(5), 480–486. https://doi.org/10.1177/0022219411432690

Journals, P. E., Baumert, J., Nagy, G., Lehmann, R., Pfost, M., Hattie, J., … Reschly, A. L. (2011). Cumulative Advantages and the Emergence of Social and Ethnic Inequality: Matthew Effects in Reading and Mathematics Development Within Elementary Schools? *Journal of Learning Disabilities*, *189*(1), 499–514. https://doi.org/10.1177/0022219411410042

Juel, C. (1983). The Development and Use of Mediated Word Identification. *Reading Reasearch Quarterly*, *18*(3), 306–327. https://doi.org/10.2307/747391

Kaptchuk, T. (2001). Gold standard or golden calf ? *Journal of Clinical Epidemiology*, *54*, 541–549. https://doi.org/10.1016/S0895-4356(00)00347-4

Kendeou, P., Savage, R., & Broek, P. Van Den. (2009). Revisiting the simple view of reading. *British Journal of Educational Psychology*, *79*, 353–370. https://doi.org/10.1348/978185408X369020

Kendeou, P., van den Broek, P., White, M. J., & Lynch, J. S. (2009). Predicting Reading Comprehension in Early Elementary School: The Independent Contributions of Oral Language and Decoding Skills. *Journal of Educational Psychology*, *101*(4), 765–778. https://doi.org/10.1037/a0015956

Kershaw, S., & Schatschneider, C. (2012). A latent variable approach to the simple view of reading. *Reading and Writing*, *25*(2), 433–464. https://doi.org/10.1007/s11145-010-9278-3

Kieffer, M. J., & Vukovic, R. K. (2013). Growth in reading-related skills of language minority learners and their classmates : more evidence for early identification and intervention. *Reading and Writing*, *26*(7), 1159–1194.

https://doi.org/10.1007/s11145-012-9410-7

Kim, H. Y. (2013). Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. *Restorative Dentistry & Endodontics*, *38*(1), 52. https://doi.org/10.5395/rde.2013.38.1.52

Kim, Y. S., Petscher, Y., Foorman, B. R., & Zhou, C. (2010). The Contributions of Phonological Awareness and Letter-Name Knowledge to Letter-Sound Acquisition-A Cross-Classified Multilevel Model Approach. *Journal of Educational Psychology*, *102*(2), 313–326. https://doi.org/10.1037/a0018449

Kirby, J. R., & Savage, R. S. (2008). Can the simple view deal with the complexities of reading? *Literacy*, *42*(2), 75–82. https://doi.org/10.1111/j.1741-4369.2008.00487.x

Kirsch, C. (2018). *National Educational Report 2018: Mehrsprachige Bildung in der frühen Kindheit*. Luxembourg : University of Luxembourg & Service de Coordination de la Recherche et de l´Innovation pédagogiques et technologiques.

Klein, O., Biedinger, N., & Becker, B. (2014). The effect of reading aloud daily-Differential effects of reading to native-born German and Turkish-origin immigrant children. *Research in Social Stratification and Mobility*, *38*, 43–56. https://doi.org/10.1016/j.rssm.2014.06.001

Kline, R. B. (2015). *Principles and practice of structural equation modeling* (3rd ed.). New York: Guilford Press.

Knudsen, E. I., Heckman, J. J., Cameron, J. L., & Shonkoff, J. P. (2006). Economic, neurobiological, and behavioral perspectives on building America's future workforce. *Proceedings of the National Academy of Sciences*, *103*(27), 10155–10162. https://doi.org/10.1037/a0025382

Kudo, M. F., Lussier, C. M., & Swanson, H. L. (2015). Reading disabilities in children: A selective meta-analysis of the cognitive literature. *Research in Developmental Disabilities*, *40*, 51–62. https://doi.org/10.1016/j.ridd.2015.01.002

Kuo, L. J., & Anderson, R. C. (2012). Effects of early bilingualism on learning phonological regularities in a new language. *Journal of Experimental Child Psychology*, *111*(3), 455–467. https://doi.org/10.1016/j.jecp.2011.08.013

Küspert, P., & Schneider, W. (2008). *Hören, lauschen, lernen: Sprachspiele für Kinder im Vorschulalter; Würzburger Trainingsprogramm zur Vorbereitung auf den Erwerb der Schrift sprache.* Göttingen: Vandenhoeck & Ruprecht.

Lallier, M., Valdois, S., Lassus-Sangosse, D., Prado, C., & Kandel, S. (2014). Impact of orthographic transparency on typical and atypical reading development: Evidence in French-Spanish bilingual children. *Research in Developmental Disabilities*, *35*(5), 1177–1190. https://doi.org/10.1016/j.ridd.2014.01.021

Landerl, K., & Thaler, V. (2006). Reading and spelling acquisition and dyslexia in German. In R. M Joshi & P. G. Aaron (Eds.), *Handbook of orthography and literacy* (pp. 121–134). Mahwah, NJ: Lawrence Erlbaum Associates.

Landerl, K., Freudenthaler, H. H., Heene, M., De Jong, P. F., Desrochers, A., Manolitsis, G., … Georgiou, G. K. (2019). Phonological Awareness and Rapid

Automatized Naming as Longitudinal Predictors of Reading in Five Alphabetic Orthographies with Varying Degrees of Consistency. *Scientific Studies of Reading*, *23*(3), 220–234. https://doi.org/10.1080/10888438.2018.1510936

Landerl, K., Ramus, F., Moll, K., Lyytinen, H., Leppänen, P. H. T., Lohvansuu, K., … Schulte-Körne, G. (2013). Predictors of developmental dyslexia in European orthographies with varying complexity. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *54*(6), 686–694. https://doi.org/10.1111/jcpp.12029

Landerl, K., & Wimmer, H. (2008). Development of Word Reading Fluency and Spelling in a Consistent Orthography: An 8-Year Follow-Up. *Journal of Educational Psychology*, *100*(1), 150–161. https://doi.org/10.1037/0022-0663.100.1.150

Law, J. M., Vandermosten, M., Ghesquière, P., & Wouters, J. (2017). Predicting future reading problems based on pre-reading auditory measures: A longitudinal study of children with a familial risk of dyslexia. *Frontiers in Psychology*, *8*(FEB). https://doi.org/10.3389/fpsyg.2017.00124

Lenhard, W., & Schneider, W. (2006). *ELFE 1-6 Ein Leseverständnistest für Erst- bis Sechsklässler. [ELFE 1-6 A reading comprehension test for students from the 1st through the 6th grade]*. Göttingen: Hogrefe.

Lenz, T., & Heinz, A. (2018). *Das Luxemburgische Schulsystem*. Luxembourg : University of Luxembourg & Service de Coordination de la Recherche et de l´Innovation pédagogiques et technologiques.

Leppänen, U., Aunola, K., Niemi, P., & Nurmi, J. E. (2008). Letter knowledge predicts Grade 4 reading fluency and reading comprehension. *Learning and Instruction*, *18*(6), 548–564. https://doi.org/10.1016/j.learninstruc.2007.11.004

Lerner, M. D., & Lonigan, C. J. (2016). Bidirectional Relations between Phonological Awareness and Letter Knowledge in Preschool Revisited: A Growth Curve Analysis of the Relation between Two Code-Related Skills. *Journal of Experimental Child Psychology*, (144), 166–183. https://doi.org/10.1158/1940-6207

Lervåg, A., & Aukrust, V. G. (2010). Vocabulary knowledge is a critical determinant of the difference in reading comprehension growth between first and second language learners. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *51*(5), 612–620. https://doi.org/10.1111/j.1469-7610.2009.02185.x

Lervåg, A., Bråten, I., & Hulme, C. (2009). The Cognitive and Linguistic Foundations of Early Reading Development: A Norwegian Latent Variable Longitudinal Study. *Developmental Psychology*, *45*(3), 764–781. https://doi.org/10.1037/a0014132

Lervåg, A. & Hulme, C. (2009). Rapid Automatized Naming (RAN) Taps a Mechanism That Places Constraints on the Development of Early Reading Fluency. *Psychological Science*, *20*(8), 1040–1048. Retrieved from http://dx.doi.org/10.1111/j.1467-9280.2009.02405.x

Lervåg, A., Hulme, C., & Melby-Lervåg, M. (2017). Unpicking the Developmental Relationship Between Oral Language Skills and Reading Comprehension: It's Simple, But Complex. *Child Development*, pp. 1–18.

https://doi.org/10.1111/cdev.12861

Lesaux, N. K., Geva, E., Koda, K., Siegel, L. S., & Timothy, S. (2008). Development of Literacy in Second-Language Learners. In A. August & T. Shanahan (Eds.), *Developing literacy in second-language learners: Report of the National Literacy Panel on Language-Minority Children and Youth* (pp. 27–60). Mahwah, NJ: Lawrence Erlbaum Associates.

Lesaux, N. K., Rupp, A. A., & Siegel, L. S. (2007). Growth in reading skills of children from diverse linguistic backgrounds- Findings from a 5-year longitudinal study. *Journal of Educational Psychology*, *99*(4), 821–834.

Lesaux, N. K., & Siegel, L. S. (2003). The Development of Reading in Children Who Speak English as a Second Language. *Developmental Psychology*, *39*(6), 1005–1019. https://doi.org/10.1037/0012-1649.39.6.1005

Lesaux, N. K, Crosson, A. C., Kieffer, M. J., & Pierce, M. (2011). Uneven Profiles: Language Minority Learners' Word Reading. *Journal of Applied Developmental Psychology*, *31*(6), 475–483. https://doi.org/10.1016/j.appdev.2010.09.004.Uneven

Leseman, P. P. M., & Jong, P. F. (1998). Home Literacy: Opportunity, Instruction, Cooperation and Social-Emotional Quality Predicting Early Reading Achievement. *Reading Research Quarterly*, *33*(3), 294–318. https://doi.org/10.1598/rrq.33.3.3

Levy, B. A., Gong, Z., Hessels, S., Evans, M. A., & Jared, D. (2006). Understanding print: Early reading development and the contributions of home literacy experiences. *Journal of Experimental Child Psychology*, *93*(1), 63–93. https://doi.org/10.1016/j.jecp.2005.07.003

Lindsey, K. A., Manis, F. R., & Bailey, C. E. (2003). Prediction of First-Grade Reading in Spanish-Speaking English-Language Learners. *Journal of Educational Psychology*, *95*(3), 482–494. https://doi.org/10.1037/0022-0663.95.3.482

Lipka, O., & Siegel, L. S. (2007). The Development of Reading Skills in Children With English as a Second Language. *Scientific Studies of Reading*, *11*(2), 105–131. https://doi.org/10.1080/10888430709336555

Little, R. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*. https://doi.org/10.1080/01621459.1988.10478722

Lloyd, S. (1998). *The Phonics Handbook*. Chigwell: Jolly Learning Ltd.

Loizou, M., & Stuart, M. (2003). Phonological awareness in monolingual and bilingual English and Greek five-year-olds. *Journal Of Research In Reading*, *26*(1), 3–18. https://doi.org/doi:10.1111/1467-9817.261002

Lonigan, C J, Burgess, S. R., & Anthony, J. L. (2000). Development of emergent literacy and early reading skills in preschool children: evidence from a latent-variable longitudinal study. *Developmental Psychology*, *36*(5), 596–613. https://doi.org/Doi 10.1037//0012-1649.36.5.596

Lonigan, C. J., Farver, J. M., Nakamoto, J., & Eppe, S. (2013). Developmental trajectories of preschool early literacy skills: A comparison of language-minority and monolingual-english children. *Developmental Psychology*, *49*(10), 1943–1957. https://doi.org/10.1037/a0031408

Lonigan, C. J., Purpura, D. J., Wilson, S. B., Walker, P. M., & Clancy-Menchetti, J. (2013). Evaluating the components of an emergent literacy intervention for preschool children at risk for reading difficulties. *Journal of Experimental Child Psychology*, *114*(1), 111–130. https://doi.org/10.1016/j.jecp.2012.08.010

López, L. M., & Greenfield, D. B. (2004). The cross-language transfer of phonological skills of hispanic head start children. *Bilingual Research Journal*, *28*(1), 1–18. https://doi.org/10.1080/15235882.2004.10162609

Lovett, M. W., De Palma, M., Frijters, J., Steinbach, K., Temple, M., Benson, N., & Lacerenza, L. (2008). Interventions for reading difficulties: A comparison of response to intervention by ELL and EFL struggling readers. *Journal of Learning Disabilities*, *41*(4), 333–352. https://doi.org/10.1177/0022219408317859

Lukie, I. K., Skwarchuk, S. L., LeFevre, J. A., & Sowinski, C. (2014). The Role of Child Interests and Collaborative Parent-Child Interactions in Fostering Numeracy and Literacy Development in Canadian Homes. *Early Childhood Education Journal*, *42*(4), 251–259. https://doi.org/10.1007/s10643-013-0604-7

Lundberg, I. (1994). Reading difficulties can be predicted and prevented: A Scandinavian perspective on phonological awareness and reading. In C. Hulme & M. J. Snowling (Eds.), *Reading development and dyslexia*. London: Whurr.

Lundberg, I., Frost, J., & Petersen, O. P. (1988). Effects of an extensive program for stimulating phonological awareness in preschool children. *Reading Research Quarterly*, *23*(3), 263–284.

Maclean, M., Bryant, P., & Bradley, L. (1987). Rhymes, nursery rhymes, and reading in early childhood. *Merrill-Palmer Quarterly*, *33*(3), 255–281. https://doi.org/10.2307/23086536

Mancilla-Martinez, J., & Lesaux, N. K. (2010). Predictors of Reading Comprehension for Struggling Readers: The Case of Spanish-speaking Language Minority Learners. *Journal of Educational Psychology*, *102*(3), 701–711. https://doi.org/10.1037/a0019135.Predictors

Mangione, T. W. (1995). *Mail Surveys. Improving the Quality*. London: SAGE.

Manis, F. R., Lindsey, K. A., & Bailey, C. E. (2004). Development of reading in grades K–2 in Spanish‐speaking English‐language learners. *Learning Disabilities Research & Practice*, *19*(4), 214–224.

Manis, R. M., Seidenberg, M. S., & Doi, L. M. (1999). See Dick RAN: Rapid Naming and the Longitudinal Prediction of Reading Subskills in First and Second Graders. *Scientific Studies of Reading*, *3*(2), 129–157. https://doi.org/10.1207/s1532799xssr0302

Mann, V., & Wimmer, H. (2002). Phoneme awareness and pathways into literacy: A comparison of German and American children. *Reading and Writing: An Interdisciplinary Journal*, *15*(7/8), 653–682. https://doi.org/10.1023/A:1020984704781

Martin, R., & Brunner, M. (2012). *Épreuves Standardisées. Nationaler Bericht 2011-2012*. Luxembourg.

Masoura, E. V., & Gathercole, S. E. (1999). Phonological Short-term Memory and Foreign Language Learning. *International Journal of Psychology*, *34*(5–6), 383–388. https://doi.org/10.1080/002075999399738

Maxwell, S. E., Kelley, K., & Rausch, J. R. (2008). Sample Size Planning for Statistical Power and Accuracy in Parameter Estimation. *Annual Review of Psychology*, *59*(1), 537–563. https://doi.org/10.1146/annurev.psych.59.103006.093735

May, P. (2002). *Hamburger Schreibprobe (HSP 1+) [Hamburg Writing Sample]*. Hamburg: Verlag für pädagogische Medien.

Mayo, A. Y., & Leseman, P. (2008). Off to a good start?: Vocabulary development and differences in early family and classroom experiences of children from native-Dutch and immigrant families in the Netherlands. *Educational and Child Psychology*, *25*(3), 70–83. https://doi.org/10.4135/9781446274750.n5

McCarney, R., Warner, J., Iliffe, S., Van Haselen, R., Griffin, M., & Fisher, P. (2007). The Hawthorne Effect: A randomised, controlled trial. *BMC Medical Research Methodology*, *7*, 1–8. https://doi.org/10.1186/1471-2288-7-30

McDowell, K. D., Lonigan, C. J., & Goldstein, H. (2007). Relations Among Socioeconomic Status, Age, and Predictors of Phonological Awareness. *Journal of Speech Language and Hearing Research*, *50*(4), 1079. https://doi.org/10.1044/1092-4388(2007/075)

McMaster, K. L., Fuchs, D., Fuchs, L. S., & Compton, D. L. (2005). Responding to nonresponders: An experimental field trial of identification and intervention methods. *Exceptional Children*, *71*(4), 445–463. https://doi.org/10.1177/001440290507100404

Melby-Lervåg, M., & Lervåg, A. (2011). Cross-linguistic transfer of oral language, decoding, phonological awareness and reading comprehension: A meta-analysis of the correlational evidence. *Journal of Research in Reading*, *34*(1), 114–135. https://doi.org/10.1111/j.1467-9817.2010.01477.x

Melby-Lervåg, M., & Lervåg, A. (2014). Reading comprehension and its underlying components in second-language learners: A meta-analysis of studies comparing first- and second-language learners. *Psychological Bulletin*, *140*(2), 409–433. https://doi.org/10.1037/a0033890

Melby-Lervåg, M., Lyster, S. A., & Hulme, C. (2012). Phonological skills and their role in learning to read: A meta-analytic review. *Psychological Bulletin*, *138*(2), 322–352. https://doi.org/10.1037/a0026744

MENFP. (2011). *Plan d'Études École Fondamentale*. Luxembourg : Ministère de l'Education Nationale et de la Formation Professionnelle.

MENFP. (2015). *Les Chiffres Clés de l'Éducation nationale*. Luxembourg.

MENJE: Ministère de l'Éducation nationale de l'Enfance et de la Jeunesse. (2017). *Statistiques globales et analyses des résultats scholaires: Enseignement fondamental. Cycles 1 à 4, Année scolaire 2015/2016.* Luxembourg : MENJE.

Moll, K., & Landerl, K. (2010). *Salzburger Lese- und Rechtschreibtest. [Salzburg* *reading and spelling test]*. Bern: Hans Huber.

Moll, K., Fussenegger, B., Willburger, E., & Landerl, K. (2009). RAN is not a measure of orthographic processing. Evidence from the asymmetric German orthography. *Scientific Studies of Reading*, *13*(1), 1–25. https://doi.org/10.1080/10888430802631684

Moll, K., Thompson, P. A., Mikulajova, M., Jagercikova, Z., Kucharska, A., Franke, H., … Snowling, M. J. (2016). Precursors of Reading Difficulties in Czech and Slovak Children At-Risk of Dyslexia. *Dyslexia*, *22*(2), 120–136. https://doi.org/10.1002/dys.1526

Morris, Z. S., Wooding, S., & Grant, J. (2011). The answer is 17 years, what is the question. *Journal of the Royal Society of Medicine*, *104*(12), 510–520. https://doi.org/10.1258/jrsm.2011.110180

Morton, S. M. B., Bandara, D. K., Robinson, E. M., & Atatoa Carr, P. E. (2012). In the 21st Century, what is an acceptable response rate? *Australian and New Zealand Journal of Public Health*, *36*(2), 106–108. https://doi.org/10.1111/j.1753-6405.2012.00854.x

Murphy, V. A., & Unthiah, D. (2015). *A systematic review of intervention research examining English language and literacy development in children with English as an additional language (EAL)*. Report for the The Education Endowment Foundation (EEF), Unbound Philanthropy and The Bell Foundation.

Muter, V., & Diethelm, K. (2001). The contribution of phonological skills and letter knowledge to early reading development in a multilingual population. *Language Learning*, *51*(2), 187–219. https://doi.org/10.1111/1467-9922.00153

Muter, V., Hulme, C., Snowling, M. J., & Stevenson, J. (2004). Phonemes, Rimes, Vocabulary, and Grammatical Skills as Foundations of Early Reading Development: Evidence From a Longitudinal Study. *Developmental Psychology*, *40*(5), 665–681. https://doi.org/10.1037/0012-1649.40.5.665

Nation, K. (2019). Children's Reading Difficulties, Language, and Reflections on the Simple View of Reading. *PsyARiv*, 1–50.

Nation, K., & Hulme, C. (1997). Phonemic Segmentation, Not Onset-Rime Segmentation, Predicts Early Reading and Spelling Skills. *Reading Reseearch Quarterly*, *32*(2), 154–167.

Nation, K., & Snowling, M. (1997). Assessing reading difficulties: the validity and utility. *British Journal of Educational Psychology*, *67*(3), 359–370.

Nation, K., & Snowling, M. J. (2004). Beyond phonological skills: Broader language skills contribute to the development of reading. *Journal of Research in Reading*, *27*(4), 342–356. https://doi.org/10.1111/j.1467-9817.2004.00238.x

National Early Literacy Panel. (2008). *Developing early literacy: report of the national early literacy panel*. Washington.

National Reading Panel. (2000). *Teaching children to read. Reports of the subgroups.* Washington, DC: National Institute of Child Health and Human Development. https://doi.org/10.1002/ppul.1950070418

Netten, A., Droop, M., & Verhoeven, L. (2011). Predictors of reading literacy for first and second language learners. *Reading and Writing*, *24*(4), 413–425.

https://doi.org/10.1007/s11145-010-9234-2

Network Promising Practices. (2007). *How programs are considered*. Available from: http://www.promisingpractices.ne t/criteria.asp [last accessed 10 September 2018].

NICHD Early Child Care Research Network. (2005). Pathways to reading: The role of oral language in the transition to reading. *Developmental Neuropsychology, 41*(2), 428–442.

Niklas, F., & Schneider, W. (2013). Home Literacy Environment and the beginning of reading and spelling. *Contemporary Educational Psychology*, *38*(1), 40–50. https://doi.org/10.1016/j.cedpsych.2012.10.001

Norton, E. S., & Wolf, M. (2011). Rapid Automatized Naming (RAN) and Reading Fluency: Implications for Understanding and Treatment of Reading Disabilities. *Annual Review of Psychology*, *63*(1), 427–452. https://doi.org/10.1146/annurev-psych-120710-100431

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.

O'Connor, R., Bocian, K., Beebe-Frankenberger, M., & Linklater, D. . (2010). Responsiveness of Students With Language Difficulties to Early Intervention in Reading. *Journal of Special Education*, *43*(4), 220–235.

Oakhill, J., Cain, K., & Elbro, C. (2015). *Understanding and teaching reading comprehension*. *English Teaching Forum* (Vol. 50). London: Routledge. Retrieved from http://files.eric.ed.gov/fulltext/EJ971235.pdf

Oakhill, J. V., & Cain, K. (2012). The precursors of reading ability in young readers: Evidence from a four-year longitudinal study. *Scientific Studies of Reading*, *16*(2), 91–121. https://doi.org/10.1080/10888438.2010.529219

OECD. (2012). *PISA 2012 Results in Focus*. Paris.

OECD. (2018). *The Resilience of Students with an Immigrant Background*. OECD Reviews of Migrant Education, OECD Publishing, Paris: OECD. https://doi.org/10.1787/9789264292093-en

Oxley, E., & de Cat, C. (2019). A systematic review of language and literacy interventions in children and adolescents with English as an additional language (EAL). *Language Learning Journal*, 1–23. https://doi.org/10.1080/09571736.2019.1597146

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

Páez, M. M., Tabors, P. O., & López, L. M. (2007). Dual language and literacy development of Spanish-speaking preschool children, *28*, 85–102. https://doi.org/10.1016/j.appdev.2006.12.007

Panel, N. E. L. (2008). *Developing early literacy: Report of the National Early Literacy Panel*. Washington, DC: National Institute for Literacy.

Papadopoulos, T. C., & Georgiou, G. K. (2009). Investigating the Double-Deficit Findings From a Longitudinal Study. *Journal of Learning Disabilities*, *42*(8037), 528–548.

Papadopoulos, T. C., Spanoudis, G. C., & Georgiou, G. K. (2016). How is RAN related to reading fluency? A comprehensive examination of the prominent theoretical accounts. *Frontiers in Psychology*, *7*(1217), 1–15. https://doi.org/10.3389/fpsyg.2016.01217

Peng, P., Barnes, M., Wang, C. C., Wang, W., Li, S., Swanson, H. L., … Tao, S. (2018). Meta-analysis on the relation between reading and working memory. *Psychological Bulletin*, *144*(1), 48–76. https://doi.org/10.1037/bul0000124

Perfetti, C. A. (1997). The psycholinguistics of spelling and reading. In CA Perfetti, L. Rieben, & M. Fayol (Eds.), *Learning to spell: Research, theory, and practice across languages* (pp. 21–38). Mahwah, NJ: Lawrence Erlbaum.

Perfetti, C. A., Landi, N., & Oakhill, J. (2005). The acquisition of reading comprehension. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 227–247). Oxford, England: Blackwell.

Perfetti, C. A. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, *11*(4), 357–383. https://doi.org/10.1080/10888430701530730

Peterson, R. L., & Pennington, B. F. (2012). Developmental dyslexia. In *The Lancet* (Vol. 379, pp. 1997–2007). Elsevier Ltd. https://doi.org/10.1016/S0140-6736(12)60198-6

Phillips, B. M., & Torgesen, J. K. (2006). Phonemic awareness and reading: Beyond the growth of initial reading accuracy. In S. B. Neuman & D. K. Dickinson (Eds.), *Handbook of early literacy research* (2nd ed.). Chicago: Guilford Press.

Piasta, S. B., Purpura, D. J., & Wagner, R. K. (2010). Fostering alphabet knowledge development: A comparison of two instructional approaches. *Reading and Writing*, *23*(6), 607–626. https://doi.org/10.1007/s11145-009-9174-x

Piasta, S. B., & Wagner, R. K. (2010). Developing early literacy skills- A meta-analysis of alphabet learning and instruction. *Reading Research Quarterly*, *45*(1), 8–38. https://doi.org/10.1598/RRQ.45.1.2.Developing

Plaut, D. C. (2005). Connectionist Approaches to Reading. In Margaret J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook*. Oxford: Blackwell Publishing.

Protopapas, A., Parrila, R., & Simos, P. G. (2016). In Search of Matthew Effects in Reading. *Journal of Learning Disabilities*, *49*(5), 499–514. https://doi.org/10.1177/0022219414559974

Puolakanaho, A., Ahonen, T., Aro, M., Eklund, K., Leppänen, P. H. T., Poikkeus, A., … Lyytinen, H. (2008). Developmental Links of Very Early Phonological Strong to Accuracy but Only Minor to Fluency. *Journal of Learning Disabilities*, *41*(4), 353–370. https://doi.org/10.1177/0022219407311747

Puranik, C. S., Lonigan, C. J., & Kim, Y. S. (2011). Contributions of emergent literacy skills to name writing, letter writing, and spelling in preschool children. *Early Childhood Research Quarterly*, *26*(4), 465–474.

https://doi.org/10.1016/j.ecresq.2011.03.002

Rack, J., Hulme, C., Snowling, M., & Wightman, J. (1994). The role of phonology in young children learning to read words: The direct-mapping hypothesis. *Journal of Experimental Child Psychology*. https://doi.org/10.1006/jecp.1994.1003

Ramus, F. (2001). Dyslexia: Talk of two theories. *Nature*, *412*(6845), 393–395. https://doi.org/10.1038/35086683

Raudszus, H., Segers, E., & Verhoeven, L. (2018). Lexical quality and executive control predict children's first and second language reading comprehension. *Reading and Writing*, *31*(2), 405–424. https://doi.org/10.1007/s11145-017-9791-8

Rhoads, C. H. (2011). The Implications of "Contamination" for Experimental Design in Education. In *Journal of Educational and Behavioral Statistics* (Vol. 36, pp. 76–104). https://doi.org/10.3102/1076998610379133

Richards-Tutor, C., Baker, D. L., Gersten, R., Baker, S. K., & Smith, J. M. (2016). The effectiveness of reading interventions for English learners: A research synthesis. *Exceptional Children*, *82*(2), 144–169. https://doi.org/10.1177/0014402915585483

Ricketts, J., Nation, K., & Bishop, D. V. M. (2007). Vocabulary is Important for ome, but not all reading skills. *Scientific Studies of Reading*, *11*(3), 235–257. https://doi.org/10.1080/10888430701344306

Rinne, L., Gregory, E., Yarmolinskaya, J., & Hardiman, M. (2011). Why Arts Integration Improves Long-Term Retention of Content. *Mind, Brain, and Education*, *5*(2), 89–96. https://doi.org/10.1111/j.1751-228X.2011.01114.x

Rose, J. (2006). *Independent review of the teaching of early reading: Final report*. London, England.

Rosowsky, A. (2001). Decoding as a cultural practice and its effects on the reading process of bilingual pupils. *Language and Education*, *15*(1), 56–70. https://doi.org/10.1080/09500780108666799

Saksida, A., Iannuzzi, S., Bogliotti, C., Chaix, Y., Dï¿½monet, J. F., Bricout, L., … Ramus, F. (2016). Phonological skills, visual attention span, and visual stress in developmental dyslexia. *Developmental Psychology*, *52*(10), 1503–1516. https://doi.org/10.1037/dev0000184

San Francisco, A. R., Mo, E., Carlo, M., August, D., & Snow, C. (2006). The influences of language of literacy instruction and vocabulary on the spelling of Spanish-English bilinguals. *Reading and Writing*, *19*(6), 627–642. https://doi.org/10.1007/s11145-006-9012-3

Saretsky, G. (1972). The OEO PC experiment and the John Henry effect. *The Phi Delta Kappan*, *53*(9), 579–581.

Savage, R., McBreen, M., Genesee, F., Erdos, C., Haigh, C., & Nair, A. (2018). Rapid automatic naming predicts more than sublexical fluency: Evidence from English-French bilinguals. *Learning and Individual Differences*, *62*, 153–163. https://doi.org/10.1016/j.lindif.2018.02.001

Savage, R., Frederickson, N., Goodwin, R., Patni, U., Smith, N., & Tuersley, L.

(2005). Relationships among rapid digit naming, phonological processing, motor automaticity, and speech perception in poor, average, and good readers and spellers. *Journal of Learning Disabilities*, *38*(1), 12–28. https://doi.org/10.1177/00222194050380010201

Savage, R., Lavers, N., & Pillay, V. (2007). Working memory and reading difficulties: What we know and what we don't know about the relationship. *Educational Psychology Review*, *19*(2), 185–221. https://doi.org/10.1007/s10648-006-9024-1

Sawilowsky, S. S. (2009). New Effect Size Rules of Thumb. *Journal of Modern Applied Statistical Methods*, *8*(2), 597–599. https://doi.org/10.22237/jmasm/1257035100

Scanlon, D. M., Gelzheiser, L. M., Vellutino, F. R., Schatschneider, C., & Sweeney, J. M. (2008). Reducing the incidence of early reading difficulties: Professional development for classroom teachers versus direct interventions for children. *Learning and Individual Differences*, *18*(3), 346–359.

Scarborough, H. S. (2001). Connecting early language and literacy to later reading (dis)abilities: Evidence, theory, and practice. In S. Neuman & D. Dickinson (Eds.), *Handbook of early literacy research* (pp. 97–110). New York: Guilford Press.

Scarborough, H. S., & Parker, J. D. (2003). Matthew Effects in Children with Learning Disabilities: Development of Reading, IQ, and Psychological Problems from Grade 2 to Grade 8. *Annals of Dyslexia*.

Scarborough, H. S. (1998). Early identification of children at risk for reading disabilities: Phonological awareness and some other promising predictors. In B. K. Shapiro, A. J. Capute, & B. Shapiro (Eds.), *Specific Reading Disability. A View of the Spectrum* (pp. 75–119). Hillsdale, NJ: Erlbaum.

Schaefer, B., Fricke, S., Szczerbinski, M., Fox-Boyer, A. V., Stackhouse, J., & Wells, B. (2009). Development of a test battery for assessing phonological awareness in German-speaking children. *Clinical Linguistics and Phonetics*, *23*(6), 404–430. https://doi.org/10.1080/02699200902770187

Schatschneider, C., Fletcher, J. M., Francis, D. J., Carlson, C. D., & Foorman, B. R. (2004). Kindergarten prediction of reading skills: A longitudinal comparative analysis. *Journal of Educational Psychology*, *96*(2), 265–282. https://doi.org/10.1037/0022-0663.96.2.265

Scheele, A. F., Leseman, P. P. M., & Mayo, A. Y. (2010). The home language environment of monolingual and bilingual children and their language proficiency. *Applied Psycholinguistics*, *31*(1), 117–140. https://doi.org/10.1017/S0142716409990191

Schiff, R., & Lotem, E. (2011). Effects of phonological and morphological awareness on children's word reading development from two socioeconomic backgrounds. *First Language*, *31*(2), 139–163. https://doi.org/10.1177/0142723710393098

Schneider, W., Roth, E., & Ennemoser, M. (2000). Training phonological skills and letter knowledge in children at risk for dyslexia: A comparison of three kindergarten intervention programs. *Journal of Educational Psychology*, *92*(2), 284–295. https://doi.org/10.10371/002243663.92.2.284

Schneider, W., Küspert, P., Roth, E., Visé, M., & Marx, H. (1997). Short- and long-term effects of training phonological awareness in kindergarten: Evidence from two German studies TT - Kurz- und langfristige phonologische Übungseffekte im Kindergartenalter: Evidenz anhand zweier deutscher Studien. *Journal of Experimental Child Psychology*, *66*, 311–340. https://doi.org/10.1006/jecp.1997.2384

Seidenberg, M. S. (2005). Connectionist models of word reading. *Current Directions in Psychological Science*, *14*(5), 238–242. https://doi.org/10.1111/j.0963-7214.2005.00372.x

Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003a). Foundation literacy acquisition in European orthographies. *British Journal of Psychology (London, England : 1953)*, *94*(Pt 2), 143–174. https://doi.org/10.1348/000712603321661859

Share, D. L. (2008). On the Anglocentricities of Current Reading Research and Practice: The Perils of Overreliance on an “;Outlier” Orthography. *Psychological Bulletin*, *134*(4), 584–615. https://doi.org/10.1037/0033-2909.134.4.584

Sheridan, S. M., Knoche, L. L., Kupzyk, K. A., Edwards, C. P., & Marvin, C. A. (2011). A randomized trial examining the effects of parent engagement on early language and literacy: The Getting Ready intervention. *Journal of School Psychology*, *49*(3), 361–383. https://doi.org/10.1038/mp.2011.182.doi

Silvén, M., Poskiparta, E., Niemi, P., & Voeten, M. (2007). Precursors of Reading Skill From Infancy to First Grade in Finnish: Continuity and Change in a Highly Inflected Language. *Journal of Educational Psychology*, *99*(3), 516–531. https://doi.org/10.1037/0022-0663.99.3.516

Smith, E. E., & Jonides, J. (1997). Working memory: A view from neuroimaging. *Cognitive Psychology*, *33*(1), 5–42. https://doi.org/10.1166/jnn.2017.13451

Snowling, M., Gallagher, A., & Frith, A. (2003). Family Risk of Dyslexia is continuous: Individual Difference in the Precursors of Reading Skill. *Child Development*, *74*(2), 358–373. https://doi.org/10.1111/1467-8624.7402003

Snowling, M. J., & Hulme, C. (2011). Evidence-based interventions for reading and language difficulties: Creating a virtuous circle. *British Journal of Educational Psychology*, *81*(1), 1–23. https://doi.org/10.1111/j.2044-8279.2010.02014.x

Snowling, M. J., & Hulme, C. (2012). Interventions for children’s language and literacy difficulties. *International Journal of Language and Communication Disorders*, *47*(1), 27–34. https://doi.org/10.1111/j.1460-6984.2011.00081.x

Snowling, M. J., & Melby-Lervåg, M. (2016). Oral language deficits in familial dyslexia: A meta-analysis and review. *Psychological Bulletin*, *142*(5), 498–545. https://doi.org/10.1037/bul0000037

Solari, E. J., & Gerber, M. M. (2008). Early Comprehension Instruction for Spanish-Speaking English Language Learners: Teaching Text-Level Reading Skills While Maintaining Effects on Word-Level Skills. *Learning Disabilities Research & Practice*, *23*(4), 155–168. https://doi.org/10.1111/j.1540-5826.2008.00273.x

Sonnleitner, P., Krämer, C., Gamo, S., Reichert, M., Mulle, C., Keller, U., & Ugen, S. (2018). *National Educational Report 2018: Schülerkompetenzen im Längsschnitt*. Luxembourg : University of Luxembourg & Service de Coordination de la Recherche et de l´Innovation pédagogiques et technologiques.

Stackhouse, J., & Wells, B. (1997). *Children's speech and literacy difficulties 1: A psycholinguistic framework*. London: Whurr Publishers.

Stahl, S. A., & Murray, B. A. (1994). Defining phonological awareness and its relationship to early reading. *Journal of Educational Psychology*, *86*(2), 221–234. https://doi.org/10.1037/0022-0663.86.2.221

Stanovich, K. E. (1986). Matthew Effects in Reading : Some Consequences of Individual Differences in the Acquisition of Literacy Matthew effects in reading : Some consequences of individual differences in the acquisition of literacy. *Reading Research Quaterly*, *21*(4), 360–407.

Storch, S. A., & Whitehurst, G. J. (2002). Oral Language and Code-Related Precursors to Reading : Evidence From a Longitudinal Structural Model. *Developmental Psychology*, *38*(6), 934–947. https://doi.org/10.1037//0012-1649.38.6.934

Streiner, D. L. (2003). Starting at the Beginning : An Introduction to Coefficient Alpha and Internal Consistency Starting at the Beginning : An Introduction to Coefficient Alpha and Internal Consistency. *Journal of Personality Assessment*, *80*(1), 99–103. https://doi.org/10.1207/S15327752JPA8001

Strong, G. K., Torgerson, C. J., Torgerson, D., & Hulme, C. (2011). A systematic meta-analytic review of evidence for the effectiveness of the "Fast ForWord" language intervention program. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *52*(3), 224–235. https://doi.org/10.1111/j.1469-7610.2010.02329.x

Stuart, M. (1999). Getting ready for reading: Early phoneme awareness and phonics teaching improves reading and spelling in inner-city second language learners. *British Journal of Educational Psychology*, *69*(4), 587–605. https://doi.org/10.1348/000709999157914

Suggate, S. P. (2016). A Meta-Analysis of the Long-Term Effects of Phonemic Awareness, Phonics, Fluency, and Reading Comprehension Interventions. *Journal of Learning Disabilities*, *49*(1), 77–96. https://doi.org/10.1177/0022219414528540

Swanson, L. (2008). Working Memory and Intelligence in Children: What Develops? *Journal of Educational Psychology*, *100*(3), 581–602. https://doi.org/10.1037/0022-0663.100.3.581

Swanson, L., Orosco, M. J., & Lussier, C. M. (2015). Growth in literacy, cognition, and working memory in English language learners. *Journal of Experimental Child Psychology*, *132*(May 2018), 155–188. https://doi.org/10.1016/j.jecp.2015.01.001

Swanson, L., Orosco, M. J., Lussier, C. M., Gerber, M. M., & Guzman-Orth, D. A. (2011). The Influence of Working Memory and Phonological Processing on English Language Learner Children's Bilingual Reading and Language

Acquisition. *Journal of Educational Psychology*, *103*(4), 838–856. https://doi.org/10.1037/a0024578

Swanson, L., Sáez, L., Gerber, M., & Leafstedt, J. (2004). Literacy and Cognitive Functioning in Bilingual and Nonbilingual Children at or Not at Risk for Reading Disabilities. *Journal of Educational Psychology*, *96*(1), 3–18. https://doi.org/10.1037/0022-0663.96.1.3

Tabachnick, B., & Fidell, L. (2012). *Using multivariate statistics* (6th ed.). Boston, MA: Pearson.

Tan, L. H., Spinks, J. A., Eden, G. F., Perfetti, C. A., & Siok, W. T. (2005). Reading depends on writing, in Chinese. *Proceedings of the National Academy of Sciences*, *102*(24), 8781–8785. https://doi.org/10.1073/pnas.0503523102

Tilstra, J., McMaster, K., Van Den Broek, P., Kendeou, P., & Rapp, D. (2009). Simple but complex: Components of the simple view of reading across grade levels. *Journal of Research in Reading*, *32*(4), 383–401. https://doi.org/10.1111/j.1467-9817.2009.01401.x

Torgerson, C. J., Brooks, G., & Hall, J. (2006). *A Systematic Review of the Research Literature on the Use of Phonics in the Teaching of Reading and Spelling*. Nottingham: Department for Education and Skills Publications.

Torgesen, J. K. (2000). Individual differences in response to early interventions in reading: The lingering problem of treatment resisters. *Learning Disabilities Research & Practice*. https://doi.org/10.1207/SLDRP1501_6

Torppa, M., Parrila, R., Niemi, P., Lerkkanen, M. K., Poikkeus, A. M., & Nurmi, J. E. (2013). The double deficit hypothesis in the transparent Finnish orthography: A longitudinal study from kindergarten to Grade 2. *Reading and Writing*, *26*(8), 1353–1380. https://doi.org/10.1007/s11145-012-9423-2

Tracey, D. H., & Morrow, L. M. (2017). *Lenses on reading: An introduction to theories and models*. New York: Guilford Press.

Treiman, R., & Bourassa, D. C. (2000). The development of spelling skill. *Topics in Language Disorders*, *20*(3), 1–18.

Treiman, R., & Kessler, B. (2005). Writing systems and spelling development. In Margaret J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 120–134). Oxford: Blackwell Publishing.

Treiman, R., & Kessler, B. (2013). Learning to Use an Alphabetic Writing System Access. *Language Learning and Development*, *9*(4), 317–330. https://doi.org/10.1016/j.neuron.2009.10.017.A

Treiman, R. (1993). *Beginning to spell*. New York: Oxford University Press.

Treiman, R. (2018). Teaching and Learning Spelling. *Child Development Perspectives*, *12*(4), 235–239. https://doi.org/10.1111/cdep.12292

Trieman, R. (2000). The Foundations of Literacy. *Current Directions in Psychological Science*, *Vol 9(3),* 89–92.

Tummer, W., & Chapman, J. (1998). Language prediction skill, phonological reoding ability, and beginning reading. In C. Hulme & R. Joshi (Eds.), *Reading and Spelling: Development and disorders* (pp. 33–67). Mahwah, NJ: Lawrence Erlbaum Associates.

Tunmer, W. E., & Hoover, W. A. (2019). The cognitive foundations of learning to read: a framework for preventing and remediating reading difficulties. *Australian Journal of Learning Difficulties*, *00*(00), 1–19. https://doi.org/10.1080/19404158.2019.1614081

Ullman, J. B. (2006). Structural equation modeling: Reviewing the basics and moving forward. *Journal of Personality Assessment*, *87*(1), 35–50.

Vadasy, P. F., & Sanders, E. A. (2010). Efficacy of Supplemental Phonics-Based Instruction for Low-Skilled Kindergarteners in the Context of Language Minority Status and Classroom Phonics Instruction. *Journal of Educational Psychology*, *102*(4), 786–803. https://doi.org/10.1037/a0019639

Van den Broek, P., Mouw, J., & Kraal, A. (2016). Individual differences in reading comprehension. In P. Afflerbach (Ed.), *Handbook of individual differences in reading: Reader, text, and context* (pp. 138–150). New York: Routledge.

Van Tuijl, C., Leseman, P. P. M., & Rispens, J. (2001). Efficacy of an intensive home-based educational intervention programme for 4- to 6-year-old ethnic minority children in the Netherlands. *International Journal of Behavioral Development*, *25*(2), 148–159. https://doi.org/10.1080/01650250042000159

Van Viersen, S., de Bree, E. H., Zee, M., Maassen, B., van der Leij, A., & de Jong, P. F. (2018). Pathways Into Literacy: The Role of Early Oral Language Abilities and Family Risk for Dyslexia. *Psychological Science*, *29*(3), 418–428. https://doi.org/10.1177/0956797617736886

Vaughn, S., Cirino, P. T., Linan-Thompson, S., Mathes, P. G., Carlson, C. D., Hagan, E. C., … Francis, D. J. (2006). Effectiveness of a Spanish Intervention and an English Intervention for English-Language Learners at Risk for Reading Problems. *American Educational Research Journal*, *43*(3), 449–487. https://doi.org/10.3102/00028312043003449

Vaughn, S., Mathes, P., Linan-Thompson, S., Cirino, P., Carlson, C., Pollard-Durodola, S., … Francis, D. (2006). Effectiveness of an English Intervention for First-Grade English Language Learners at Risk for Reading Problems. *The Elementary School Journal*, *107*(2), 153–180. https://doi.org/10.1086/510653

Vellutino, F. R., Fletcher, J. M., Snowling, M. J., & Scanlon, D. M. (2004). Specific reading disability (dyslexia): what have we learned in the past four decades? *Journal of Child Psychology and Psychiatry*, *1*, 2–40.

Verhoeven, L. (2000). Components in Early Second Language Reading and Spelling. *Scientific Studies of Reading*, *4*(4), 331–353. https://doi.org/10.1207/S1532799XSSR0404

Verhoeven, L. (2007). Early bilingualism , language transfer , and phonological awareness. *Applied Psycholinguistics*, *28*(3), 425–439.

Verhoeven, L. (1987). Literacy in a Second Language Context: Teaching immigrant children to read. *Educational Review*, *39*(3), 245–261. https://doi.org/10.1080/0013191870390306

Verhoeven, L. (2000). Components in Early Second Language Reading and Spelling. *Scientific Studies of Reading*, *4*(4), 331–353. https://doi.org/10.1207/S1532799XSSR0404

Verhoeven, L. (2011). Second language reading acquisition. In M. L. Kamil, P. D. Pearson, E. B. Moje, & P. Afflerbach (Eds.), *Handbook of reading research* (pp. 661–683). New York: Taylor & Francis.

Verhoeven, L., & van Leeuwe, J. (2012). The simple view of second language reading throughout the primary grades. *Reading and Writing*, *25*(8), 1805–1818. https://doi.org/10.1007/s11145-011-9346-3

Viladot, L., Hilton, C., Casals, A., Saunders, J., Carrillo, C., Henley, J., … Welch, G. (2018). The integration of music and mathematics education in Catalonia and England: perspectives on theory and practice. *Music Education Research*, *20*(1), 71–82. https://doi.org/10.1080/14613808.2017.1290595

Von Allmen, R. S., Weiss, S., Tevaearai, H. T., Kuemmerli, C., Tinner, C., Carrel, T. P., … Dick, F. (2015). Completeness of follow-up determines validity of study findings: Results of a prospective repeated measures cohort study. *PLoS ONE*, *10*(10), 1–13. https://doi.org/10.1371/journal.pone.0140817

Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1994). Development of Reading-Related Phonological Processing Abilities : New Evidence of Bidirectional Causality From a Latent Variable Longitudinal Study. *Developmental Psychology*, *30*(1), 73–87.

Walberg, H. J., & Tsai, S. (1983). ' Matthew ' Effects in Education. *Amercian Educational Research Journal*, *20*(3), 359–373.

Walton, P. (2014). Using Singing and Movement to Teach Pre-reading Skills and Word Reading to Kindergarten Children: An Exploratory Study. *Language and Literacy*, *16*(3), 54–77. Retrieved from http://socialiststudies.com/index.php/langandlit/article/view/23441

Wawire, B. A., & Kim, Y. S. G. (2018). Cross-language transfer of phonological awareness and letter knowledge: causal evidence and nature of transfer. *Scientific Studies of Reading*, *00*(00), 1–19. https://doi.org/10.1080/10888438.2018.1474882

Webb, S. (2007). Learning word pairs and glossed sentences : the effects of a single context on vocabulary knowledge. *Language Teaching Research*, *11*(1), 63–81.

Weber, J., Marx, P., & Schneider, W. (2007). Prävention von Lese-Rechtschreibschwierigkeiten: bei Kindern mit nichtdeutscher Herkunftssprache durch ein Training der phonologischen Bewusstheit. *Zeitschrift Für Pädagogische Psychologie*, *21*(1), 65–75.

Wechsler, D. (2002). *Wechsler preschool and primary scale of intelligence*. San Antonio, TX: The Psychological Corporation.

Weigel, D. J., Martin, S. S., & Bennett, K. K. (2006). Contributions of the home literacy environment to preschool-aged children's emerging literacy and language skills. *Early Child Development and Care*, *176*(3–4), 357–378. https://doi.org/10.1080/03004430500063747

Welch, B. L. (1951). On the Comparison of Several Mean Values: An Alternative Approach. *Biometrika*, *38*(3–4), 330–336. https://doi.org/10.2307/2332579

Werker, J. F., & Byers-Heinlein, K. (2008). Bilingualism in infancy: first steps in perception and comprehension. *Trends in Cognitive Sciences*, *12*(4), 144–

151. https://doi.org/10.1016/j.tics.2008.01.008

Westwood, P. (2018). Learning to spell: enduring theories, recent research and current issues. *Australian Journal of Learning Difficulties*, *23*(2), 137–152. https://doi.org/10.1080/19404158.2018.1524391

What Works Clearing House. (2007). *Review process: Standards*. Available from: https://ies.ed.gov/ncee/wwc/Docs/ReferenceResources/wwc_brief_baseline_080715.pdf [last accessed 10 September 2018].

Whitehurst, G. J., & Lonigan, C. J. (1998). Child Development and Emergent Literacy. *Child Development*, *69*(3), 848–872. https://doi.org/10.1111/j.l467-8624.2010.01529.x

Whitehurst, G. J., & Lonigan, C. J. (2001). Emergent literacy: Development from pre-readers to readers. In S. B. Neuman & D. K. Dickinson (Eds.), *Handbook of early literacy research* (pp. 11–29). New York: Guilford Press.

Wilkinson, L. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, *54*(8), 594–604. https://doi.org/10.1037/0003-066X.54.8.594

Wilson, B. A. (2005). Foundations: Wilson language basics for K–3. Oxford, MA: Wilson Language Training Corporation.

Wimmer, H., & Landerl, K. (1997). How learning to spell German differs from learning to spell English. In C. A. Perfetti, L. Rieben, & M. Fayol (Eds.), *Learning to spell: Research, theory, and practice across languages* (pp. 81–96). Mahwah, NJ: Erlbaum.

Wimmer, H., & Goswami, U. (1994). The influence of orthographic consistency on reading development: word recognition in English and German children. *Cognition*, *51*(1), 91–103. https://doi.org/10.1016/0010-0277(94)90010-8

Wimmer, H., Mayringer, H., & Landerl, K. (2000). The double-deficit hypothesis and difficulties in learning to read a regular orthography. *Journal of Educational Psychology*, *92*(4), 668–680. https://doi.org/10.1037/0022-0663.92.4.668

Wolery, M. (2011). Intervention Research: The Importance of Fidelity Measurement. *Topics in Early Childhood Special Education*, *31*(3), 155–157. https://doi.org/10.1177/0271121411408621

Wolf, M., & Bowers, P. G. (1999). The double-deficit hypothesis for the developmental dyslexias. Journal of educational psychology. *Journal of Educational Psychology*, *91*(3), 415–438.

Yeong, S. H. M., & Rickard Liow, S. J. (2011). Cognitive-Linguistic Foundations of Early Spelling Development in Bilinguals. *Journal of Educational Psychology*, *103*(2), 470–488. https://doi.org/10.1037/a0022437

Yeung, S. S., Siegel, L. S., & Chan, C. K. K. (2013). Effects of a phonological awareness program on English reading and spelling among Hong Kong Chinese ESL children. *Reading and Writing*, *26*(5), 681–704. https://doi.org/10.1007/s11145-012-9383-6

Young, A., & Bowers, P. G. (1995). Individual difference and text difficulty determinants of reading fluency and expressiveness. *Journal of Experimental*

Child Psychology, *60*(3), 428–454.

Zhou, L., Duff, F. J., & Hulme, C. (2015). Phonological and Semantic Knowledge Are Causal Influences on Learning to Read Words in Chinese. *Scientific Studies of Reading*, *19*(6), 409–418. https://doi.org/10.1080/10888438.2015.1068317

Ziegler, J. C., Bertrand, D., Tóth, D., Csépe, V., Reis, A., Faísca, L., … Blomert, L. (2010). Orthographic depth and its impact on universal predictors of reading: A cross-language investigation. *Psychological Science*, *21*(4), 551–559. https://doi.org/10.1177/0956797610363406

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: a psycholinguistic grain size theory. *Psychological Bulletin*, *131*(1), 3–29. https://doi.org/10.1037/0033-2909.131.1.3

Ziegler, M., Beyer, L., Schmider, E., Danay, E., & Bühner, M. (2010). Is It Really Robust? *Methodology*, *6*(4), 147–151. https://doi.org/10.1027/1614-2241/a000016